Springer Handbook *of*

Augmented Reality

Nee
Ong
Editors

Springer

Springer Handbooks

Springer Handbooks maintain the highest standards of references in key areas of the physical and applied sciences for practitioners in industry and academia, as well as graduate students. Designed to be useful and readable desk reference books, but also prepared in various electronic formats, these titles allow fast yet comprehensive review and easy retrieval of essential reliable key information. Springer Handbooks cover methods, general principles, functional relationships and fundamental data and review established applications.

All Springer Handbooks are edited and prepared with great care by editors committed to harmonizing the content. All chapters are written by international experts in their field.

Indexed by SCOPUS. The books of the series are submitted for indexing to Web of Science.

Andrew Yeh Ching Nee • Soh Khim Ong
Editors

Springer Handbook of Augmented Reality

With 526 Figures and 79 Tables

Springer

Editors
Andrew Yeh Ching Nee
Department of Mechanical Engineering
National University of Singapore
Singapore, Singapore

Soh Khim Ong
Department of Mechanical Engineering
National University of Singapore
Singapore, Singapore

ISSN 2522-8692 ISSN 2522-8706 (electronic)
Springer Handbooks
ISBN 978-3-030-67821-0 ISBN 978-3-030-67822-7 (eBook)
https://doi.org/10.1007/978-3-030-67822-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

I'm happy to have the honor of introducing the Springer Handbook of Augmented Reality.

Augmented Reality (AR) has been a topic of research for decades, but recent commercial interest and investments have made certain forms of AR available to hundreds of millions of users. AR generates the illusion, in real time, of 3D virtual objects appearing to exist in the real world around the user, either supplementing or modifying how users see their surroundings. Unlike Virtual Reality, which completely replaces what a user sees with a purely virtual 3D environment, AR retains the user's view of the real world and modifies or supplements that with 3D digital objects and information. AR offers a new paradigm of how people view and interact with both digital information and the reality around them. It enables new interfaces and experiences that require combining real and virtual. It can provide the experience of having "supersenses," where a person can see, hear, and otherwise sense things that were previously invisible and undetectable. The long-run potential of AR glasses is to replace smartphones as the dominant mobile computing platform. That is why companies are investing billions of dollars today to make these visions a reality.

This handbook provides a detailed introduction to the field of AR, through chapters contributed by many notable researchers and practitioners. The first two chapters set the context by describing AR itself and the history of the field. Chapters 3–8 describe key algorithms needed to build effective AR systems, including tracking, mapping, and interaction techniques. Chapters 9–12 focus on the hardware of AR systems, including optics, tracking systems, and wearable devices. Chapters 13–20 introduce many different consumer usages, including gaming, education, theatrical performances, museums, and tourism. However, the first usages that were explored for AR were enterprise usages, such as the maintenance and repair of complex equipment, medical usages, and military applications. Chapters 21–27 cover many of these enterprise usages, while Chaps. 28–31 focus on medical and health applications. Finally, many people are excited about the concept of the "metaverse," as an open, interoperable virtual 3D world that replaces the digital platforms we use today. A specific type of metaverse is a "mirror world" that is a digital twin of the real world, at 1:1 scale, where the Internet of Things enables the real-time digitization and connection between reality and its digital twin. AR will be the interface to these mirror worlds, so it is a key technology for making this type of metaverse successful. Chapters 32–35 describe this emerging vision of the future.

I hope you find this handbook invaluable for learning about AR and in guiding your own contributions to this exciting field.

Intel Labs Ronald Azuma

Foreword

Augmented Reality (AR) has the potential to create one of the most significant shifts in how humans and computers interact. Unlike Virtual Reality, which aims to immerse humans in a computer-generated space, AR augments human activity in the real world by seamlessly blending the digital and physical space. One of the overarching goals of Human–Computer Interaction is to make the computer invisible and for people to interact with digital content as easily as they can do with objects in the real world. AR makes this possible.

Over the last 50 years AR technology has moved from the research laboratory to the mobile phones in the pockets of billions of people and so is now widely available. AR is ideally suited for applications that have a strong spatial element, need real-time interaction, and are connected to the real world. So, there are many compelling use cases in different domains such as Engineering, Education, Entertainment and more. However, the technology still hasn't reached its full potential. For example, mobile phone-based AR doesn't provide an immersive AR view of the real world, and current head-mounted AR displays are bulky and difficult to use. Books like this are necessary to educate the next generation of people who will drive the field forward and overcome these limitations.

In terms of content, the book contains 5 sections with 35 chapters written by over a hundred authors. The first section is an overview of historical development, including a definition of what Augmented Reality is, how it is placed in context with different realities, and a history of AR technology development. The second contains chapters that outline various principles and fundamentals of AR, such as object tracking and mapping, hand pose estimation, interaction techniques, and privacy and security issues. Next is a section on hardware and peripherals, including chapters on AR displays, tracking, interaction and networking. The bulk of the book contains chapters describing a wide range of applications in the fields of Arts and Education, Engineering and Science, and Health Science. The nineteen chapters describing applications in each of these areas show the huge potential that AR has to change many aspects of our lives. Finally, the last section describes the convergence of AR with other emerging technologies of the Internet of Things and Digital Twins, both important research fields that have significant overlap with the AR space.

Many of the chapters are written by leading researchers in AR, and taken together the material provides an ideal introduction to the field. The book is suitable for people unfamiliar with AR who want to quickly understand the basics, or those already working in the field who want to refresh and fill in the gaps of some of their knowledge. Most importantly, the content provides a snapshot of the current state of the art for AR and a research roadmap for the future. It will be an excellent resource for students, researchers, and developers in this space and for all those interested in the major trends in AR.

Finally, I want to congratulate the editors of this work and all of the chapter authors. In a fast-moving field like Augmented Reality, it can be extremely challenging to create a book that will be relevant in the years after it has been published. What they have achieved is outstanding and should be on the bookshelf of anyone seriously interested in the field. I can’t wait to see what the readers of this book will create and how the future of Augmented Reality will develop as a result.

Empathic Computing Lab
University of South Australia
October 2021

Prof. Mark Billinghurst

Foreword

More than 20 years have passed since the first scholarly compilations were published about Augmented Reality (AR). Each collection marked a waterline that summarized the state of the art at the time and an updated view of the research horizon. The editors of this compilation deserve credit for selecting authors and chapters that span the current state-of-the-art achievements in AR, with each chapter serving as both an introduction and a reference for deeper study. The field is now mature enough for the opening section to summarize a significant history for a new generation of readers and researchers. The technical section presents in-depth details for a wide range of key topics, including new directions for tracking with machine learning and neural networks. The depth and breadth of the application sections are both surprising and inspiring. Several applications, such as Digital Twins, are fresh new additions to the scope of AR research. This book truly marks a new waterline for its time.

My personal history with AR began over 40 years ago. My recent research focus is on other topics, so I approach this book with a fresh perspective. While I glanced away, the ingenuity of AR researchers produced new technologies, algorithms, and system advances in the key areas of tracking, displays, and interaction. These long-standing challenge areas set the pace of progress in the field. The diversity of applications is a pleasant surprise to me. Applications provide the motivating forces for both the economic and cerebral resources focused on AR. Readers should appreciate and be inspired by the creative and challenging efforts described herein. Overall, the book leaves me personally excited and optimistic about the current state and the future of augmented reality.

Computer Science Department at the
University of Southern California
October 2021

Prof. Ulrich Neumann

Preface

Due to the rapid advancement and pervasive applications of AR technologies in various fields, Mr. Anthony Doyle mooted publishing an AR handbook, akin to many other Springer handbooks which are premium products within Springer's portfolio. Anthony made a trip to Singapore to discuss this project in mid-2018. The editors accepted the invitation and embarked on this immensely challenging project. After much discussion and planning, the call for chapters started in July 2019. It is not exactly an open call as the editors searched through AR publications in the recent years and identified potential contributors in the various fields with minimal overlapping.

The project, unfortunately, met with much delay due the worldwide onset of COVID-19. To be exact, it was like close to 2 years behind the original target. Along the way, some authors gave up and others struggled to meet timeline to wrap up their chapters. Completed chapters were rigorously reviewed and revised to meet the high standard expected of the Springer Handbook Series.

In its final form, the AR Handbook consists of 35 chapters classified into 4 categories, viz., Principles and Fundamentals, Hardware and Peripherals, Applications, and Convergence with Emerging Technologies. There are 119 contributing authors from various organizations in different countries. Many authors are trailblazers in AR research and have focused on applications in specialized fields with seminal findings. It is indeed most encouraging to see the large fraternity of AR researchers under one book cover.

It will be impossible to summarize all the contributions in one or two paragraphs, but the editors are most fortunate and honored to have received forewords from pre-eminent forerunners of AR research and technologies Ronald Azuma, Mark Billinghurst, and Ulrich Newmann. Their insights shed great lights and further illumine the handbook.

The editors sincerely hope this handbook can add to the collection of the AR community. As AR technology is continuously evolving, it is expected that newer editions of the handbook will be prepared periodically to update this exciting technology.

The editors would like to acknowledge the great assistance provided by Springer during the preparation and production of this handbook, viz., Anthony Doyle, Judith Hinterberg, Kate Emery, Heather King, and Ursula Barth.

National University of Singapore — Andrew Yeh Ching Nee
December 2022 — Soh Khim Ong

Contents

Part I Historical Developments 1

1 **Fundamentals of All the Realities: Virtual, Augmented, Mediated, Multimediated, and Beyond** 3
Steve Mann, Phillip V. Do, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang

2 **History of Augmented Reality** 35
Raffaele Vertucci, Salvatore D'Onofrio, Stefano Ricciardi, and Maurizio De Nino

Part II Principles and Fundamentals, Software Techniques, and Developments 51

3 **Principles of Object Tracking and Mapping** 53
Jason Rambach, Alain Pagani, and Didier Stricker

4 **3D Object and Hand Pose Estimation** 85
Vincent Lepetit

5 **Mixed Reality Interaction Techniques** 109
Jens Grubert

6 **Interaction with AI-Controlled Characters in AR Worlds** 131
Christoph Bichlmeier

7 **Privacy and Security Issues and Solutions for *Mixed Reality* Applications** 157
Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne

Part III Hardware and Peripherals 185

8 **The Optics of Augmented Reality Displays** 187
Aaron Bauer and Jannick P. Rolland

9 **Tracking Systems: Calibration, Hardware, and Peripherals** 211
Alexander Plopski, Naoto Ienaga, and Maki Sugimoto

10 **Embodied Interaction on Constrained Interfaces for Augmented Reality** 239
Lik-Hang Lee, Tristan Braud, and Pan Hui

11 **Networking and Cyber Foraging for Mobile Augmented Reality** 273
Tristan Braud, Wenxiao Zhang, Benjamin Finley, and Pan Hui

Part IV Applications in Arts, Education, and Culture 303

12 Augmented Reality in Arts Education 305
Chiara Panciroli, Manuela Fabbri, Alessandro Luigini, Anita Macauda, Laura Corazza, and Veronica Russo

13 Augmented Reality's Application in Education and Training 335
Stephen Marshall

14 Augmented Reality in Sports and Physical Education 355
Jia Zhang and Yang-Sheng Huang

15 Potentiating Learning Through Augmented Reality and Serious Games 369
J. M. Cerqueira, B. Cleto, J. M. Moura, C. Sylla, and L. Ferreira

16 Augmented Reality for Cultural Heritage 391
Bruno Fanini, Alfonsina Pagano, Eva Pietroni, Daniele Ferdani, Emanuel Demetrescu, and Augusto Palombini

17 Augmented Reality in Holocaust Museums and Memorials 413
Jennifer Challenor and Minhua Ma

18 Augmented Reality into Live Theatrical Performance 433
Dan Lisowski, Kevin Ponto, Shuxing Fan, Caleb Probst, and Bryce Sprecher

19 How do Tourists Evaluate Augmented Reality Services? Segmentation, Awareness, Devices and Marketing Use Cases 451
Katrin E. Schein, Marc Herz, and Philipp A. Rauschnabel

Part V Applications in Engineering and Science 471

20 Augmented Reality Uses and Applications in Aerospace and Aviation 473
Maryam Safi and Joon Chung

21 Augmented Reality for Building Maintenance and Operation 495
Lorena C. S. Moreira, Regina C. Ruschel, and Amir H. Behzadan

22 An Augmented Reality Platform for Interactive Finite Element Analysis 533
J. M. Huang, Soh Khim Ong, and Andrew Yeh Ching Nee

23 Augmented Reality in Maintenance: A Review of the State-of-the-Art and Future Challenges 575
Soh Khim Ong, C. Y. Siew, and Andrew Yeh Ching Nee

24 Augmented Reality for Maintenance and Repair 597
Salvatore D'Onofrio, Raffaele Vertucci, Stefano Ricciardi, and Maurizio De Nino

25 Augmented Reality for Naval Domains 617
Mark A. Livingston, Joshua Kvavle, Corey J. Countryman, Kendy K. Vierling, Nathan Jones, Michael J. Confessore, Ryan Bowersock, and Aristides Staikos

26 Augmented and Mixed Reality for Shipbuilding 643
Tiago M. Fernández-Caramés and Paula Fraga-Lamas

27 Augmented Reality for Remote Assistance (ARRA) 669
Riccardo Palmarini, Iñigo Fernández del Amo, Dedy Ariansyah, John Ahmet Erkoyuncu, and Rajkumar Roy

Part VI Applications in Health Science 687

28 Augmented Reality for Computer-Guided Interventions 689
Stephane Cotin and Nazim Haouchine

29 Youth and Augmented Reality 709
Elana R. Kriegel, Bojan Lazarevic, Deborah S. Feifer, Christian E. Athanasian, Nelson Chow, Jamie P. Sklar, Yaa O. Asante, Clara S. Goldman, and Ruth L. Milanaik

30 Augmented Reality-Assisted Healthcare Exercising Systems 743
Soh Khim Ong, M. Y. Zhao, and Andrew Yeh Ching Nee

31 Augmented Reality for Cognitive Impairments 765
Zhanat Makhataeva, Tolegen Akhmetov, and Huseyin Atakan Varol

Part VII Convergence with Emerging Technologies 795

32 The Augmented Reality Internet of Things: Opportunities of Embodied Interactions in Transreality 797
Kangsoo Kim, Nahal Norouzi, Dongsik Jo, Gerd Bruder, and Gregory F. Welch

33 Convergence of IoT and Augmented Reality 831
Alexis Morris and Georgina Yeboah

34 Digital Twin and Extended Reality: Strategic Approach and Practical Implementation 853
Roberto Rocca, Roman Felipe Bastidas Santacruz, Claudio Sassanelli, Paolo Rosa, Luca Fumagalli, and Elisa Negri

35 Digital Twins as Foundation for Augmented Reality Applications in Aerospace 881
Georgia Albuquerque, Philipp M. Fischer, Syed Muhammad Azeem, Andreas-Christoph Bernstein, Sebastian Utzig, and Andreas Gerndt

Index 901

About the Editors

Andrew Yeh Ching Nee is a professor at NUS. He received his PhD and DEng from Manchester and UMIST respectively. He is Fellow of the International Academy for Production Engineering (CIRP) and was elected its president in 2012, as well as Fellow of the US Society of Manufacturing Engineers and its Gold Medal recipient in 2014. He is a pioneering leader in manufacturing research and education. Currently, he serves on 20 editorial boards of refereed journals in the field of manufacturing. He is Editor-in-Chief of Springer's longstanding *International Journal of Advanced Manufacturing Technology* and Executive Editor-in-Chief of *Advances in Manufacturing*. He has published over 400 papers in peer-reviewed international journals, 250 papers in conference proceedings, and has authored and edited 22 books and 25 book chapters. He has graduated 53 PhD and 47 Master's students. He has received over 19,569 GS citations and an H-Index of 72 (June 2021). He holds honorary professorship from Tianjin, BUAA, NUAA, Shanghai, and HUST. He is Fellow of the Academy of Engineering Singapore (2012). Some of the major international awards he has received include: SME Gold Medal (2014), IEEE Kayamori Award (1999), IJPR Norman A Dudley Award (2003), IMechE Joseph Whitworth Prize (2009), and SME Outstanding Young Manufacturing Engineer Award (1982).

Soh Khim Ong is an Associate Professor in the Mechanical Engineering Department at the National University of Singapore. Her research interests are virtual and augmented reality applications in manufacturing, ubiquitous manufacturing, assistive technology, and rehabilitation engineering. She has published 7 books and over 320 international refereed journal and conference papers. She has graduated 31 PhD students and has received 9313 GS citations and an H-Index of 52 (June 2021). S.K. Ong received the 2004 Eugene Merchant Outstanding Young Manufacturing Engineer Award from the US Society of Manufacturing Engineers. In August 2012, she was elected a Fellow of CIRP, The International Academy for Production Engineering. She has also received various other awards including the 2004 Singapore Youth Award for the Science & Technology Category, the Outstanding Young Person of Singapore Award for the Science & Technology Category, the 2009 Emerging Leaders Award in Academia by the US Society for Women Engineers, and the 2013 Singapore Youth Award Medal of Commendation.

Contributors

Tolegen Akhmetov Department of Robotics, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan City, Kazakhstan

Georgia Albuquerque Institute for Software Technology, DLR Braunschweig, Braunschweig, Germany

Dedy Ariansyah School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK

Yaa O. Asante Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Christian E. Athanasian Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Syed Muhammad Azeem Institute for Software Technology, DLR Braunschweig, Braunschweig, Germany

Aaron Bauer The Institute of Optics, University of Rochester, Rochester, NY, USA

Amir H. Behzadan Department of Construction Science, Texas A&M University, College Station, TX, USA

Andreas-Christoph Bernstein Institute for Software Technology, DLR Braunschweig, Braunschweig, Germany

Christoph Bichlmeier Fakultät Informatik, Studiengänge Game Engineering (BA) und Game Engineering & Visual Computing (MA), Hochschule für angewandte Wissenschaften Kempten, Kempten, Germany

Ryan Bowersock Naval Air Warfare Center, Aircraft Division, Patuxent River, MD, USA

Tristan Braud Division of Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong, China

Gerd Bruder University of Central Florida, Orlando, FL, USA

J. M. Cerqueira Polytechnic Institute of Cávado and Ave, Barcelos, Portugal

Jennifer Challenor Games & Visual Effects, Staffordshire University, Stoke-On-Trent, UK

Nelson Chow Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Joon Chung Department of Aerospace Engineering, Ryerson University, Toronto, ON, Canada

B. Cleto Polytechnic Institute of Cávado and Ave, Barcelos, Portugal

Michael J. Confessore Naval Air Warfare Center, Aircraft Division, Patuxent River, MD, USA

Laura Corazza Department of Education Studies "Giovanni Maria Bertin", University of Bologna, Bologna, Italy

Stephane Cotin MIMESIS Team, Inria, Strasbourg, France

Corey J. Countryman Naval Undersea Warfare Center Division Keyport, Keyport, WA, USA

Jaybie A. De Guzman University of the Philippines Diliman, College of Engineering Electrical and Electronics Engineering Institute, Quezon City, Philippines

Iñigo Fernández del Amo School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK

Emanuel Demetrescu CNR ISPC, Rome, Italy

Maurizio De Nino Digitalcomoedia S.r.l, Naples, Italy

Research & Developments, Digitalcomoedia S.r.l., Naples, Italy

Phillip V. Do University of Toronto, Toronto, ON, Canada

Salvatore D'Onofrio Leonardo S.p.A, Giugliano in Campania, Italy

John Ahmet Erkoyuncu School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK

Manuela Fabbri Department of Education Studies "Giovanni Maria Bertin", University of Bologna, Bologna, Italy

Shuxing Fan Department of Theater and Drama, University of Wisconsin, Madison, WI, USA

Bruno Fanini CNR ISPC, Rome, Italy

Deborah S. Feifer Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Daniele Ferdani CNR ISPC, Rome, Italy

Tiago M. Fernández-Caramés Dept. of Computer Engineering, CITIC Research Center, Universidade da Coruña, A Coruña, Spain

L. Ferreira 2Ai – School of Technology, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal

Benjamin Finley Edge Cloud Innovation Lab, Huawei Technologies Co. Ltd., Shenzhen, China

Philipp M. Fischer Institute for Software Technology, DLR Braunschweig, Braunschweig, Germany

Luca Fumagalli Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy

Tom Furness University of Washington, Seattle, WA, USA

Paula Fraga-Lamas Dept. of Computer Engineering, CITIC Research Center, Universidade da Coruña, A Coruña, Spain

Andreas Gerndt Institute for Software Technology, DLR Braunschweig, Braunschweig, Germany

Clara S. Goldman Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Jens Grubert Coburg University of Applied Sciences, Department of Electrical Engineering and Computer Science, Coburg, Germany

Nazim Haouchine Surgical Planning Laboratory, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

Marc Herz kleinundplaecking, Berlin, Germany

J. M. Huang Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore

Yang-Sheng Huang Intelligent Interactive Technology Learning Laboratory, Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan

Pan Hui Computational Media and Arts Thrust, Hong Kong University of Science and Technology, Hong Kong, China

Division of Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong, China

Naoto Ienaga University of Tsukuba, Tsukuba, Japan

Jay Iorio IEEE Standards Association, Piscataway, NJ, USA

Dongsik Jo School of IT Convergence, University of Ulsan, Ulsan, Republic of Korea

Nathan Jones U.S. Marine Corps Systems Command, Training Systems, Quantico, VA, USA

Kangsoo Kim Department of Electrical and Software Engineering, University of Calgary, Calgary, AB, Canada

Elana R. Kriegel Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Joshua Kvavle Naval Information Warfare Center, Pacific, San Diego, CA, USA

Bojan Lazarevic Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Lik-Hang Lee Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

Vincent Lepetit LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France

Dan Lisowski Department of Theater and Drama, University of Wisconsin, Madison, WI, USA

Mark A. Livingston Naval Research Laboratory, Washington, DC, USA

Alessandro Luigini Faculty of Education, Free University of Bozen-Bolzano, Brixen-Bressanone, Italy

Minhua Ma Falmouth University, Penryn, Cornwall, UK

Anita Macauda Department of Education Studies "Giovanni Maria Bertin", University of Bologna, Bologna, Italy

Zhanat Makhataeva Department of Robotics, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan City, Kazakhstan

Steve Mann University of Toronto, Toronto, ON, Canada

Stephen Marshall Centre for Academic Development, Victoria University of Wellington, Wellington, New Zealand

Ruth L. Milanaik Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Lorena C. S. Moreira Laboratory of Advanced Studies in City, Architecture and Digital Technologies (LCAD), Faculty of Architecture, Federal University of Bahia, Salvador, Brazil

Alexis Morris Adaptive Context Environments Lab, OCAD University, Toronto, ON, Canada

J. M. Moura 2Ai – School of Technology, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal

Andrew Yeh Ching Nee Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore

Elisa Negri Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy

Nahal Norouzi University of Central Florida, Orlando, FL, USA

Soh Khim Ong Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore

Alain Pagani Augmented Vision, German Research Center for Artificial Intelligence, Kaiserslautern, Germany

Alfonsina Pagano CNR ISPC, Rome, Italy

Riccardo Palmarini School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK

Augusto Palombini CNR ISPC, Rome, Italy

Chiara Panciroli Department of Education Studies "Giovanni Maria Bertin", University of Bologna, Bologna, Italy

Eva Pietroni CNR ISPC, Rome, Italy

Alexander Plopski Graz University of Technology, Graz, Austria

Kevin Ponto Wisconsin Institute for Discovery/Design Studies, University of Wisconsin, Madison, WI, USA

Caleb Probst Curriculum and Instruction/Theater and Drama, University of Wisconsin, Madison, WI, USA

Jason Rambach Augmented Vision, German Research Center for Artificial Intelligence, Kaiserslautern, Germany

Philipp A. Rauschnabel Digital Marketing & Media Innovation, College of Business, Universität der Bundeswehr München, Neubiberg, Germany

Stefano Ricciardi Department of Biosciences and Territory, University of Molise, Pesche, Italy

Roberto Rocca Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy

Jannick P. Rolland The Institute of Optics, University of Rochester, Rochester, NY, USA

Paolo Rosa Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy

Rajkumar Roy School of Mathematics, Computer Science and Engineering, City, University of London, London, UK

Regina C. Ruschel School of Civil Engineering, Architecture and Urban Design, University of Campinas, Campinas, Brazil

Veronica Russo Department of Education Studies "Giovanni Maria Bertin", University of Bologna, Bologna, Italy

Maryam Safi Department of Aerospace Engineering, Ryerson University, Toronto, ON, Canada

Roman Felipe Bastidas Santacruz Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy

Claudio Sassanelli Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy

Katrin E. Schein Digital Marketing & Media Innovation, College of Business, Universität der Bundeswehr München, Neubiberg, Germany

Aruna Seneviratne University of New South Wales, School of Electrical Engineering and Telecommunications, Sydney, NSW, Australia

C. Y. Siew Mechanical Engineering Department, National University of Singapore, Singapore, Singapore

Jamie P. Sklar Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA

Bryce Sprecher Wisconsin Institute for Discovery/Virtual Environments Group, University of Wisconsin, Madison, WI, USA

Aristides Staikos Naval Air Warfare Center, Aircraft Division, Patuxent River, MD, USA

Didier Stricker Augmented Vision, German Research Center for Artificial Intelligence, Kaiserslautern, Germany

Maki Sugimoto Keio University, Tokyo, Japan

C. Sylla CIEC – ITI/LARSyS, University of Minho, Braga, Portugal

Kanchana Thilakarathna The University of Sydney, School of Computer Science Faculty of Engineering, Sydney, NSW, Australia

Sebastian Utzig Institute for Software Technology, DLR Braunschweig, Braunschweig, Germany

Huseyin Atakan Varol Department of Robotics, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan City, Kazakhstan

Raffaele Vertucci Leonardo S.p.A, Giugliano in Campania, Italy

Kendy K. Vierling Naval Education Division, Office of the Assistant Secretary of the Navy (Manpower and Reserve Affairs), Washington, DC, USA

Zixin Wang Sun Yat-sen University, Guangzhou, China

Gregory F. Welch University of Central Florida, Orlando, FL, USA

Georgina Yeboah Adaptive Context Environments Lab, OCAD University, Toronto, ON, Canada

Yu Yuan IEEE Standards Association, 0xSenses Corporation, Piscataway, NJ, USA

Jia Zhang Intelligent Interactive Technology Learning Laboratory, Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan

Wenxiao Zhang Computational Media and Arts Thrust, Hong Kong University of Science and Technology, Hong Kong, China

M. Y. Zhao Mechanical Engineering Department, National University of Singapore, Singapore, Singapore

Part I

Historical Developments

Check for updates

Fundamentals of All the Realities: Virtual, Augmented, Mediated, Multimediated, and Beyond

1

Steve Mann, Phillip V. Do, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang

Contents

1.1 What Is (Augmented) Reality? ... 4

1.2 Historical Background and Context ... 7
1.2.1 A Confusing Mess of Different Realities: Virtual, Augmented, Mixed, and X-Reality ... 7
1.2.2 Mediated Reality (XY-Reality) ... 9
1.2.3 Deliberately Mediated Reality ... 10
1.2.4 Unintentionally Mediated Reality ... 10
1.2.5 The Mediated Reality (X,Y) Continuum ... 10

1.3 Multimediated Reality ... 11
1.3.1 Technologies for Sensory Attenuation ... 11
1.3.2 Multimedia in Photographic Darkrooms ... 11
1.3.3 Multimediated Reality Darkroom ... 11
1.3.4 Comparison with Existing Measuring Instruments ... 17

1.4 Multimediated Reality Is Multiscale, Multimodal, Multisensory, Multiveillant, and Multidimensional ... 17
1.4.1 Multisensory Synthetic Synesthesia ... 19
1.4.2 Multidimensional Multimediated Reality ... 19

1.5 Multimediated Reality Continuum ... 21
1.5.1 Multimediated Reality is "★R" (All R) ... 23

1.6 Other Forms of Phenomenological Augmented Reality ... 23
1.6.1 History of the SWIM ... 23
1.6.2 SWIM Principle of Operation ... 24
1.6.3 Visualizing Radio Waves with SWIM and Software-Defined Radio ... 25
1.6.4 Electric Machines and SWIM ... 27

1.7 Summary and Conclusions ... 29

References ... 31

Abstract

This chapter introduces the different "realities": virtual, augmented, (deliberately) diminished, mixed, mediated, multimediated, and phenomenological. The contributions of this chapter include (1) a taxonomy, framework, conceptualization, etc., of all of the "realities"; (2) a new kind of "reality" that comes from nature itself, which expands our notion beyond synthetic realities to include also phenomenological realities; and (3) methods of using phenomenological reality as a means of visualizing as well as understanding hidden phenomena present in the world around us. VR (virtual reality) replaces the real world with a simulated experience (a "virtual" world). AR (augmented reality) allows the real world to be experienced while at the same time, adding to it, a virtual world. Mixed reality provides blends that *interpolate* between real and virtual worlds in various proportions, along a "virtuality" axis, and *extrapolate* to an "X-axis" defined by "X-Reality" (eXtended reality). Mediated reality goes a step further by mixing/blending *and also modifying* reality, including, for example, deliberate diminishing of reality (e.g., a computerized welding helmet that darkens bright subject matter while lightening dark subject matter). This modifying of reality introduces a second axis called "mediality." Mediated reality is useful as a seeing aid (e.g., modifying reality to make it easier to understand) and for psychology experiments like George Stratton's 1896 upside-down eyeglasses experiment. Multimediated reality ("all reality" or "All R") is a multidimensional multisensory mediated reality that includes not just interactive multimedia-based "reality" for our five senses but also

S. Mann (✉) · P. V. Do (✉)
University of Toronto, Toronto, ON, Canada
e-mail: mann@eecg.toronto.edu; phillip.do@alum.utoronto.ca

T. Furness
University of Washington, Seattle, WA, USA
e-mail: tfurness@uw.edu

Y. Yuan
IEEE Standards Association, 0xSenses Corporation, Piscataway, NJ, USA
e-mail: y.yuan@ieee.org

J. Iorio
IEEE Standards Association, Piscataway, NJ, USA
e-mail: j.iorio@ieee.org

Z. Wang
Sun Yat-sen University, Guangzhou, China
e-mail: wangzix@mail.sysu.edu.cn

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_1

includes additional senses (like sensory sonar, sensory radar, etc.), as well as our human actions/actuators. These extra senses are mapped to our human senses using synthetic synesthesia. This allows us to directly experience real (but otherwise invisible) phenomena, such as wave propagation and wave interference patterns, so that we can see radio waves and sound waves and how they interact with objects and each other, i.e., phenomenological reality. Moreover, multimediated reality considers not just multiple axes in addition to the X-Reality axis but also that the origin of the axes exists at zero sensory stimuli. In this way, we can account for various (virtual, augmented, etc.) realities in a sensory deprivation float tank. Consider, for example, wearing an underwater VR headset while floating in a sensory deprivation tank. Then consider an augmented reality headset while floating in an immersive multimedia pool. Consider Internet-connected immersive multimedia water therapy and music therapy pools that allow multiple users to play a hydraulophone (underwater pipe organ) while meditating together in a hydraulically multimediated collective. These are examples of what we call "fluid (user) interfaces," and they fall outside the range of any previously existing "reality taxonomy" or conceptualization.

Keywords

Virtual reality (VR) · Augmented reality (AR) · Phenomenological augmented reality (PAR) · Mixed reality · Mediated reality · Multimediated reality · Taxonomy · Wave propagation · Education · Standing waves · Sitting waves

1.1 What Is (Augmented) Reality?

Let's begin with a question that's at once both broad and deep: What is reality? What is augmented reality? There are also many new realities emerging: mediated, multimediated, mixed, phenomenal, X, Y, Z, …, so we ask more generally, "What is *reality?", where "*" is any kind of reality. Definitions of reality are usually rather circular *tau*tologies. For example, Merriam-Webster dictionary defines "reality" as "the quality or state of being real" or as something that is real.

As scientists, engineers, and mathematicians, if we take this definition quite literally, we'd define reality as that which can be expressed as a weighted sum of cosine functions, leaving the sine functions for something that isn't real, giving us something equally absurd like: "Imaginality is that which can be expressed as a sum of sine functions" (perhaps the Hilbert transform of reality?).

That same dictionary defines AR (augmented reality) as "an enhanced version of reality created by the use of technology to overlay digital information on an image of something viewed through a device (such as a smartphone camera)." Now take a look at Fig. 1.1, which is the exact opposite of this definition! The figure depicts an augmented reality sculpture. The "reality" is a discarded wireless telephone salvaged from the garbage. That reality is our wasteful society generating e-waste on a massive scale. The "augmented" part of the sculpture is the wooden plaque upon which the phone is affixed, and, in particular, the radio waves from the phone were burned onto the wood using a computer-controlled laser and lock-in amplifier. Thus we can see the radio waves that are or were being emitted from the phone. Notice how the waves are stronger (of greater amplitude) closer to the phone and weaker further from it. Note also that this augmented reality overlay is not digital, and it does not require being viewed through the (or any!) device other than normal biological unaided naked human eyes.

Thus an ordinary person with ordinary unaided vision can see the augmented reality overlay of the otherwise invisible radio waves coming from the phone, by just looking at the wooden board. We call this phenomenological augmented reality (PAR) [22] or just "phenomenal reality."

The point of this sculpture and, more generally, our point here, firstly, is that many early AR systems were completely analog. For example, early biofeedback devices [18] used miniature cathode-ray tubes and moving coil meter movements mounted in eyeglasses to create a "dashboard" of one's own body's functioning. Many of these early ideas persist in modern AR eyeglass dashboards based on pie charts (circle charts), pi charts (semicircle charts), and tau

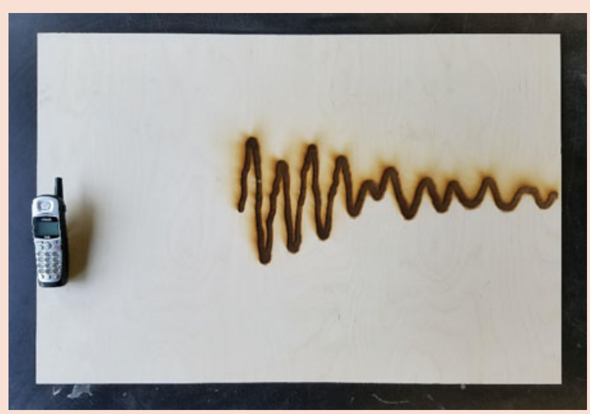


Fig. 1.1 "Burnerphone" is an augmented reality sculpture created by Mann in 2016 that pokes fun at those who try to define augmented reality. A wireless phone mounted to a board upon which is burned electromagnetic radio waves from the phone

charts (quarter-circle charts) that mimic early moving coil meter movements. Secondly, AR need not be limited to the visual or to viewing through a device, e.g., some AR can be experienced phenomenologically.

One of the simplest forms of augmented reality, for example, is phenomenological reality allowing us to see and understand otherwise invisible wave phenomena by making it pictorially visible, as shown in Fig. 1.2. The inverse discrete Fourier transform (reconstruction) may be thought of as matrix multiplication, and the rows (or columns) of the matrix are waves. For simplicity, the 4 by 4 matrix is shown, though in practice many more points are used. The inset photograph shows an oscilloscope vibrating back and forth while a dot travels around in a circle. The expression of the mathematics itself is also a form of augmented reality, i.e., through the use of concrete poetry of sorts, at the character level, i.e. *pi* as double-tau, then the introduction of tri-tau, quad-tau, etc., and the development of a visual mathematical language that leads naturally to the complex waves. Consider the following very simple example. Suppose we are trying to teach students (or our children) about sound waves and radio waves, periodic functions, Fourier series, Hilbert transforms, and the like. A very simple way to think of a sound wave is to think of it as the real part of circular motion in the Argand (complex) plane.

In 1729, Leonhard Euler proposed the use of the Greek letter π to denote "the circumference of a circle whose radius = 1", i.e., *pi* $\approx$ 6.2831853... [8]. Later he used π to denote half that quantity (approx. 3.14159...), i.e., the circumference of a circle whose diameter = 1, as well as 1/4 that quantity (approx. 1.57...). Thus the symbol was used for a full cycle, half a cycle, or a quarter of a cycle, as appropriate.

Some have proposed the Greek letter τ for a full circle, but we have suggested τ to denote a quarter of a wheel, i.e., $pi = 2\tau$, as shown in Fig. 1.2 (where we note how when π is cut in half, each half resembles τ), and the Phoenician (or Paleo Hebrew) letter teth, ⊗, for the full circle [31,34] since it originally meant "spinning wheel" [1] and in visually looks like a wheel with spokes. We have also suggested using half the wheel to denote Pi and a quarter of the wheel to denote $pi/2$, and so on [34].

However, in the true Eulerian tradition of redefining symbols to suit a given task, let us here, instead, define a visual fraction of a circle to represent the natural exponential base, e, raised to the square root of minus one, times that fraction of the circle, to represent the corresponding number in the Argand (complex) plane.

This way of thinking is very much visually inspired, for it lends itself to seeing and understanding waves in terms of rotating circles.

Continuing along this line of reasoning, we may visualize arbitrary periodic waveforms. Consider, for example, an approximation of a square wave. The typical approach is to consider rotating phasors that add up to generate a waveform whose real part is the square wave. Typically the phasors all rotate counterclockwise, and thus the imaginary part is the Hilbert transform of the square wave. However, we choose to consider phasors traveling alternately clockwise and counterclockwise such as to generate an approximation to a square wave with the imaginary part also a square wave but shifted a quarter cycle. Such a construct may be regarded as an augmented reality conceptualization of how, for example, a stepper motor works with square waves in quadrature (see Fig. 1.3). More generally, consider

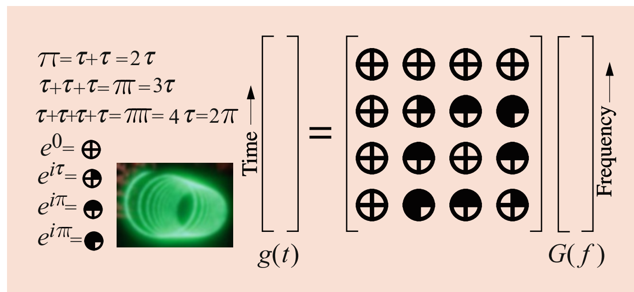


Fig. 1.2 A very simple form of augmented reality that simply allows us to see periodic functions like sound waves by way of spinning light sources vibrating back and forth in a darkened room

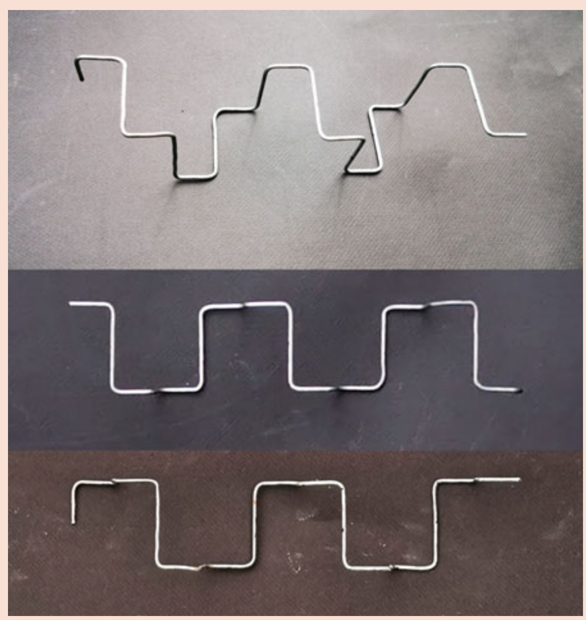


Fig. 1.3 Augmented reality wire sculpture by Mann. A steel wire was bent into a shape such that when viewed from above produces a square wave. When viewed from the front/side a square wave shifted one-quarter of a period is seen. This is a physical embodiment of the augmented reality depicted in Fig. 1.4

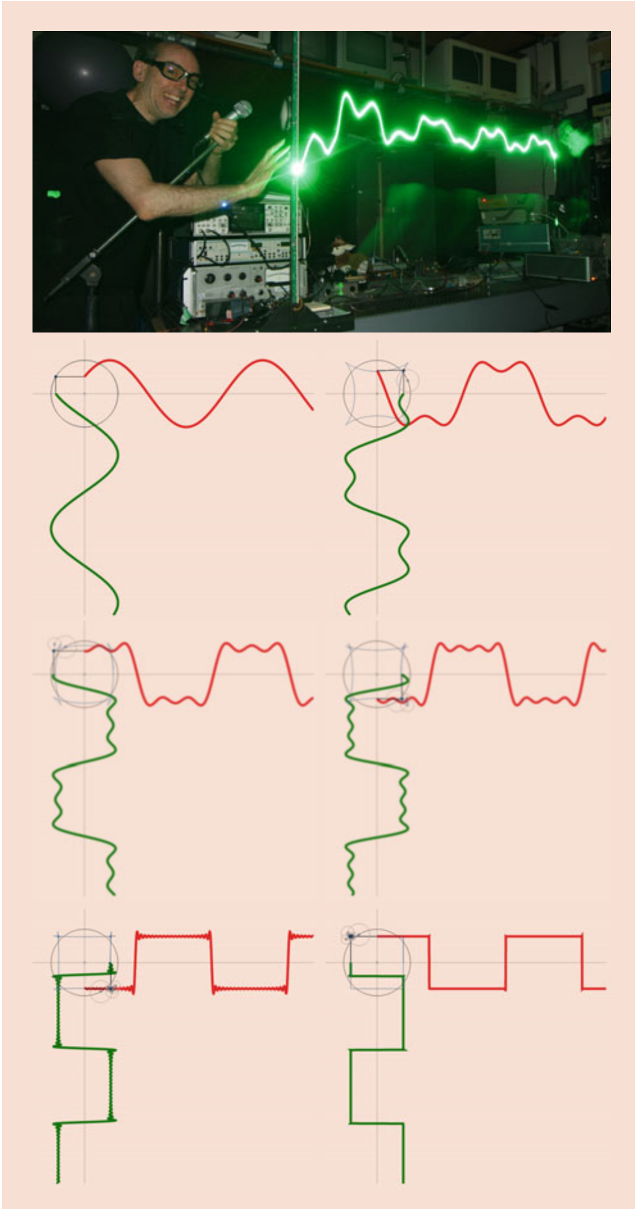


Fig. 1.4 Phenomenal augmented reality showing a microphone's capacity to "listen," e.g., to a square wave stimulus, showing the first two harmonics. Because the augmented reality is complex, we proffer a "truthful sines" approach, wherein the imaginary part of the reality ("imaginality") is in accordance with Eqs. 1.1, 1.2, and 1.3 rather than the Hilbert transform of the real part. Drawings are shown for up to 1st, 3rd, 5th, 7th, 39th, and 3999th harmonics, wherein red denotes the real part and green the imaginary part. S. Mann and C. Pierce modified a Javascript program from Pierre Guilleminot's bl.ocks.org/jinroh to bring it in line with S. Mann's Eqs. 1.1, 1.2, and 1.3

$$\mathcal{M}_n = \cos(n\omega t) + i\sin(n\omega(t - \text{ʊ}/4) + \text{ʊ}/4), \quad (1.1)$$

where $i = \sqrt{-1}$ and $\text{ʊ} = 2\pi = 360\,\text{deg} = 1$ cycle, i.e., we're shifting by 1/4 cycle, scaling, then unshifting by 1/4 cycle.

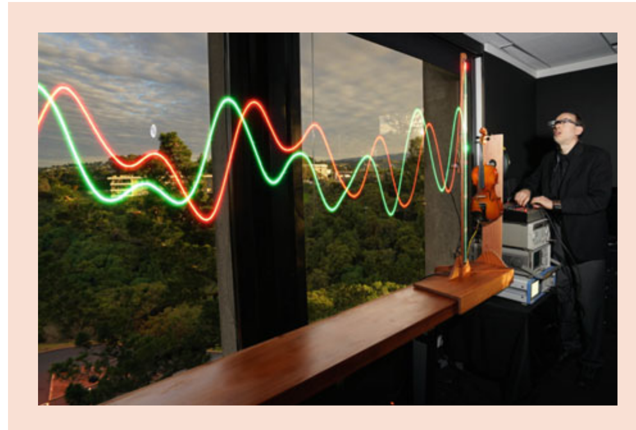


Fig. 1.5 Photograph showing actual sound waves coming from a violin. This form of augmented reality uses a row of LEDs that move back and forth to make visible sound waves as if they were sitting still [22]. The real part of the sound wave is out by red LEDs. The imaginary part which is traced out by green LEDs is *not* the Hilbert transform of the real part but, rather, follows the philosophy outlined in Figs. 1.3 and 1.4

This philosophy is also embodied in the SYSUxMannLab Model 1024SO Scientific Outstrument™ lock-in amplifier, as shown in the photograph of Fig. 1.5. This amplifier is specifically designed for phenomenal augmented reality, i.e., augmented reality of or pertaining to physical phenomena. In particular, unlike other lock-in amplifiers, it uses as its reference signal the function:

$$\bar{\mathcal{M}}_n = \sum_{k=1}^{n} \mathcal{M}_k. \quad (1.2)$$

Consider, for example, a sound wave of the form $\cos\omega t$ that is to be set in motion as a traveling wave, i.e., we replace t with $t - \delta t$ to cause a time shift to the right by δt, and, in particular, the time shift is to depend on space as $\delta t = x/c$ where x is the spatial variable (consider for simplicity a wave that is a function of one spatial and one-time dimension), and c is the speed of propagation (e.g., speed of sound in the air if the sound wave is in air, or speed of sound in water if the sound wave is in water, or speed of light if the augmented reality is an electromagnetic radio wave), etc. Now $\cos(\omega(t - x/c)) = \cos(\omega t = kx)$ where k is the wave vector, as is the familiar case. Where this becomes more interesting is when we consider a multicomponent wave having various harmonics in it, e.g., $\cos(n\omega t)$ becomes $\cos(n\omega(t - x/c)) = \cos(n\omega t - nkx)$.

Consider a specific example, such as a square wave having Fourier series $s(t) = \cos(\omega t - kx) - 1/3\cos(3\omega t - 3kx) + 1/5\cos(5\omega t - 5kx) - 1/7\cos(7\omega t - 7kx) + \ldots$.

The special augmented reality lock-in amplifier multiples this sound wave by the reference signal $\bar{\mathcal{M}}_n$, and in fact we can adjust n so as to be able to examine the waveform up to a certain number of harmonics (and more generally the amplifier has settings to adjust the weights of each harmonic if desired).

Consider as a very simple example, $\overline{\bowtie}_3$. Multiplying the various terms together, only the low-frequency terms that are primarily a function of space, rather than time, appreciably come through the low pass filter of the amplifier, thus giving us

$$\overset{\sqcap}{\bowtie}_3 = \cos(kx) - \frac{1}{3}\cos(3kx) + i(\sin(kx) + \frac{1}{3}\sin(3kx)), \quad (1.3)$$

i.e., a function of space rather than of time, which (as described in [23]) is a form of augmented reality in which the spacetime continuum is sheared in such a way that the speed of sound (or light or the like) is exactly zero, such that we can see and photograph sound waves (or radio waves) sitting still.

Finally, consider the sculpture shown in Fig. 1.6. Since the waves are in fact sitting still, we may physically realize them in a fixed form such as furniture, art, sculpture, or the like. This sculpture, named Wavetable™, is a piece of wooden furniture (a table) that provides an augmented reality overlay of otherwise invisible sound waves from a musical instrument located to the left of the table.

The purpose of these very simple examples is to provide a broader intellectual landscape in regard to what is meant by augmented reality and, more broadly, by reality itself.

Summarizing the last section, we have carefully chosen some simple examples of realities that make visible phenomena that are otherwise invisible. We've also made visible some phenomena which are complex, i.e., by definition, that which is not real!

We aim here not so much to answer the question "What is reality?", but, perhaps more importantly, as Baldwin once said, "The purpose of art is to lay bare the questions that have been hidden by the answers." However, let us proffer the following definition:

> Augmented Reality (AR) is a means, apparatus, or method of real-time superposition (adding, overlaying) of technologically-generated content onto, and aligned with, or pertaining to, a user's perception of reality. There is no requirement that that content be digital, or that it even be transient or ephemeral.

Fig. 1.6 A wavetable sculpture was carved robotically following the sound waves of a musical instrument sitting to the left of the table. In this way, the table forms an augmented reality overlay of the instrument's sound waves. Notice that the augmented reality experience is not digital and does not even require electricity

1.2 Historical Background and Context

Historically, augmented realities were not necessarily digital. For example, biofeedback systems built into eyeglasses often used miniature CRTs (cathode-ray tubes) and miniature moving coil meters. Moving coil meters, for example, sweep out a certain angular range of motion and thus give rise to a display that is part of a circle. Most D'Arsonval/Weston-type moving coil meters have a needle indicator that swings through a 90° angle, but some embody a unique design that allows an approximately 180° swing, and these were favored for augmented reality displays and form the basis for some modern display aesthetics, using what we proffer to call a "pi chart" (where a half circle denotes "all" or "maximum" = 100%). For a more modern take on this concept, see Fig. 1.7.

1.2.1 A Confusing Mess of Different Realities: Virtual, Augmented, Mixed, and X-Reality

Recently there has been an explosion of a plethora of different "realities" each with a variety of different and sometimes conflicting definitions.

VR (virtual reality) is a computer-generated simulation of a realistic experience. Typically VR blocks out the real world ("reality") and replaces it with a "virtual" world. The virtual world may be generated by a computer or by interactively playing back recorded media. An example of the latter is the Aspen Movie Map of 1978 that used computers to play back analog laser disk recordings to render an interactive virtual world as hypermedia [42] or, more recently, Google Street View with Earth VR.

AR (augmented reality) is a similar concept, but instead of blocking out reality, the computer-generated content is added onto, or embedded into, the real-world experience, so that both can be experienced together [4].

In 1994, Milgram and Kishino [40] conceptualized that augmented reality exists along a continuum between the real and virtual worlds, giving rise to "mixed reality." This continuum is called the mixed reality continuum (see Fig. 1.8). The blue arrow is suggestive of a one-dimensional "slider" or "fader" that "mixes" between the real world and the virtual world, as illustrated in Fig. 1.9, where a disk jockey (DJ)

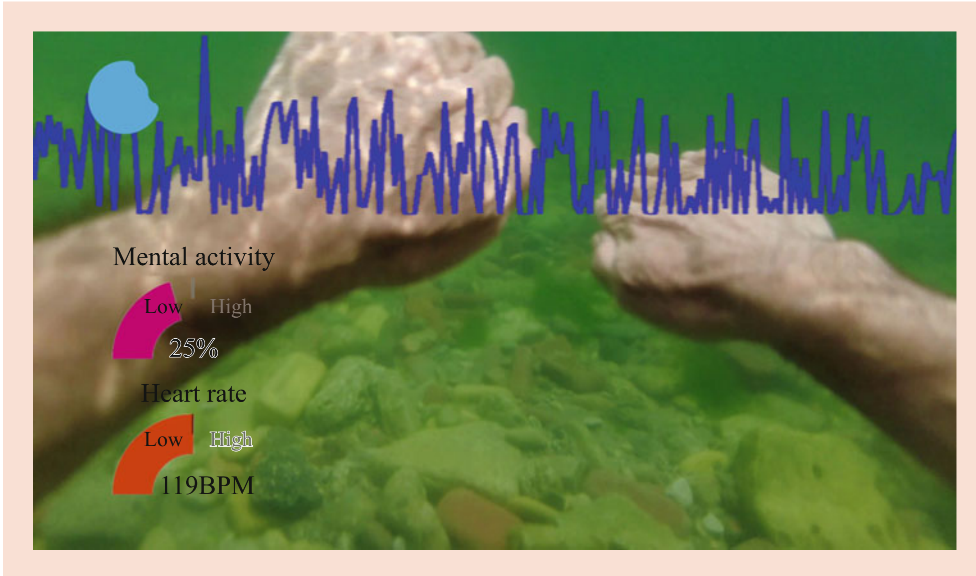


Fig. 1.7 Biofeedback of the EyeTap / BlueberryX brain-and-world-sensing eyeglass, which is an example of the new field of Water-Human-Computer-Interaction (WaterHCI) that originated in Canada in the 1960s and 1970s. Note the use of pi charts (π-charts), half circles, where π is 100%

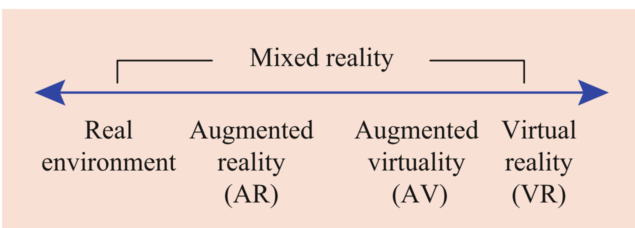


Fig. 1.8 Mixed reality continuum, adapted from Milgram and Kishino [40]

mixer is used as a metaphor. Imagine two record players (turntables), feeding into an audio/video mixer. Real-world and virtual world mixtures are selected by sliding a one-dimensional "fader" left or right. This allows us to choose various points along an "$\mathbb{X}$" axis between the extremes of reality, "$\mathbb{R}$", and virtuality, "$\mathbb{V}$." In this context we can think of AR as a setting on a "mixer" or "fader" or "slider" that is somewhere between reality and virtuality.

This "slider" is analogous to the "X-axis" of an X-Y plot or graph, treating "X" as a mathematical variable that can assume any quantity on the real number line. Thus mixed reality is sometimes referred to as "X-Reality" or "XR" [5, 18, 43]. Specifically, a 2009 special issue of IEEE "PERVASIVE computing" on "Cross-Reality Environments" defines X-Reality as a proper subset of mixed reality. Paradiso and Landay define "cross-reality" as:

> The union between ubiquitous sensor/actuator networks and shared online virtual worlds…We call the ubiquitous mixed reality environment that comes from the fusion of these two technologies cross-reality. [43].

In that same issue of IEEE "PERVASIVE computing," Coleman defines "cross-reality" and "X-Reality" as being identical:

> Cross-reality (also known as X-Reality) is an informational or media exchange between real-and virtual-world systems. [5].

XR as extrapolation ("extended response") dates back as early as 1961 when Charles Wyckoff filed a patent for his silver halide photographic film [55]. This extended response film allowed people to see nuclear explosions and other phenomena, which have high dynamic range beyond the range of normal human vision [14, 54, 56]. In 1991, Mann and Wyckoff worked together to introduce the concept of "X-Reality" and "XR vision" in the context of wearable computers (AR/VR headsets, etc.) for human augmentation and sensory extension by way of high dynamic range (HDR) imaging combined with virtual/augmented reality [18, 39].

The terms "XR," "X-Reality," "X-REALITY," and "XREALITY" appear as trademarks registered to Sony Corporation, filed in 2010, and used extensively in the context of mobile augmented reality across Sony's "Xperia" X-Reality™ for mobile products, as shown in Fig. 1.10 below:
Sony's use of XR and X-Reality is consistent with the Wyckoff-Mann conceptualization of extended human sensory perception through high dynamic range imaging.

There is some confusion, though, since XR (X-Reality) now has at least three definitions, one in which it is a proper *superset* of mixed reality, another in which it *is* mixed reality, and another in which it is a proper *subset* of mixed reality. We shall enumerate and classify these, chronologically, as follows:

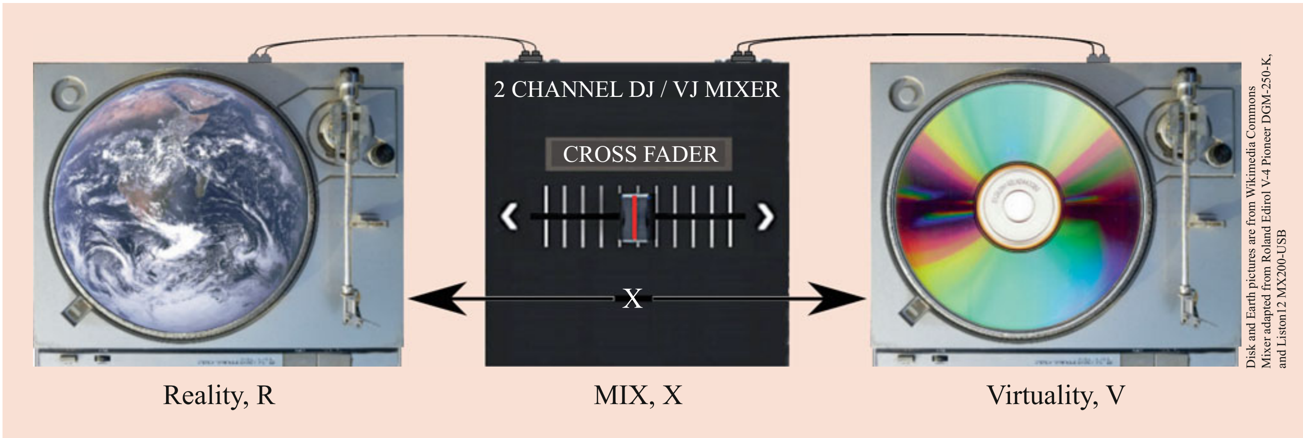


Fig. 1.9 Disk jockey (DJ) mixer metaphor for the mixed reality continuum

X-Reality™ SONY XPERIA XR™

Fig. 1.10 Sony's trademarked X-Reality and XR

- **Type 1 XR/X-Reality** in which "X" as a mathematical variable, i.e., any number on the real number line, that defines an axis for either:
  - **Type 1a XR/X-Reality: extrapolation**, i.e., XR/X-Reality in Wyckoff-Mann sense [39], as technologies that extend, augment, and expand human sensory capabilities through wearable computing. In this sense "X" defines an axis that can reach past "reality."
  - **Type 1b XR/X-Reality: interpolation**, i.e., XR/X-Reality in the Milgram sense [40], as technologies that augment human senses by creating a blend (mixture) between the extremes of reality and virtuality. In this sense "X" defines an axis that miXes (interpolations) between reality and virtuality, without reaching past reality.
- **Type 2 XR/X-Reality** in which "X" means "cross" in the Paradiso-Landay/Coleman sense [43], i.e., as a form of mixed reality (a proper subset of mixed reality) in which the reality portion comes from sensor/actuator networks and the virtuality portion comes from shared online virtual worlds.

The taxonomy of these three definitions of XR (X-Reality) is summarized as a Venn diagram in Fig. 1.11, showing also XY-Reality (XYR) which will be introduced in the next section. Leftmost: XR1a, introduced in 1991, is the most general of the X-realities. XR1b is identical to mixed reality. XR2, introduced in 2009, is a subset of mixed reality and specifically refers to the combination of "wearables" and "smart cities."

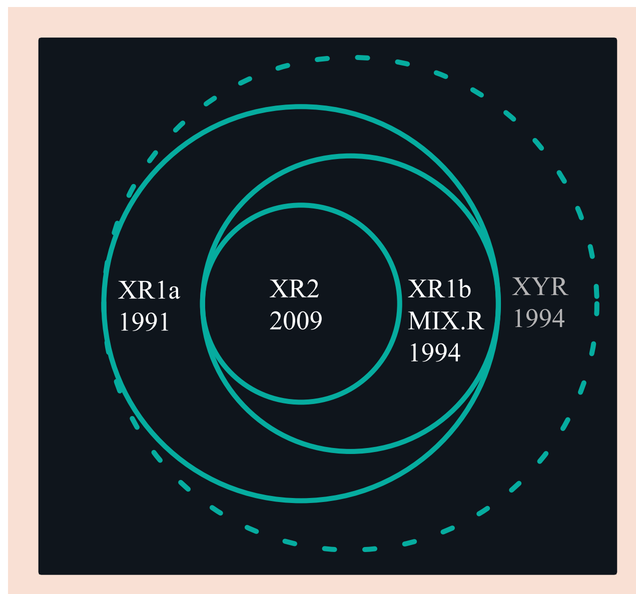


Fig. 1.11 Taxonomy (Venn diagram) of the three definitions of XR (X-Reality) as well as XYR (XY-Reality)

What these definitions of XR/X-Reality all have in common is that XR/X-Reality defines an "X-axis" defining a number line that passes through both "reality" and "virtuality," consistent with the Mann-Wyckoff 1991 conceptualization.

1.2.2 Mediated Reality (XY-Reality)

Many technologies function as an intermediary between us and the environment around us. Technology can modify or change (mediate) our "reality," either as a result of deliberate design choices of the technology to mediate reality or sometimes as an accidental or unintended side effect. These two variants of mediated reality are further discussed below.

1.2.3 Deliberately Mediated Reality

Examples of deliberate modification of reality include the upside-down eyeglass invented in 1896 by George Stratton to study the effects of optically mediated vision on the brain [50]. More recently others have done similar experiments with deliberate mediation of reality, such as left-right image reversing eyeglasses [7]. Multimedia devices such as hand-held camera viewfinders have also been used to study long-term adaptation to a photographic negative view of the world in which light areas of the image are made dark, and dark areas of the image are made light [2]. Computer-mediated reality has also been explored [15, 19].

Mediated reality is not just for psychology experiments, though. It has many practical everyday applications such as eyeglasses that filter out advertisements and, more generally, helping people see better by getting rid of visual clutter. HDR (high dynamic range) welding helmets use computer vision to *diminish* the otherwise overwhelming brightness of an electric arc while *augmenting* dark shadow detail. In addition to this mediated reality, the HDR welding helmet also adds in some virtual content as well [30].

Mediated reality has also been examined in the context of wearable computing, prosthesis, and veillance (surveillance, sousveillance, metaveillance, and dataveillance) [9, 51].

1.2.4 Unintentionally Mediated Reality

In augmented reality, there is often an attempt made to not alter reality at all. But when we experience augmented reality through a smartphone or tablet, or by using video see-through eyeglasses, the simple fact that we have a technology between us and our outside world means that the virtual objects overlaid onto reality are being overlaid onto an unintentionally modified reality (i.e., both the virtual and real objects are presented by a video display device). Thus the mediated reality framework is directly applicable to the research and understanding of video see-through implementations of AR.

It has also been observed that head-mounted displays (HMDs) can cause semipermanent or permanent and lasting harm (e.g., brain damage) or result in good, such as in the treatment of PTSD (posttraumatic stress disorder), phobias, and in treating (reversing the effects of) brain damage [44, 45, 47, 49, 53]. The fact that HMDs can both damage the brain and treat brain damage suggests that we need to be extra careful when using technology as an intermediary and that there is special importance to mediality in general.

1.2.5 The Mediated Reality (X,Y) Continuum

Summarizing the previous two subsections, consider either of the following:

- Devices and systems designed to intentionally modify reality.
- The unintended modification of reality occurs whenever we place any technology between us and our surroundings (e.g., video see-through augmented reality).

Both of these situations call for at least one other axis beyond the mix between reality and virtuality.

Moreover, the above are just two of the many more examples of "reality" technologies that do not fit into the one-dimensional "mixer" of Fig. 1.8. Thus we need at least one additional axis when describing technology that specifically modifies reality. For this reason, mediated reality [11, 15, 51, 52] has been proposed. See Fig. 1.12. In this mediated reality taxonomy (continuum), there are two axes: the virtuality axis ("X"), exactly as proposed by Milgram, and a second axis, the mediality axis ("Y"). This allows us to consider other possibilities like augmented mediated reality ("augmediated reality") such as HDR welding helmets [30], as well as

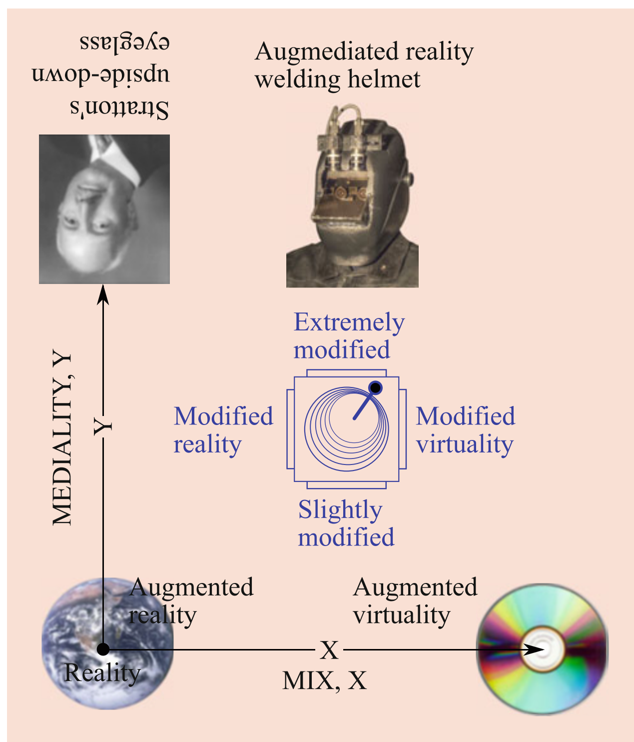


Fig. 1.12 Mediated reality (X,Y) continuum: There exists a continuum in both the degree to which reality can be virtualized and the degree to which it can be modified (e.g., the "upside-down eyeglass" experiments George Stratton did in 1896 and 1897) [33]

mediated virtuality (e.g., taking an existing VR system and then flipping the image upside-down, to allow us to repeat George Stratton's 1896 upside-down eyeglass experiment but in a virtual world).

The mediated reality (X,Y) continuum in Fig. 1.12 illustrates that there exists a continuum in both the degree to which reality can be virtualized and the degree to which it can be modified. The "MIX" axis ("$\mathbb{X}$" axis) runs left to right (reality to virtuality). The "MEDIALITY" axis ('$\mathbb{Y}$") runs bottom to top (slightly modified to extremely modified). George Stratton's upside-down glass is an example of a wearable eyeglass technology that involves no virtuality but a great deal of mediality and thus occupies an area in the upper left. The EyeTap [27] HDR welding helmet [30] is an example of extreme reality modification (mediality) that also involves a moderate amount of virtuality. The amount of virtuality it has is about the same as a typical augmented reality setup, so it exists near the top middle of the space. This top middle area of the continuum is sometimes called "augmediated reality" (augmented mediated reality) [12,41].

1.3 Multimediated Reality

The Milgram continuum [40] and the Mann continuum [15], both introduced in 1994, are shown in Figs. 1.8 and 1.12, respectively. Both continuums place reality at the left or the lower left, i.e., the "origin" in Cartesian coordinates.

Neither Milgram's nor Mann's continuum directly addresses visual sensory attenuation technologies, like sunglasses and sleep masks, or attenuation of other senses by technologies such as earplugs or sensory attenuation tanks (also known as "sensory deprivation tanks" or "flotation tanks").

Other visual useful sensory attenuation devices include the sun visor of a car, the brim of a baseball cap, or the "blinders" attached to a horse's bridle so that the horse is not distracted by peripheral motion cues.

1.3.1 Technologies for Sensory Attenuation

Sensory attenuation technologies form an underexplored yet richly and interesting space for technology. Consider, for example, some of the following possibilities:

- Interactive sleep masks for shared lucid dreaming.
- Interactive multimedia bathing environments like computer-mediated sensory attenuation tanks and interactive relaxation tanks that use water for sensory attenuation [33] and "immersive multimedia" and "fluid user interfaces" [21] that use water to alter the senses in conjunction with interactive multimedia. See Figs. 1.13 and 1.14.
- Interactive darkroom experiences such as interactive light painting with reality-based media such as persistence of exposure and "phenomenological augmented reality" (e.g., being able to see radio waves and see sound waves by way of a darkroom environment with eyes adjusted to the dark).

1.3.2 Multimedia in Photographic Darkrooms

A light bulb waved around in a dark room will tend to create the visual appearance or impression of shapes, curves, patterns, etc., by way of a "persistence of exposure" effect in human vision, in photography, or videographic media. There is a long history of photographic "light painting" (http://lpwa.pro/event/15). There is also a well-established "flow arts" community doing artistic dances in a dark environment with light sources, e.g., LED (light-emitting diodes), and "fire spinning" and juggling light-emitting objects as a medium of creative expression. Flow art is similar to light painting but for direct viewing rather than through photography. Some tools (specialized light sources) and devices are used for both light painting and flow arts.

The tradition of darkness (sensory attenuation) combined with sensory media (e.g., controlled lighting) dates back to the early days of theatre. Theatrical productions typically take place in a space in which all or most of the walls are painted black, and there is usually a black floor, and black curtains, such that lighting can be controlled carefully. In fact, the world's first use of the term "virtual reality" came from theatre in 1938 [3], where Antonin Artaud uses the term to describe theatre as being fictitious and illusory (e.g., lighting, stage settings, etc.). Artaud's usage of VR is akin to modern VR, where modern VR is a computer-generated simulation of reality and is therefore illusory as well.

The tradition of sensory attenuation and controlled sensory stimulus connects well to multimedia: Morton Heilig produced the "Sensorama" (U.S. Pat. #3050870), the world's first multimodal, multisensory experience. The Sensorama was an early example of "3D film" combined with other sensory modalities, i.e., what might now be called "multimedia." It also included an oscillating fan so the participant would a simulated wind blowing.

1.3.3 Multimediated Reality Darkroom

In the 1970s, the idea of an interactive darkroom was taken a step further, by conducting a series of experiments to make otherwise invisible phenomena visible. These experiments

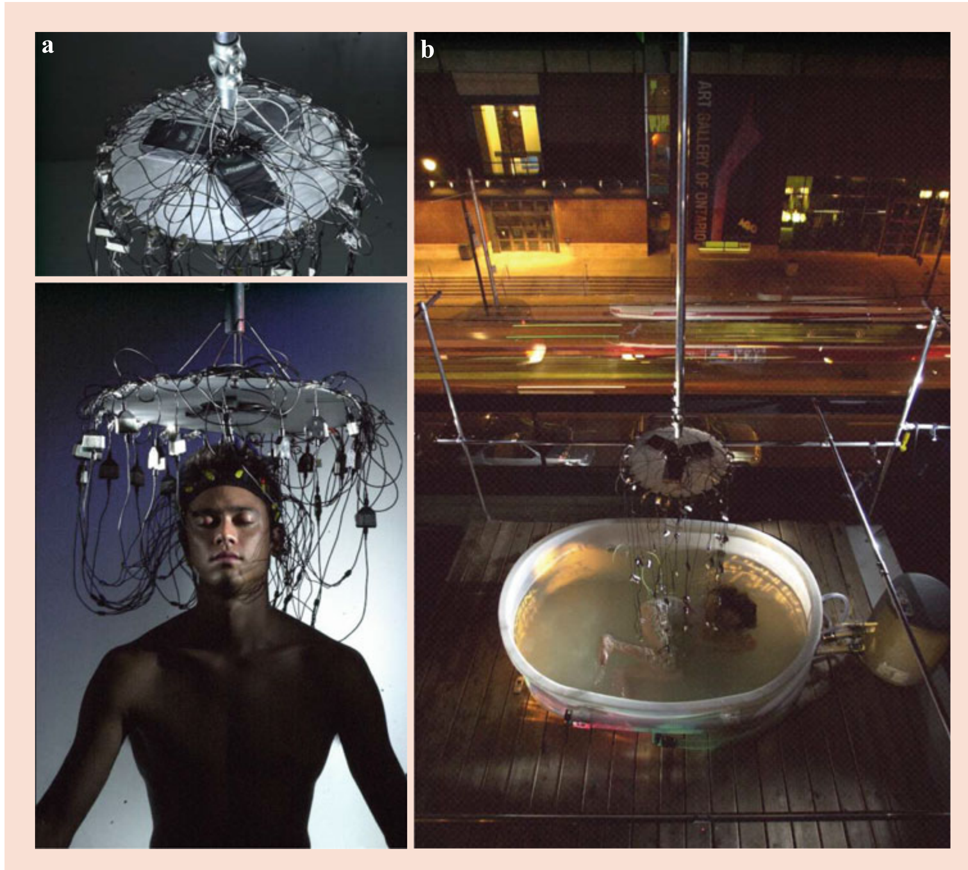


Fig. 1.13 Submersive reality (SR): meditation in a VR/AR/MR Flotation tank. (**a**) A waterproof 24-electrode underwater EEG (electroencephalogram) system is part of the many multidimensional inputs to a real-time computer system. (**b**) An interactive multimedia environment with water pumps, waves, vibration, lighting, and data projectors controlled by brainwave mediation/meditation [20, 29, 33]

involved light bulbs connected to the output of considerably powerful amplifiers that were driven by transducers or antennae that sensed a physical quantity of interest.

A special kind of lock-in amplifier (LIA) was designed specifically for multimediated reality. This special LIA can be used to scan a space using a 3D positioner, and then the resulting wave fronts are experienced on a wearable computing system, as well as on video displays in the environment. Wearers of the computer system can experience a multidimensional space (X, Y, Z, time, phase, real, imaginary, etc.). Those not wearing the computer system can still experience a reduced dimensional slice through this multidimensional space using ambient display media.

In one example, a light bulb was used to "sweep" for video "bugs." The light bulb glowed more brightly when in the field of view of a surveillance camera than it did when not in the camera's field of view [25]. This method works very simply by using video feedback: a receive antenna is connected to the input of a very sensitive lock-in amplifier with extremely high gain. The lock-in amplifier was specially designed to drive high capacity loads such as powerful electric light bulbs (1500–2500 watts).

A very sensitive lock-in amplifier picks up weak radio signals, amplifies them, and drives a light bulb, causing video feedback. The amplifier "locks in" on the signal that is due to itself and thus glows brilliantly whenever it is within the field of view of the surveillance camera. In this way, waving the light bulb back and forth "paints" out an image that reveals the otherwise hidden "sight field" of the camera.

When the light shines on a camera, it causes the camera to exhibit small but measurable changes, causing video feedback. Waving the light bulb back and forth in front of the camera makes the "sight field" of the camera visible. See Fig. 1.15. This is considered to be a special form of augmented reality [25] in the sense that it comes directly from nature itself. Unlike many other forms of realities, it does not come from a computer simulation. Here the light bulb filament has a dual role: it is both the mechanism by which a physical quantity is sensed, and it is also the display mechanism. Therefore, due to the fundamental physics of the situation, the alignment between the "real" physical world and the "augmented" world is exact; there is no need for any tracking mechanism since the process itself is self-tracking.

We proffer to call this "phenomenological reality" or "phenomenal reality" because it makes visible the true physical quantities of otherwise invisible physical phenomena. This is done by way of direct physical means, for example, a direct connection between a physical quantity and the sensed quantity of various phenomena.

When the quantity we wish to sense is not the same thing as light itself, we can use a separate sensor but mount it directly to the light bulb. For example, Fig. 1.16 shows a

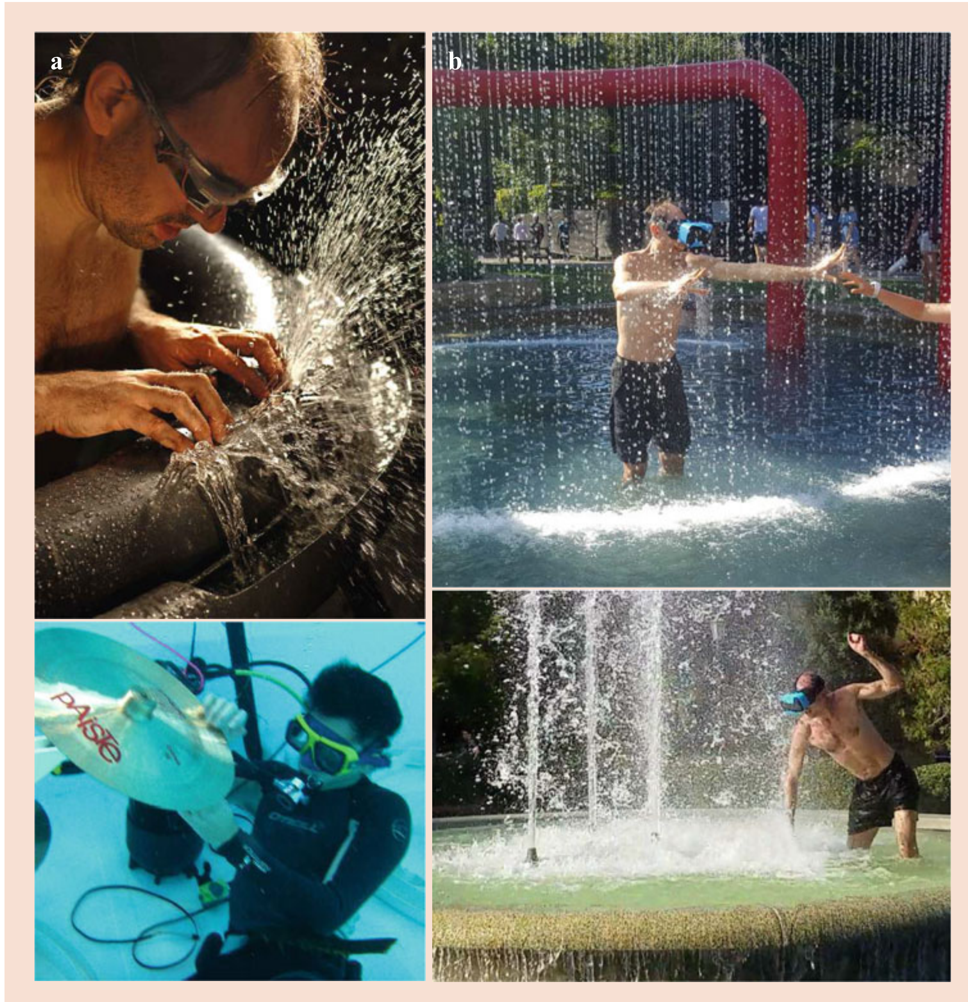


Fig. 1.14 Also, submersive reality (SR) was developed for (**a**) musical performances, underwater concerts, etc., and (**b**) physical fitness, e.g., swimming in virtual environments and interaction with water jets [20, 29, 33]

light bulb being moved by an XY plotter, driven by the output of a lock-in amplifier. It is important to emphasize here that the path the light bulb takes is in perfect alignment with the physical quantity and thus provides us with a direct glimpse into the physical nature of the signal that is being shown. For example, we can measure the distance between cycles of the waveform as being exactly (300,000,000 m/s)/(10.525 GHz) = 2.85 cm. Note the hand grasping the bulb, making it possible for a human user to feel the radio wave in addition to seeing it. Thus we have a visual and haptic multimediated reality experience that extends our senses beyond the five human senses by mapping this "sixth sense" (radio) onto two or three (we can also hear it) existing senses. Here, the quantity that we're measuring is electromagnetic energy, phase-coherently, with an antenna affixed directly to the light bulb. A special technique is used to transform the space in such coordinates that the speed of light is exactly zero, and thus we can see the radio wave sitting still. The first such photograph of radio waves was captured using a linear array of light bulbs, controlled by a wearable computer, in 1974 [22]. See Fig. 1.17 (see [23] for an explanation of the methodology of photography of radio waves and sound waves).

Figure 1.18 shows an example of phenomenological reality directed to seeing sound. Here a linear array of 600 LEDs forms a simple dot graph indicator of voltage (zero volts at the bottom and 5 volts at the top) and is connected to the output of a lock-in amplifier driven by a moving sound transducer referenced against a stationary sound transducer. The relative movement between the two transducers causes corresponding relative movement along the phase front of the sound wave, thus making it visible in coordinates in which the speed of sound is exactly zero [23]. Because the waveform is driven directly by fundamental physics (e.g., nature itself), we can use it as a scientific teaching tool from which direct measurement may be made. The array of lights traveled a distance of 1.5 m, and there are 21 cycles of the sound waveform from a 5 kHz tone that are quite visible. Thus each cycle of the 5 kHz tone travels 150 cm/21 = 7 cm in one cycle, i.e., each cycle of the waveform indicates a 7 cm wavelength. From this, we can calculate the speed of sound = (0.07 m/cycle) × (5000 cycles/second) = 350 m/s.

Since the temperature was 27 °C, we know from theory that the speed of sound is 347 m/s, and the resulting experimental error is approximately 0.86%. This multimediated reality system could be useful for teaching physics to children by making physics, in this case, wave propagation, directly visible.

Multimedia display technologies, such as video and special eyeglasses, can be used to sample and hold the data captured by a moving sensor. A multimedia darkroom setup of this type is shown in Figs. 1.19 and 1.20.

Figure 1.19 demonstrates a multimediated reality darkroom with a 3D mechanical position control device that scans a space with transducers connected to a special lock-in amplifier (visible in the lower-left portion of Fig. 1.19b). The signal generator drives the transmit transducer here at 40 kHz (signal generator waveform appears on the oscilloscope that's sitting on top of the amplifier). Figure 1.19a: with the lights on (HDR photograph); Fig. 1.19b: with the lights off, the sensory attenuation experience helps us concentrate

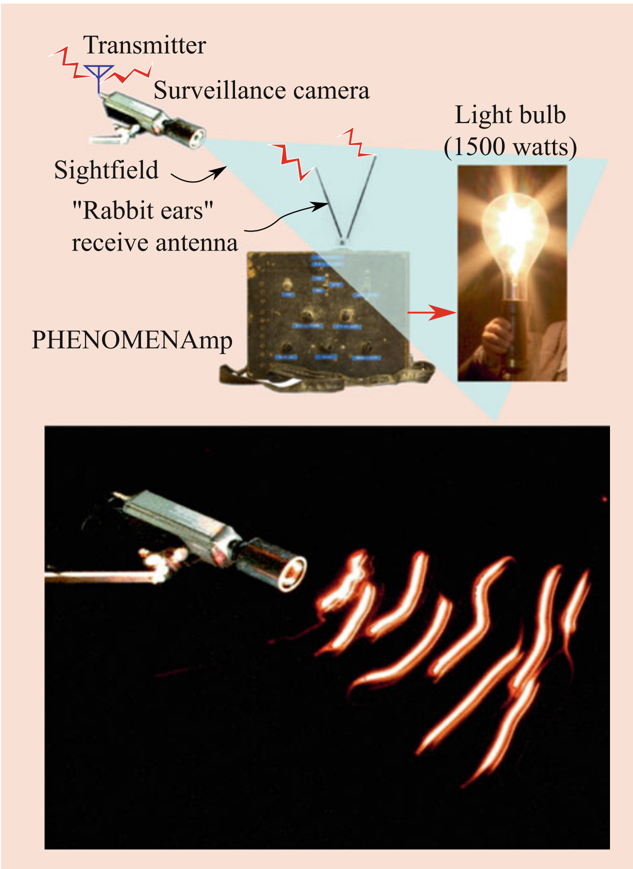


Fig. 1.15 Visualizing the "sight field" of a camera. Waving a light bulb back and forth "paints" out an image that reveals the otherwise hidden "sight field" of the camera

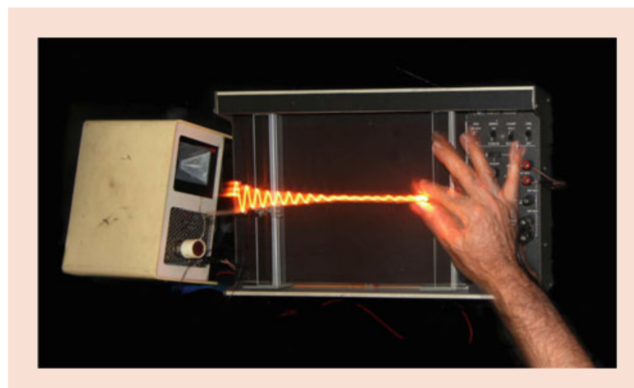


Fig. 1.16 A visual and haptic multimediated reality experience: the electrical signal from a microwave motion sensor (radar) is received by an antenna that moves together with a light bulb on an XY plotter. The bulb traces out the waveform of the received signal in exact alignment with physical reality

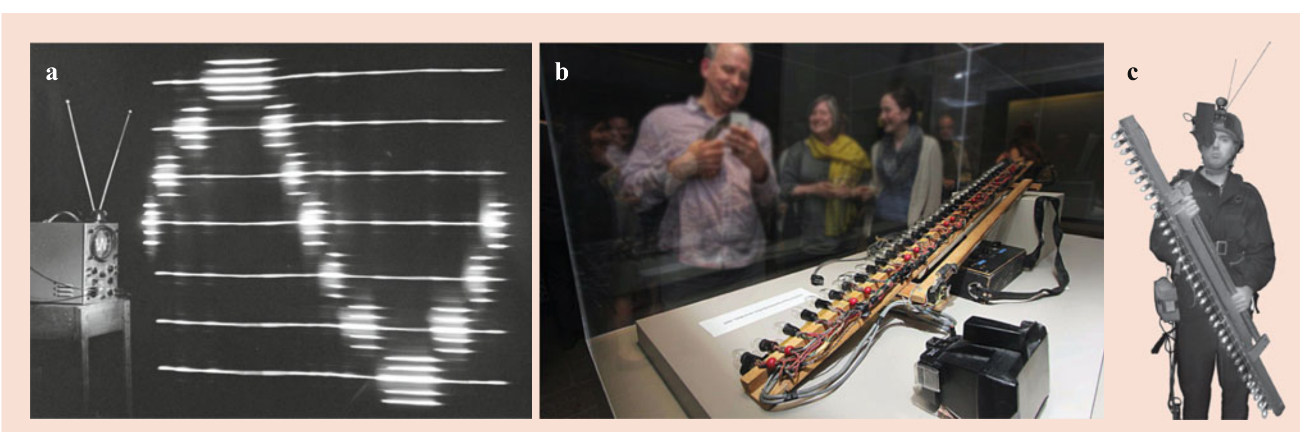


Fig. 1.17 (**a**) Early photograph of electromagnetic radio waves using the Sequential Wave Imprinting Machine (S. Mann, July 6, 1974, from Wikimedia Commons). (**b**) Sequential Wave Imprinting Machine exhibited at Smithsonian Institute and National Gallery, showing the linear array of 35 electric lights and the early wearable computer with a lock-in amplifier and antenna sensor. (**c**) Early 1980 version of the apparatus with early wearable multimedia computer prototype

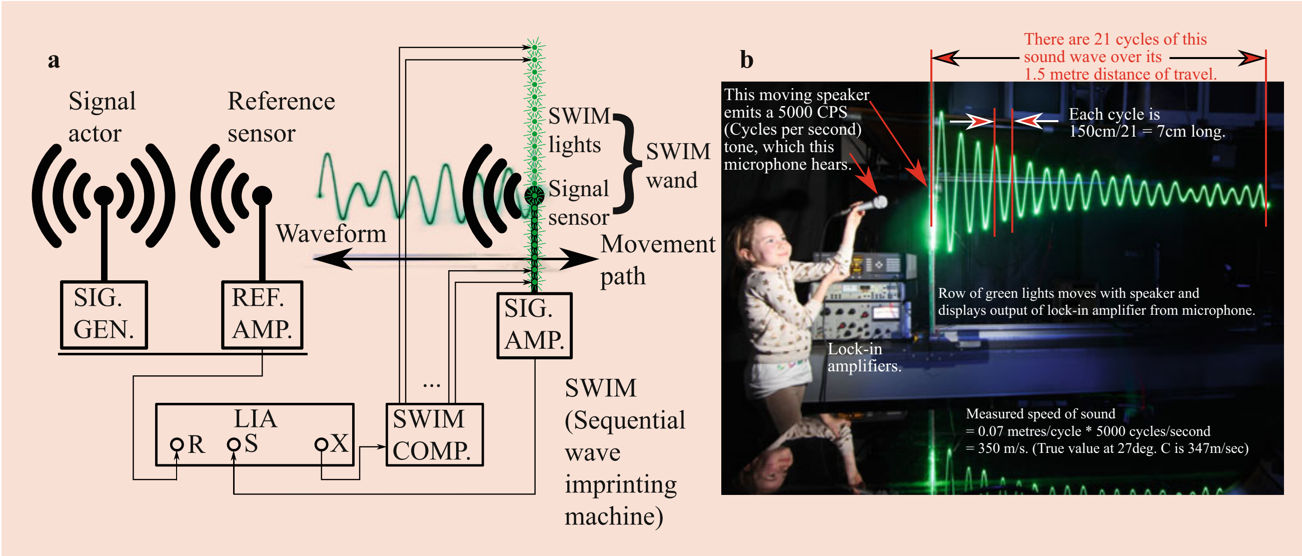


Fig. 1.18 (**a**) SWIM (Sequential Wave Imprinting Machine) principle of operation: a lock-in amplifier (LIA) compares a reference sensor with another sensor that moves together with a linear array of electric light sources. A long-exposure photograph captures the electric wave (radio wave or sound wave). (**b**) Photograph of a sound wave, from which we can measure the speed of sound

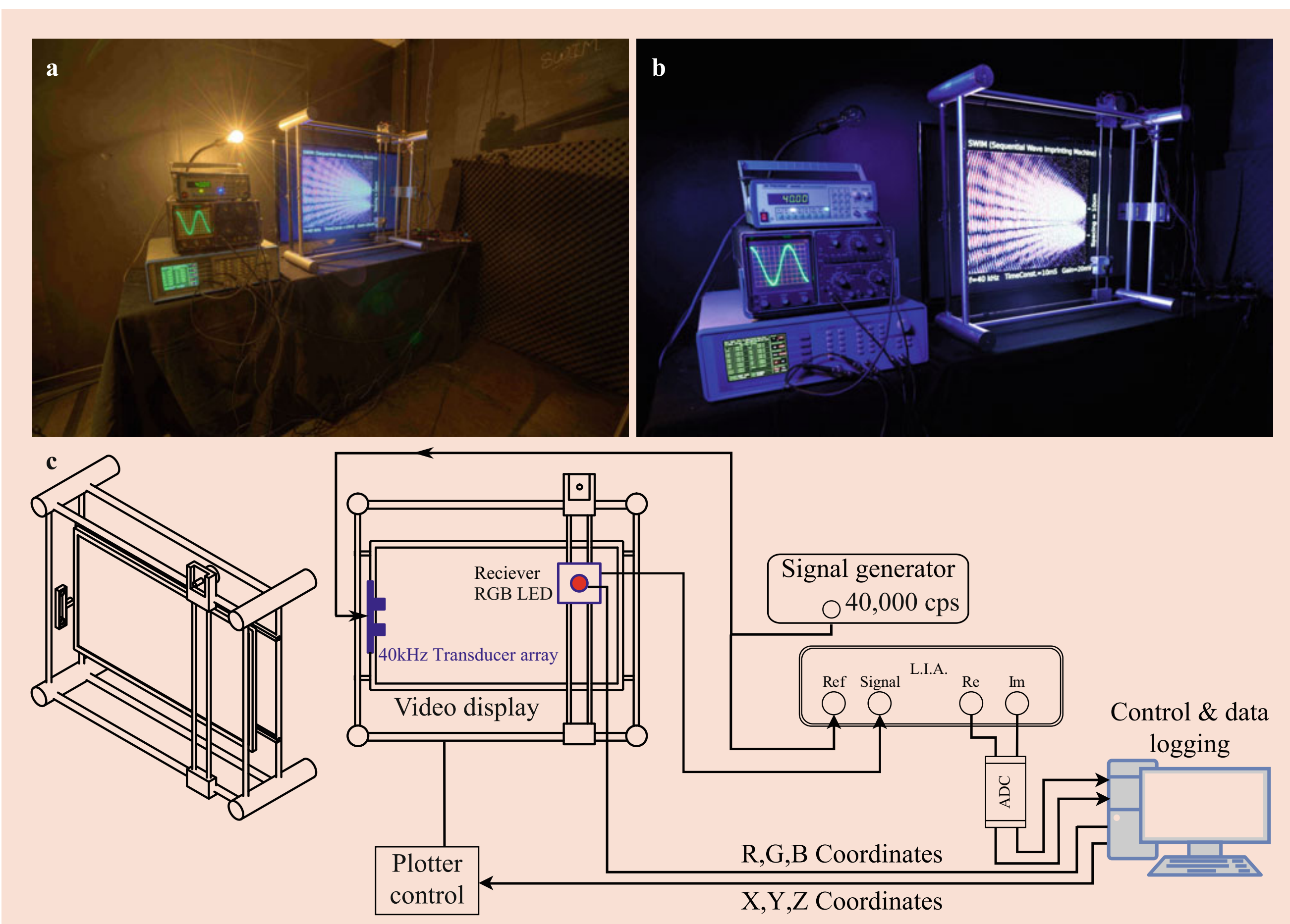


Fig. 1.19 Multimediated reality darkroom with a 3D mechanical position control device that scans a space with transducers connected to a special lock-in amplifier (**a**) darkroom setup with the room lights turned on so we can see the setup; (**b**) darkroom setup with the room lights turned off, as per actual operation; (**c**) schematic / diagram showing overall system setup

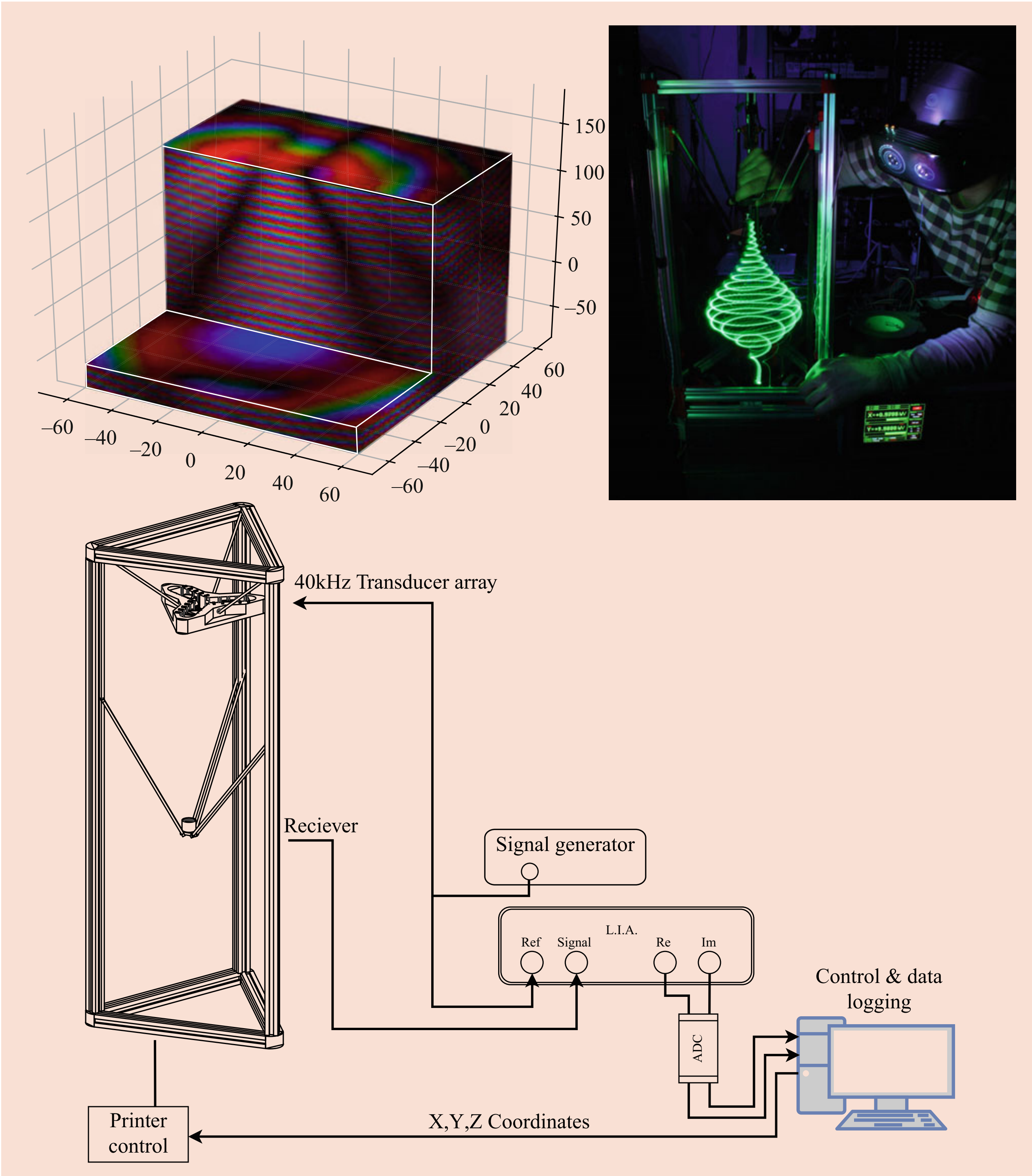


Fig. 1.20 With our multimediated reality headset, we see the multidimensional interference pattern embedded in 3D space. Here the interference pattern between three transducers (a triangular array) is shown, and we can also interact with the waveforms by way of patterns traced by the moving print head

on the multimediated reality that shows interference patterns between two sound sources. Figure 1.19c: Experimental apparatus for multimediated reality. An XY(Z) plotter carries a listening device (transducer) together with an RGB (red, green, blue) LED (light-emitting diode) through all possible positions in space. At each position, the sound is sensed phase-coherently by way of an LIA, while the sound is produced by a transmit array comprised of two transmitters, receiving the same signal to which the LIA is referenced. The outputs *Re* (Real) and *Im* (Imaginary) of the LIA are converted to RGB values for display on the LED. A picture is taken of this movement and presented to a video display, to provide persistence of exposure. Alternatively, the video display may be driven directly by data stored in the control and data logging system. In this case, it can be animated by multiplication by a complex number of unit modulus, so that the waves on the screen slowly "crawl" at any desired speed of sound (e.g., the speed of sound can be set to zero or to some small value to be able to see it clearly).

In this way, a volumetric dataset is retained as the 3D waveform interference pattern is scanned through space. While prior SWIM (Sequential Wave Imprinting Machine) techniques (e.g., long-exposure photography) can capture 2D and sometimes 3D waveforms, they are recorded from a single two-dimensional perspective. By reading and storing radio wave or sound wave information in point-cloud form, the waves can be reconstructed, manipulated, and analyzed a posteriori in three dimensions, allowing for infinite angles and slicing. This method gives a new level of richness when observing waves, and this richness can be explored using multimediated reality eyeglasses.

1.3.4 Comparison with Existing Measuring Instruments

Many measuring instruments allow us to use one of our senses to measure a quantity that pertains to another sense. For example, a photographic darkroom light meter that emits sound is used in total darkness. An oscilloscope allows us to see sound waves. A multimeter is a device that allows us to see or hear electrical signals that would otherwise be difficult or impractical to sense directly. What is unique about multimediated reality compared with these traditional measurement instruments is that multimediated reality provides a direct physical alignment between the measured quantities and the real world from which they come.

Multimediated reality allows sound waves or radio waves to be seen in perfect alignment with the world in which they exist, that is, at 1:1 scale, and "frozen" in time (i.e., visible in a set of coordinates in which the speed of sound or speed of light is zero, or a small number that allows the wave propagation to be studied easily).

1.4 Multimediated Reality Is Multiscale, Multimodal, Multisensory, Multiveillant, and Multidimensional

Multimediated reality is more than just a taxonomy of real and synthetic experience. It also considers how we

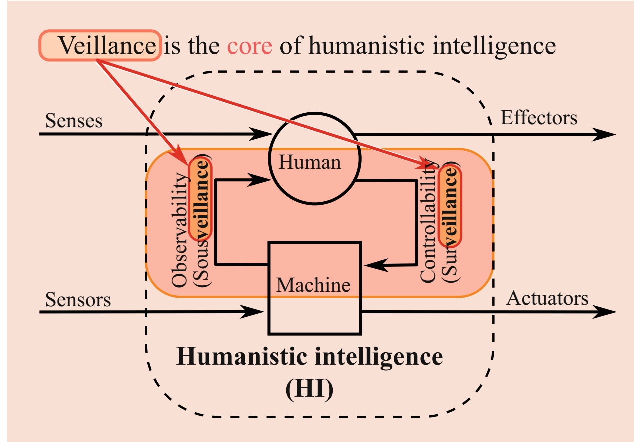


Fig. 1.21 Humanistic intelligence (HI) [16]

interact with the world around us and each other through the use of technology as a true extension of our minds and bodies. Specifically, we consider the concept of AI (artificial intelligence) and human-in-the-loop AI, also known as HI (humanistic intelligence) [41]. HI posits that technology should function as an intermediary between us and our environment in such a way that the intelligence it affords us arises through a computational feedback loop of which we are apart. Thus, HI is an important aspect of multimediated reality. See Figs. 1.21, 1.22, and 1.23. HI addresses systems in which our human senses (sight, hearing, etc.) and effectors (e.g., hands) are augmented or mediated by machine intelligence having sensors and actuators. HI is intelligence that arises by having the human in the feedback loop of a computational process. It requires sousveillance (undersight) whenever surveillance (oversight) is present, i.e., it requires a fully closed loop such that if a machine can sense us, we must also be able to sense the machine [41]. This reciprocity is the core feature of HI that enables it to form a fundamental basis for multimediated reality. Thus multimediated reality is multiveillant (in contrast to monoveillant technologies that include only surveillance without sousveillance).

We have introduced the concept of "vironment" as the union between ourselves and the "environment" (that which surrounds us). The concept of vironment is illustrated in Fig. 1.24. Consider a 2-m-diameter social distancing sphere. The "environment" is that which surrounds us, i.e., outside the sphere. The "invironment" is the orthogonal complement of the environment, i.e., the invironment is us ourselves. We usually consider our clothes to be part of us, i.e., the invironment is us plus the technology that we consider to be part of us. The "vironment" is the environment and the invironment together. Vironmentalism concerns itself with both, as well as the boundary between them, for example, wearables.

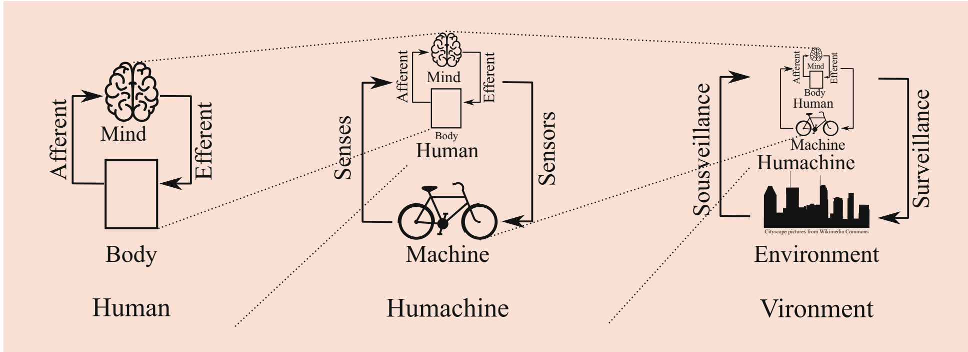


Fig. 1.22 The self-similar (Fractal) nature of HI: the body may be thought of as a machine. A human may be thought of as comprised of a mind-body feedback loop formed by efferent nerves that carry signals from the brain to the body and afferent nerves that carry signals from the body to the brain. The human may interact with (other) machines, such as a bicycle or eyeglasses, etc., to form a "cyborg" entity denoted as "humachine". The humachine interacts with its environment, such as an urban environment (e.g., a "smart city") through surveillance (surveillant signals) analogous to efferent nerves. Smart cities should also embrace sousveillance (sousveillant signals) analogous to afferent nerves [38]

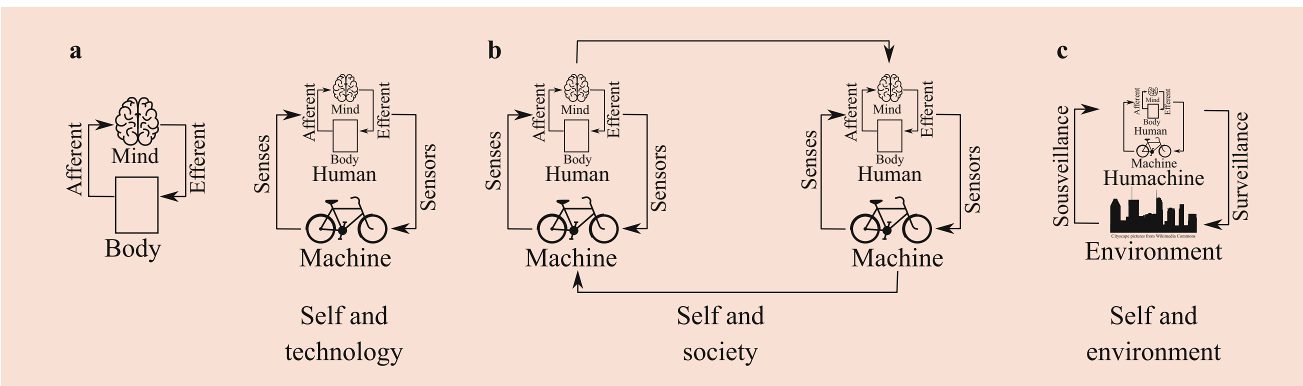


Fig. 1.23 Humanistic intelligence operates at three levels: (**a**) self and technology (e.g., "cyborgs" or "humachines"); (**b**) self and society (e.g., interaction between multiple cyborgs and non-cyborgs, etc.); and (**c**) self and environment (typically multiple cyborgs and non-cyborgs in a smart building or smart city or smart world) [38]

Vironmentalism gives rise to a new form of artistic intervention, performance art, design, etc., which plays on the idea of separation between the environment and the environment. See Fig. 1.25.

Multimediated reality involves multiple physical scales, including both wearable technology and technology in the environment around us, such as smart rooms, smart buildings, smart cities, smart worlds, etc. This multiscale and multiveillant nature of multimediated reality is illustrated in Fig. 1.26.

Multimediated reality is multiveillant (surveillance *AND* sousveillance) and multiscale (wearables *AND* smart environments). We can identify at least three axes: Firstly, a physical scale axis (of physical reality) defines the environment (that which surrounds us) and the invironment (us ourselves). At the border between the environment and invironment are things like clothes and other technological prostheses. A virtuality axis also defines a scale from "bits" all the way out to "big data". A sociopolitical or "veillance" axis defines sousveillance (individual/internal) out to surveillance (external). At the origin are "atoms," "bits," and "genes." Genes, for example, are the smallest unit of "humanness" (human expression). The axes are labeled α, β, and γ. The first of the three axes (α) is denoted pictorially, at physical scales starting from naked, to underwear, to outerwear, to a vehicle (car), to the "smart city." Let us not confuse sousveillance with counterveillance or with transparency. A person can, for example, be in favor of both veillance (sur and sous) or of just one or just the other. Thus they are distinct and not necessarily in opposition. Likewise with transparency. We might wish for a transparent society but must be careful not to wish for one-sided transparency that leaves certain officials unaccountable. We proffer to name the call for such false or illusory transparency "clearwashing," akin to the false environmentalism called "greenwashing."

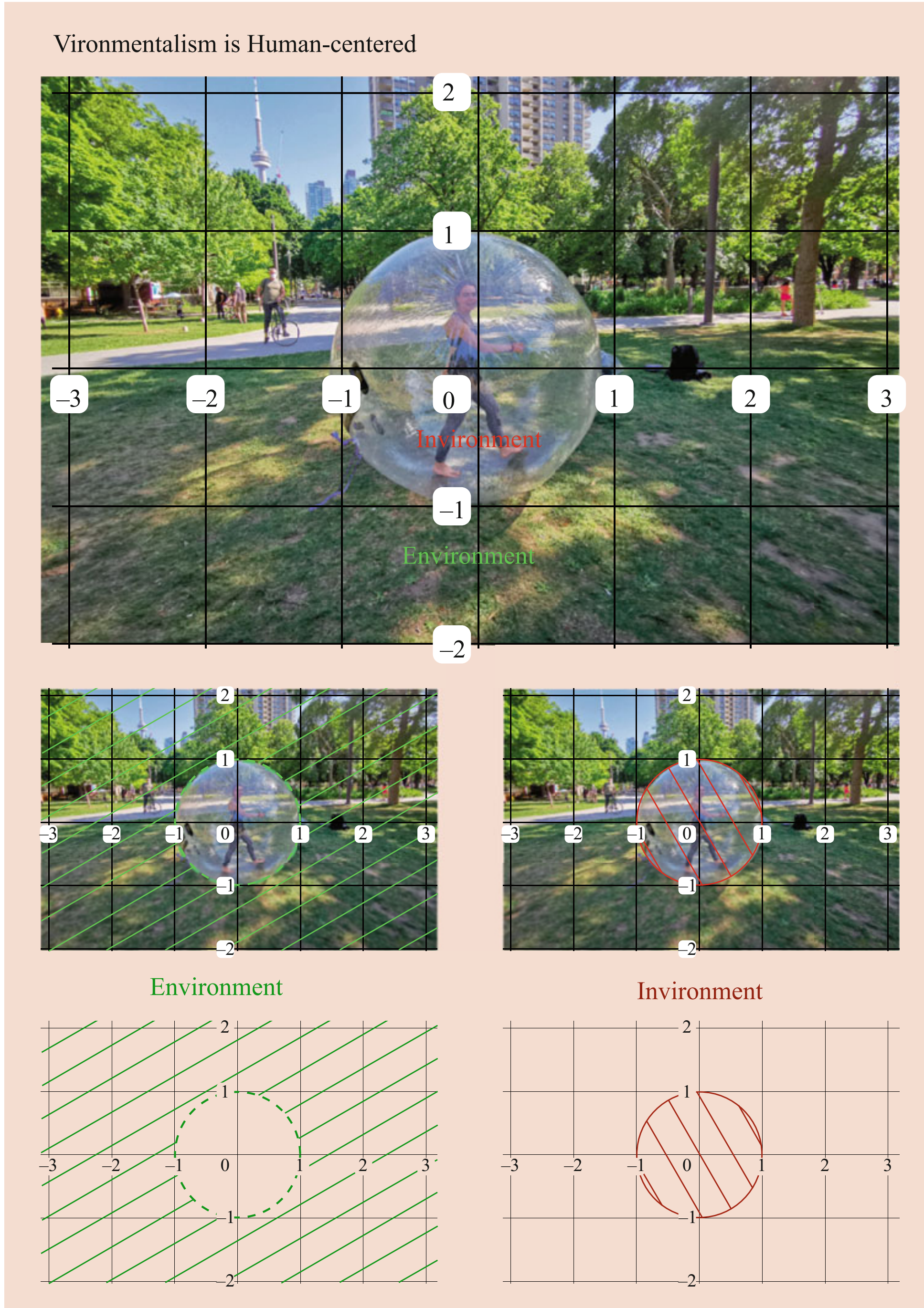


Fig. 1.24 A 2-m-diameter social distancing sphere is used to describe the “environment” and “invironment”

1.4.1 Multisensory Synthetic Synesthesia

Synesthesia is a neurological condition in which there is crosstalk between human senses, for example, chromesthesia, where sounds cause one to perceive colors. Another example of synesthesia is “The Man Who Tasted Shapes” [6].

Multimediated reality often involves a multimedia-induced (synthetic) synesthesia among and across our existing senses (e.g., seeing sound) or extrasensory, i.e., beyond our existing senses (e.g., seeing or feeling radio waves). In this way, multimediated reality is multisensory and multimodal.

1.4.2 Multidimensional Multimediated Reality

Synthetic synesthesia of extra senses like electromagnetic radio waves and radar (or sonar) provides us with a richly multidimensional perceptual space in the context of multimediated

Fig. 1.25 Vironmentalism as a form of visual art or performance art (**a**) and (**b**) real reality, HEADome, and long-spike necklace. (**c**) Augmented reality with 12 sonar units arranged in a circle to capture a full panoramic range map displayed as a plan position indicator on the EyeTap

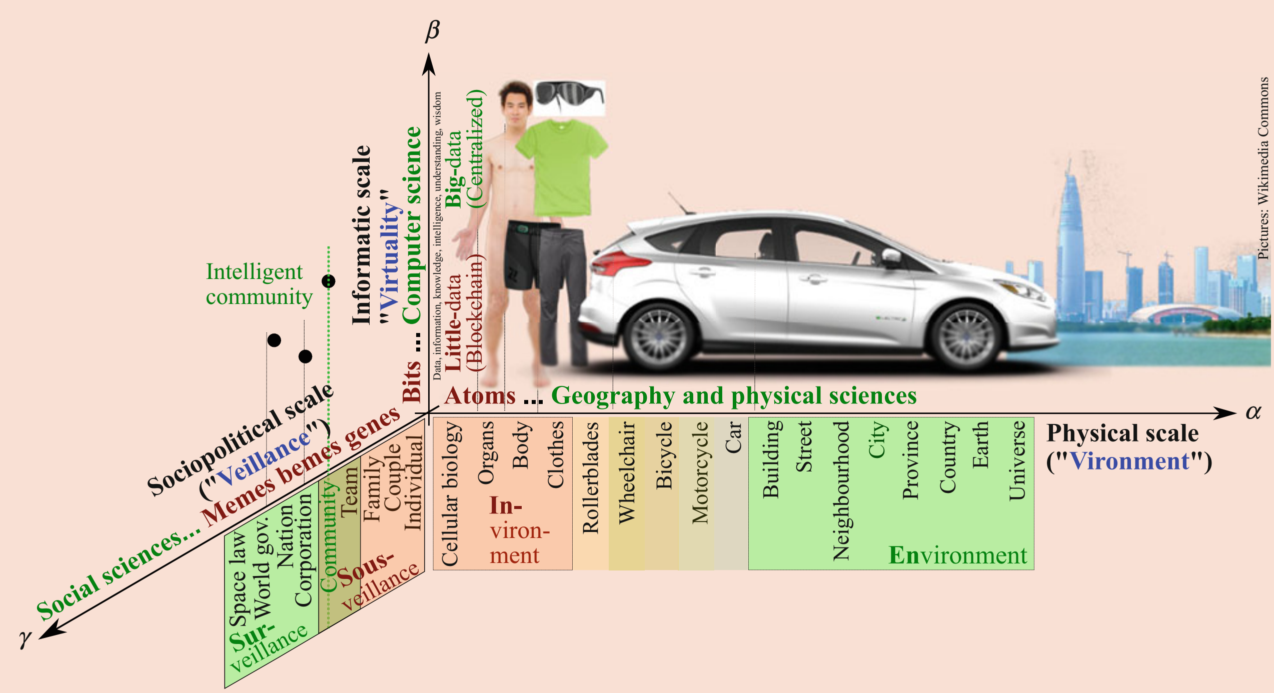


Fig. 1.26 Multimediated reality is multiveillant (surveillance *AND* sousveillance) as well as multiscale (wearables *AND* smart environments). We can identify at least three axes: physical scale, sociopolitical scale, and informatic scale

reality. Whereas existing virtual reality systems might use radio waves or sound waves for tracking and position sensing, they do not directly engage the waveforms phase coherently as part of the resulting output space. When we engage with the radio waveforms or sound waveforms directly, we have many new sensory possibilities for direct experience of multidimensional signal properties like multi-polarization and multipath propagation.

Even simple radio waves become complex valued when brought down to baseband (e.g., when experienced in coordinates where the speed of sound or light is zero). In this case, the chosen synesthesia is to be able to see complex numbers on a color wheel where phase is color and magnitude is the quantity (photoquantity [28]) of light presented [24].

We see this in Fig. 1.27, where the phase of the sound waves is displayed in color. Thus we can clearly see the

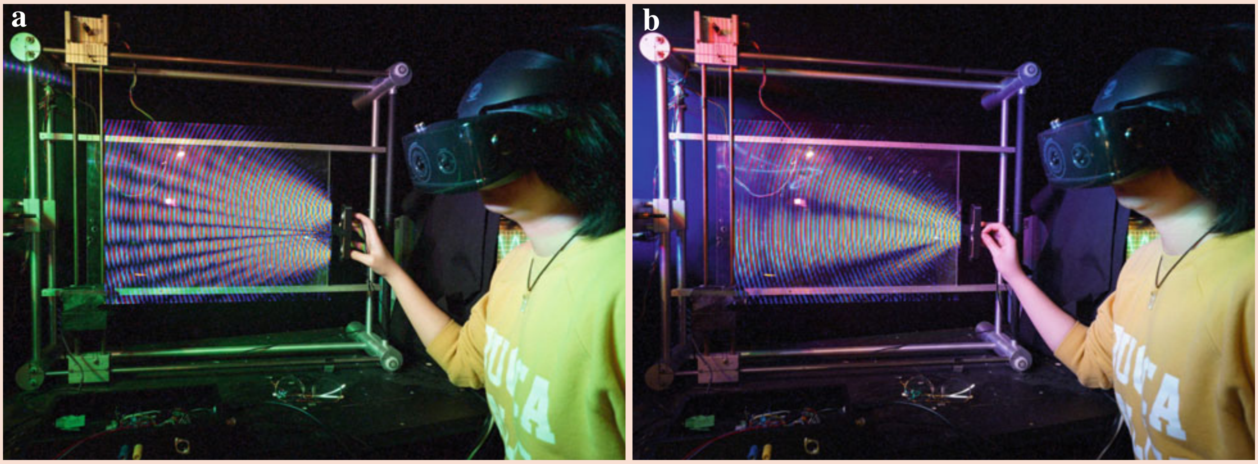


Fig. 1.27 Interference pattern of two acoustic wave fronts made visible using the multimediated reality darkroom of Fig. 1.19. (**a**) Photograph showing the apparatus when the two transducers are far apart. (**b**) Photograph showing the apparatus when the two transducers are moved closer together

interference patterns of two sound waves, where these waves interfere constructively and destructively, as variations in the quantity of light, and the phase fronts, as variations in color. Interference pattern of two acoustic wave fronts is made visible using the multimediated reality darkroom of Fig. 1.19. As the two sound sources (speakers) move closer together, their acoustic interference fringes spread further apart, confirming a simple law of physics. Left photograph: wide spacing between speakers gives a closely spaced interference pattern. Right photograph: narrow spacing between speakers results in a broad-spaced interference pattern. Sound waves from two movable speakers emitting the same pure tone are photographed with a lock-in amplifier driving an RGB light source moving in space (X, Y, Z). These results are not computer graphics! These are photographs of an actual physical process generated by nature itself and are merely facilitated (made visible) by computation. Those wearing a multimediated reality headset see patterns in a high-dimensional space, whereas those without headsets can still see the sound waves but at a reduced dimensionality.

This system may also be used to study underwater sound waves. Figure 1.28 shows a sonar array being explored. The apparatus shows underwater sound at 40 kHz from two hydrophones, toward the left (the spacing is controllable with the user's right hand), and a test/reference hydrophone toward the right (near the user's left hand). The wearer of the multimediated reality eyeglass sees in a high-dimensional space (3D spatially, plus time animation, plus complex-valued quantities, etc.). Others see only a flat 2D slice through the interference pattern. We have also constructed a multimediated reality (MR) aquatics facility for use with underwater MR eyewear. In water, the weightless state allows us to feel as if we are floating among the underwater sound waves.

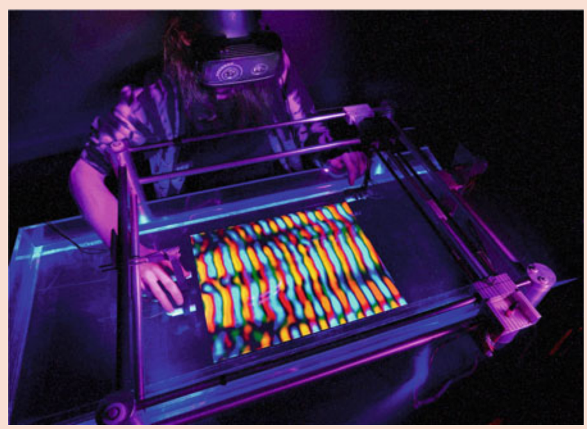


Fig. 1.28 Photograph of apparatus showing underwater sound at 40 kHz from two hydrophones

1.5 Multimediated Reality Continuum

Many of the systems presented in this chapter do not fit nicely into existing taxonomies of VR, AR, or any of the more general taxonomies of synthetic experience [46]. We proffer a more general "reality" continuum in which the space is multidimensional, in which the origin is the absence of

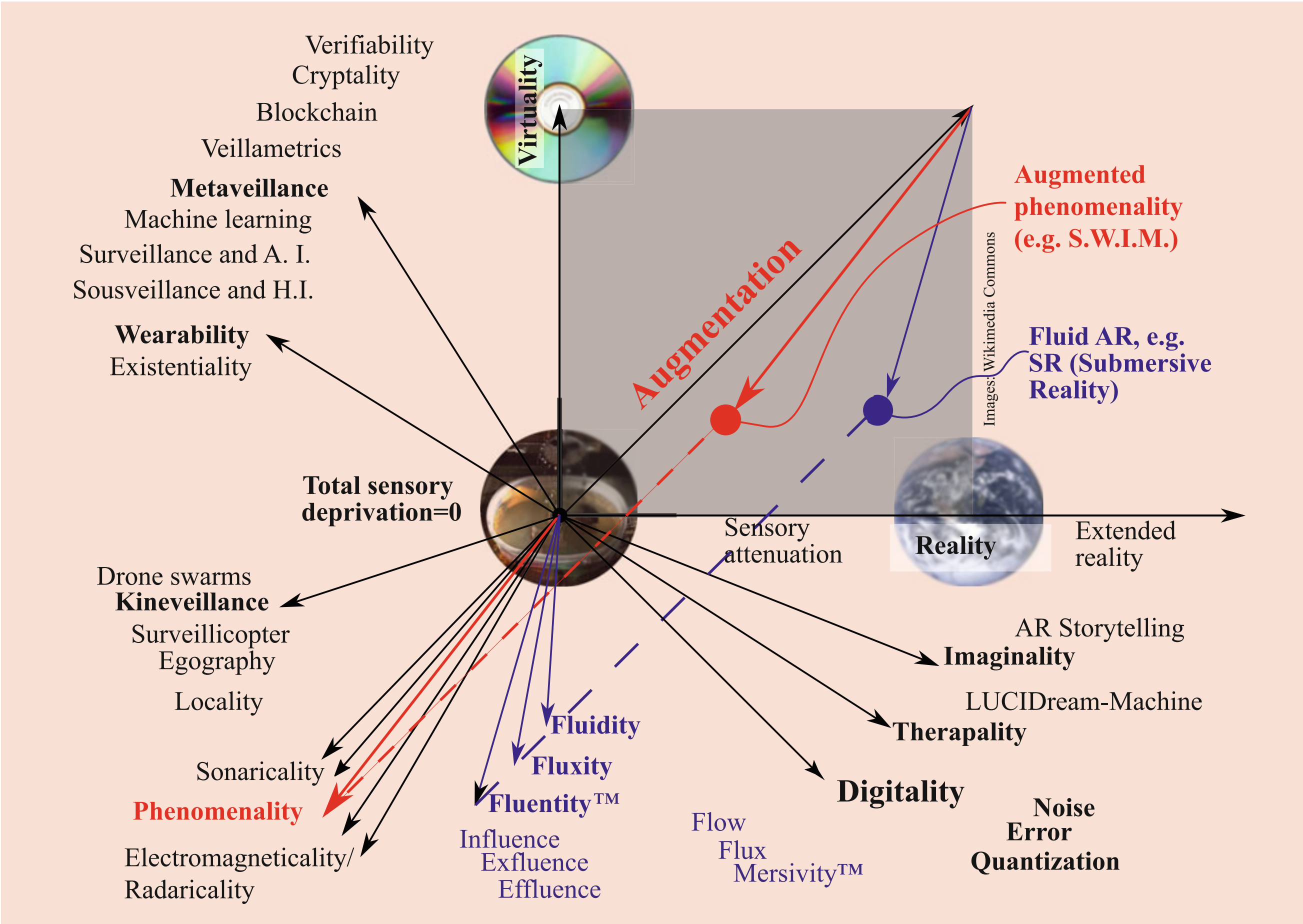


Fig. 1.29 The multimediated reality continuum: a more general "reality" continuum

sensory stimulation, allowing us to consider technologies such as sleep masks, interactive sleep masks, sensory deprivation tanks, interactive sensory deprivation tanks [20, 29], aquatics facilities, theatres, darkrooms, therapy systems, and the like, as a basis upon which to create new realities directly connected to physical or intellectual phenomena. Figure 1.29, the multimediated reality continuum, illustrates the concept of a general "reality" continuum. Reality is the main axis going from left to right, starting at "total sensory deprivation" (the origin, indicated by a sensory deprivation tank), then to "sensory attenuation," then reality, and then beyond reality to give also extended reality. Virtuality is the secondary axis pointing upwards. Augmentation exists in the two-dimensional space spanned by reality and virtuality. A third axis, phenomenality, indicates any kind of phenomenological reality, such as phase-coherent photography of radio waves or sound waves such as by Sequential Wave Imprinting Machine (SWIM). In this sense, PAR (phenomenological augmented reality) [22] is a combination of AR and P (phenomenality). A point in this space is indicated by the red dot as "augmented phenomenality." As another example, consider a point (indicated in blue) that comes out from AR along the fluentity axis. An example of this kind of reality is the Internet-based underwater virtual/augmented reality performance space [20, 29]. When we submerge ourselves in a large swimming pool, with an underwater SWIM to see sound waves (e.g., to test out an underwater sound system and see the interference patterns between two underwater speakers), we're in a reality described by adding the two vectors (red dot and blue dot) taking us into an additional higher dimension. This works like a giant version of our apparatus of Fig. 1.28, in which we submerge ourselves fully in the pool while wearing the all reality headset. The all reality continuum thus allows us to understand sensory attenuation (interpolation between sensory deprivation and reality) as well as eXtended reality (extrapolation beyond reality), in addition to the many other dimensions shown, such as metaveillance [23] (sur/sousveillance, smart cities, smart buildings, etc.), wearability (Mann's "Wearable Reality" of 1974 and Canadian Patent 2388766), kineveillance (drone and swarm-based realities [32]), imaginality (e.g., lucid dreaming in the sensory deprivation tank), and therapality (the axis of lasting effects that persist even after shutting off and removing a technology). Not all axes are desirable, e.g., digitality is the axis of quantization noise that embodies the undesirable artifacts of *being digital*, pixelation, etc. Ideally, we wish to use computers for *being undigital* [26]. Note the many dimensions and the many ways they can be combined. For example, we can have a mix of reality and virtuality that gives AR (augmented reality) and then further add some phenomenality to get PAR (phenomenological augmented reality [22]). We can add to AR some fluentity to get SR (submersive reality [20, 29, 33]). And if we use PAR while swimming fully submerged in water, we're spanning the four dimensions of reality, virtuality, phenomenality, and fluidity/fluentity.

Note also that many of the dimensions are inherently combined and thus certainly do not form an orthonormal basis. For example, metaveillance includes surveillance (oversight) and AI (artificial intelligence), as well as sousveillance (undersight) and HI (humanistic intelligence). Sousveillance and HI are very closely related to wearability. Consequently, there is a strong overlap between metaveillance and wearability. Wearability is closely correlated to existentiality [17], giving us the "wearable reality" (Canadian Pat. 2388766) proposed by Mann in 1974 as embodied by the SWIM (Sequential Wave Imprinting Machine), itself an example of phenomenality [22]. Thus these three axes, metaveillance, wearability, and phenomenality, are closely related.

Not all dimensions are desirable. For example, sensing can be done in a discrete (quantized) or continuous manner. We prefer systems in which computers are used to sense "undigitally," i.e., above the Nyquist rate spatiotemporally. Today's digital cameras do this, for the most part, but still exhibit tonal/level quantization, thus the need for *being undigital* [26] with digital cameras, e.g., HDR sensing and metasensing [22, 48]. Ideally, we would like to use computers for *undigital senses/sensors* as well as for *undigital effectors/actuators*, i.e., for all six signal flow paths of Fig. 1.21. This allows us to achieve undigital experiences, i.e., what we proffer to call a *undigital reality*, which is a multimediated reality that is free from the artifacts of digital computation. This is especially important for SR (submersive reality), in which the free-flowing nature of swimming makes it possible to forget that one is wearing a VR/AR/MR headset and imagine, in the totally weightless world, another reality that is totally fluid and free of *digitalness* (quantization noise, pixelation, etc.), a world where digital senses/sensors, or the like would be totally unacceptable.

1.5.1 Multimediated Reality is "★R" (All R)

In most search engines and computing environments, the asterisk symbol, "∗," is a "wildcard" that can mean anything. It can be replaced with any other characters, symbols, words, or phrases and dates back as far as the TOPS-10 operating system in 1967. Thus for the space defined by the multimediated reality continuum, we proffer to call it "'∗R," pronounced "All R" (all realities). In situations where there is only room for two characters, and where an asterisk cannot be used conveniently (e.g., file names in the Unix operating system), we proffer "ZR," to emphasize the possibility of complex-valued multidimensional data of the form $Z = X + iY$ where $i = \sqrt{-1}$ is the imaginary unit. We must realize, though, that there are many dimensions, and some of these many dimensions may be complex valued or beyond, i.e., not limited to the two dimensions implied by the Argand plane (https://en.wikipedia.org/wiki/Complex_plane).

1.6 Other Forms of Phenomenological Augmented Reality

Phenomenological augmented reality (PAR) arises when the user experiences the mixing of augmented reality (AR) with phenomenality or phenomenal reality. Phenomenal reality is the revelation of hidden phenomena present around us, where the inner workings of the laws of nature are revealed. PAR is the mechanism that allows us to experience phenomenal reality. Virtual reality, and in some cases augmented reality, is analogous to the fiction section in a library, while PAR is akin to the nonfiction section. While PAR was mentioned in earlier sections, this section is meant to cohesively describe and further explore this type of multimediated reality.

A PAR system can be implemented by using the Sequential Wave Imprinting Machine (SWIM), which consists of a sensor, an actuator whose invisible phenomena we wish to make visible, a lock-in amplifier for phase-coherent detection, and a light source (e.g., an array of lights or a monitor/display) to overlay the otherwise invisible phenomena onto reality (the principle of operation is described in Fig. 1.18). The sensed quantity is displayed as a function of space (and not time), also known as a sitting wave.

The SWIM was first invented in the 1970s and makes visible the otherwise invisible real-world phenomena [22]. Some early examples of the SWIM and by extension PAR were briefly described in this chapter (e.g., Figs. 1.15, 1.16, 1.17, 1.18, 1.19, 1.20, 1.27, and 1.28), where radio and sound waves are made visible to the naked eye, in real time and phase coherently.

The history of the SWIM and examples of the SWIM in various applications, such as visualizing radio waves using software-defined radios and electromagnetic waves within electric machines, are presented and discussed in the following sections.

1.6.1 History of the SWIM

Since the discovery of the original SWIM in 1974, the SWIM has been used as a way of researching, understanding, visualizing, and teaching fundamentals of scientific phenomena. The classical SWIM was based on a simple and intuitive phenomenon. A light source, which was Voltage controlled, consisted of either a linear array lights of lights or a cathode-ray tube moved back and forth and gives rise to a Doppler shift on propagating waves.

The SWIM phenomena was discovered by moving an RCA TMV-122 cathode-ray oscillograph that had a broken time base oscillator (basically it had no time base) back and forth in front of a police radar (Kustom TR-6) that was fixed at a point. The output of the radar, which was at baseband, was connected to the vertical deflection plates of the cathode-ray oscillograph. This caused the dot on the oscillograph to move vertically, up and down on the screen, while the entire body of the oscillograph was moved along a horizontal axis (i.e., left and right).

The oscillograph's metal body caused Doppler radar returns which resulted in the dot on the screen following a fixed path (i.e., following the waveform as a function of space), regardless of the speed of motion of the oscillograph. Moving the oscillograph faster resulted in a greater Doppler shift, so the dot moved up and down faster to the same extent it moved left to right, thus canceling the propagatory effects of the wave. This resulted in a phenomenological augmented reality overlay "due to a sheared spacetime continuum with time-axis at slope $1/c$" [23]. In other words, the resulting effect is as if we're moving along at the speed of wave propagation c, causing the wave to essentially "sit" still with respect to our moving reference frame [23], which is known as a sitting wave. The oscillograph was later replaced by a linear array of light sources (i.e., light bulbs and then LEDs) that sequentially illuminate as a function of space, in response to an input voltage.

1.6.2 SWIM Principle of Operation

A diagram illustrating the Sequential Wave Imprinting Machine (SWIM) principle of operation is shown in Fig. 1.18. The SWIM consists of a lock-in amplifier (LIA), display, sensor, transducer, as well as a reference receiver and a signal receiver that is required to be phase coherent with each other. This section will explain the concept of "sitting waves" and the importance of phase coherence for the SWIM and how these properties allow for visualization of phenomenology such as radio waves (e.g., from a radio) or sound waves (e.g., from a microphone).

The lock-in amplifier (LIA) is a phase-sensitive detector with three terminals: a signal terminal, reference terminal, and output terminal. The reference receiver, signal receiver, and display are connected to the reference, signal, and output terminals, respectively, on the LIA. A stationary signal source transmits a waveform to a stationary reference receiver and a signal receiver that moves together with the display. When both the reference and signal receiver are phase coherent with respect to each other, a waveform can be visualized on the display as a function of space and not as a function of time; this is a sitting wave. A vertical, linear array of lights (e.g., LEDs) is typically used as the display, where the waveform is traced out as a function of space on the LEDs as the display is moved back and forth. Note that signal source and receivers must be aligned about an axis, where the reference receiver is in the middle.

Figure 1.30 illustrates a comparison between a standing wave, crawling wave, and sitting wave. Standing waves arise due to the interference of two waves traveling in opposite directions at the same frequency and amplitude; it is a wave that oscillates in the vertical direction as a function of time but is fixed at the nodal points (Fig. 1.30a). Crawling waves are traveling waves, which move slowly in the horizontal direction as a function of time and arise due to a slight phase difference between two waves that are fed into an LIA's reference and input terminals (Fig. 1.30b). Unlike standing and crawling waves, sitting waves are completely fixed as a function of time and only change as a function of space;

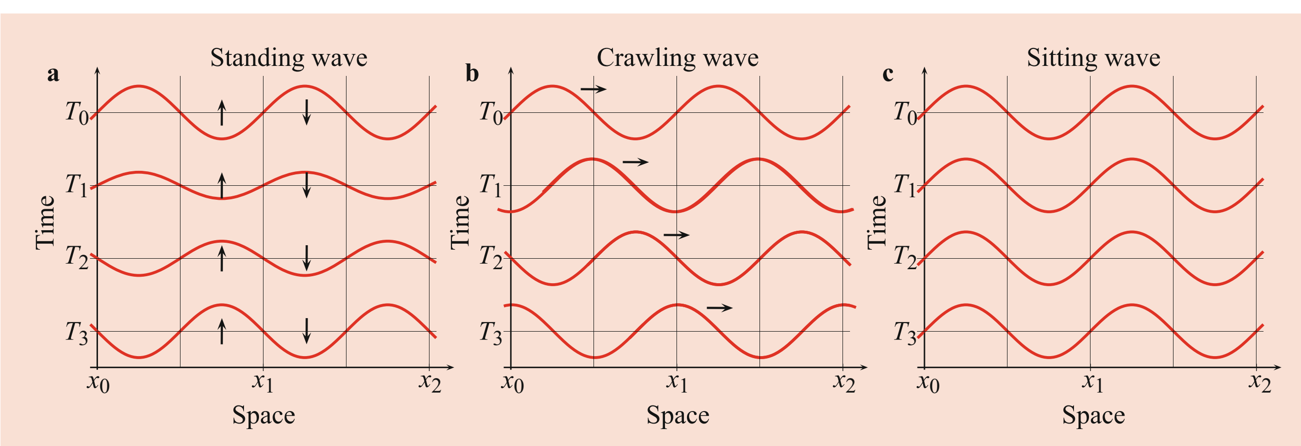


Fig. 1.30 Comparison between the following types of waves as a function of space and time. (**a**) Standing wave. (**b**) Crawling wave. (**c**) Sitting wave

this type of wave is a result of successful phase-coherent detection of the two waves fed into an LIA (Fig. 1.30c).

To gain a deeper understanding of sitting waves, we will analyze sinusoidal waves with amplitude A and angle θ together with the components of the SWIM, as explained in [23, 36]. Sinusoidal waves are described by their in-phase I (real) and quadrature Q (imaginary) components, where I is the product of the signal amplitude and the cosine of the phase angle (Eq. 1.4) and Q is the product of the signal amplitude and the sine of the phase angle (Eq. 1.5):

$$I = A \cdot \cos(\theta) \tag{1.4}$$

$$Q = A \cdot \sin(\theta) \tag{1.5}$$

With a stationary reference receiver, Eq. 1.4 can be expressed as a function of time t and angular frequency ω:

$$\mathrm{I}_{\mathrm{REF}} = A \cdot \cos(\omega t) \tag{1.6}$$

While ignoring the Doppler effect, with a moving signal receiver, Eq. 1.4 can be expressed as a function of space x and time t (where k is the wave number):

$$\mathrm{I}_{\mathrm{SIG}} = A \cdot \cos(\omega t - kx) \tag{1.7}$$

By having the stationary and moving receivers synchronized with respect to frequency and phase (i.e., phase coherent), when the two signals are mixed (i.e., multiplied together), this results in a sitting wave [23]. The signals from Eqs. 1.6 and 1.7 are inputs to the LIA. The LIA will mix these inputs together:

$$\mathrm{I}_{\mathrm{MIXED}} = \mathrm{I}_{\mathrm{REF}} \times \mathrm{I}_{\mathrm{SIG}} \tag{1.8}$$

Equation 1.8 can be expanded by applying the product of cosines identity as follows:

$$= \frac{1}{2}[\cos(2\omega t - kx) + \cos(kx)] \tag{1.9}$$

Finally, the LIA applies a low-pass filter to Eq. 1.9, eliminating the high-frequency term, resulting in a sitting wave (of the real component), where the wave is a function of space only:

$$= \frac{1}{2} \cdot \cos(kx) \tag{1.10}$$

The same steps described above can be followed to solve for the sitting wave of Eq. 1.5 (the imaginary component) using the product of sines identity.

The resulting output from the LIA (Eq. 1.10) is fed into the display. Because the signal receiver is attached to the display and both move together, the display traces out the waveform as a function of space, allowing the naked eye to observe the wave as it occurs naturally or for long-exposure photography to capture the phenomena.

1.6.3 Visualizing Radio Waves with SWIM and Software-Defined Radio

Using the SWIM together with software-defined radio (SDR) results in a novel method of visualizing the actual radio waves being received or transmitted by the radio [36]. This method improved on the early example shown in Figs. 1.17 and 1.18, where large and bulky hardware is required. SDRs are software configurable radios, where nearly all of the physical layer functions typically found in traditional radios are software defined. Thus, SDRs are compact in size, inexpensive, and easily integrate with general-purpose computers through Universal Serial Bus (USB). Digital signal processing, such as modulation and demodulation of FM radio, are easily handled by a general-purpose computer when paired with an SDR. Additionally, radio parameters such as gain, sampling rate, center frequency, etc., are all easily adjustable in software.

Figure 1.31 illustrates the SDR SWIM setup; SDRs are used for the receivers, the lock-in amplifier (LIA) is replaced by a general purpose computer, and the linear array of LEDs is replaced with a computer monitor.

The waveform shown in Fig. 1.32a is the in-phase component of a pure carrier wave at 915 MHz being received by the SDR and displayed by the SWIM. This waveform has been empirically verified to have a wavelength of approximately 30 cm. On the other hand, the ideal wavelength of this waveform is $\lambda = c/v$. Solving for this equation yields $\lambda = (3 \times 10^8\ \mathrm{m/s})/(915 \times 10^6\ \mathrm{Hz})$, resulting in a wavelength

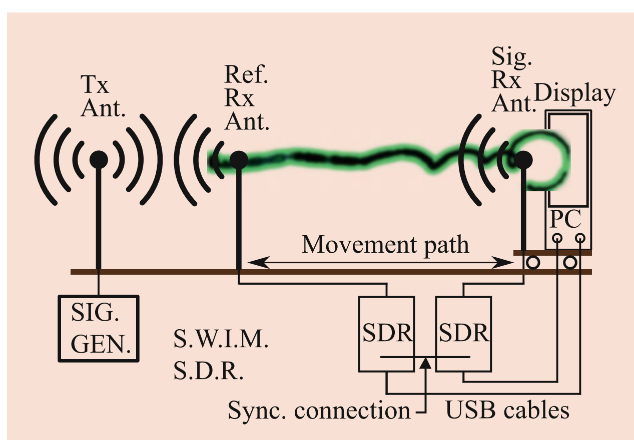


Fig. 1.31 The block diagram for the software-defined radio (SDR) Sequential Wave Imprinting Machine (SWIM) variant [36]

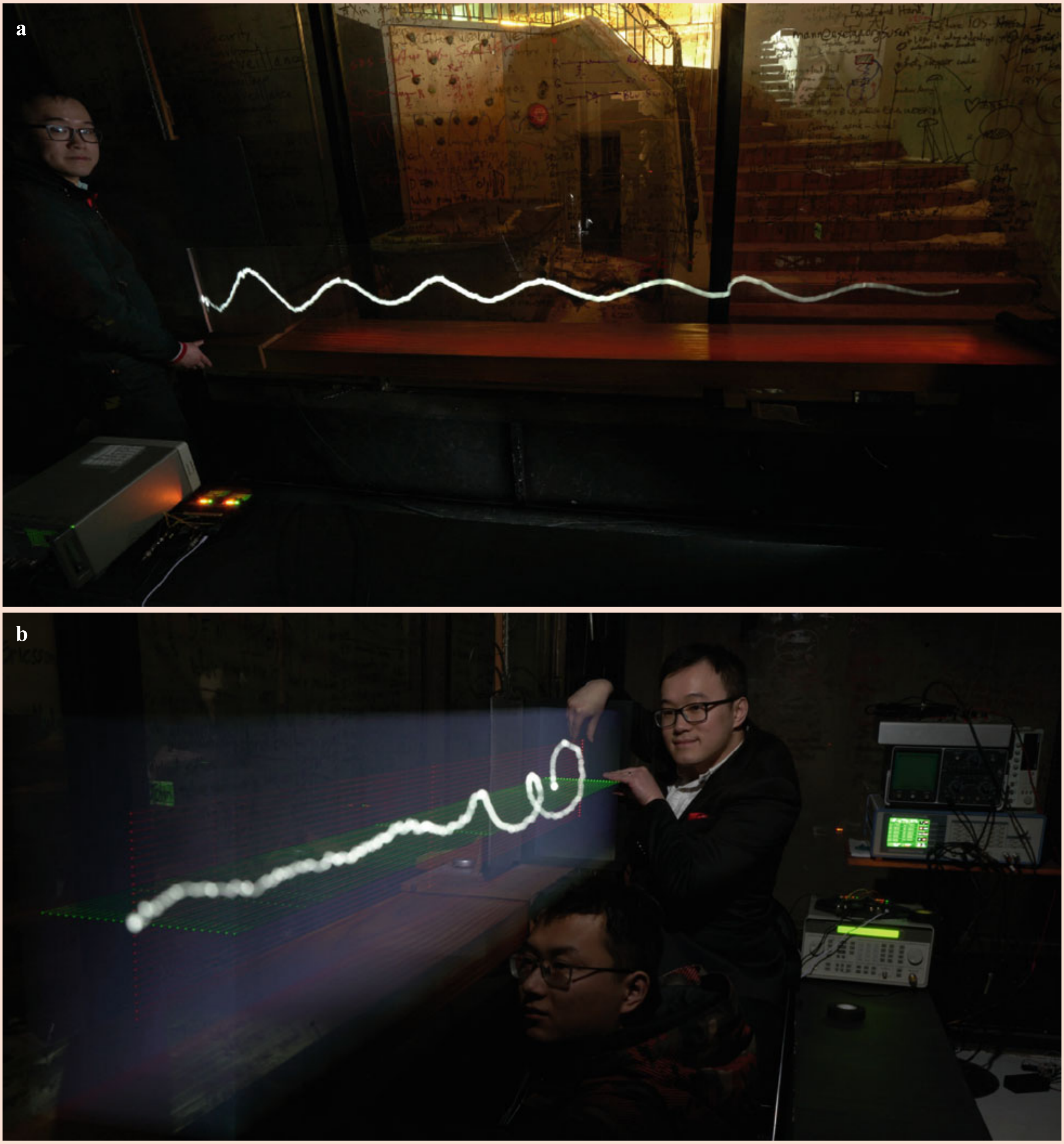


Fig. 1.32 Phenomenological augmented reality (PAR) using software-defined radio and the Sequential Wave Imprinting Machine to visualize radio waves. (**a**) A long-exposure photograph captures the real (in-phase) component of a pure carrier radio wave at 915 MHz. The transmitter can be seen in the bottom left, and the red and green glow are the two SDRs [36]. (**b**) A long-exposure photograph captures the real (in-phase) and imaginary (quadrature) components (i.e., helical wave) of a pure carrier radio wave at 915 MHz. The HP8648A transmitter can be seen on the right bottom-most [36]

of $\lambda = 32.78$ cm. This phenomenological experimental setup achieved a margin of error of 9.27%, which could be improved through calibration.

Both the in-phase and quadrature components of the waveform can be used to form a complete description of the waveform, as a helical wave, as shown in Fig. 1.32b.

This application foregoes the traditional, hardware LIA for a software-defined LIA, to maintain alignment with the SDR in terms of flexibility. Furthermore, for simplicity, an alternative method of obtaining the sitting wave is employed, by taking the phase difference. This yields the same results as Eq. 1.10. The phase angle θ and signal amplitude A of each I and Q sample are calculated using Eqs. 1.11 and 1.12, which essentially converts the I/Q data from cartesian form to polar form:

$$\theta = \arctan\left(\frac{Q}{I}\right) \tag{1.11}$$

$$A = \sqrt{I^2 + Q^2} \tag{1.12}$$

The phase angle of the reference and signal receivers, θ_r and θ_s, respectively, can be determined from Eq. 1.11, and the phase difference can be computed as

$$\theta_x = \theta_r - \theta_s \tag{1.13}$$

This method does not use a mixer and therefore does not require a low-pass filter to remove the higher-frequency term of the signal. By simply taking the phase difference, the sitting wave can be determined:

$$I_x = A_x \cdot \cos(\theta_x) \tag{1.14}$$

$$Q_x = A_x \cdot \sin(\theta_x) \tag{1.15}$$

The amplitude of the sitting wave A_x is set to the amplitude of the moving signal antenna A_S. Note that the amplitude of the reference A_r can be ignored.

The process explained above acts as a software-defined version of the LIA, where the time component of the wave from the reference receiver and the signal receiver cancel out, leaving the space component of the waveform (i.e., kx). This eliminates the need for a hardware version of the LIA by creating a software equivalent with the same working principle.

A simple smoothing function (see Eqs. 1.16, and 1.17) may be used if needed to remove noise and smooth out the signal sent from the SDRs.

$$\theta_x = (1-\beta)\cdot\theta_x(n) + \beta\cdot\theta_x(n-1) \tag{1.16}$$

$$A_x = (1-\beta)\cdot A_x(n) + \beta\cdot A_x(n-1) \tag{1.17}$$

Therefore, SDR-SWIM can be used as an inexpensive, flexible way to visualize and learn about the transmission of electromagnetic waves. Not only can the transmission of carrier waves be visualized, as presented, but interference on the waveform, i.e., due to nearby metallic objects, or data transmission and modulation schemes, could be visualized as well.

1.6.4 Electric Machines and SWIM

The electromagnetic phenomena within electric machines can be revealed using the SWIM. This allows us to visualize and understand the fundamentals of electric machines, such as current or voltage waveforms present within a motor or generator. For example, a stepper motor's current waveform is shown in Fig. 1.33 and a brushless direct current (BLDC) motor's three phases, each corresponding to the red, green, and blue waveforms shown in Fig. 1.34. Figure 1.35a is a photograph of the spinning motor with room lights on. SWIMs attached to the motor illuminate and make visible the electromagnetic fields inside the motor. Figure 1.35b is a photograph taken in the dark (i.e., with room lights off) of the spinning motor with SWIMs. For a detailed explanation of the theory and methodology of electric machines and SWIM, see [35, 37].

The application of SWIM to electric machines becomes increasingly valuable when considering the effects of loading. Figure 1.34 compares the situation when zero external load is applied and where approximately 90 g of friction force is applied by hand, for a BLDC motor. This results in a leading current waveform, where the angle difference is 1.84°.

Therefore, we can see that the SWIM is a meaningful phenomenological method for understanding three-phase motors, which allows us to visualize the rotation of magnetic fields within the motor due to varying load, in an interactive manner.

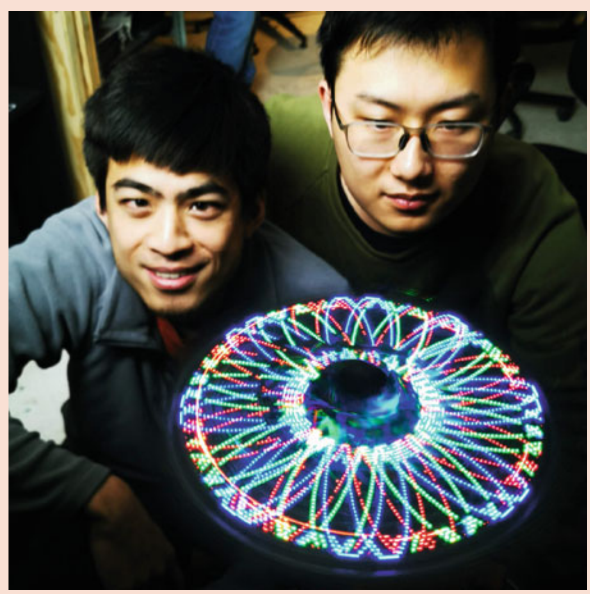


Fig. 1.33 The current waveforms inside a stepper motor are observable using the SWIM, a phenomenological augmented reality (PAR) system

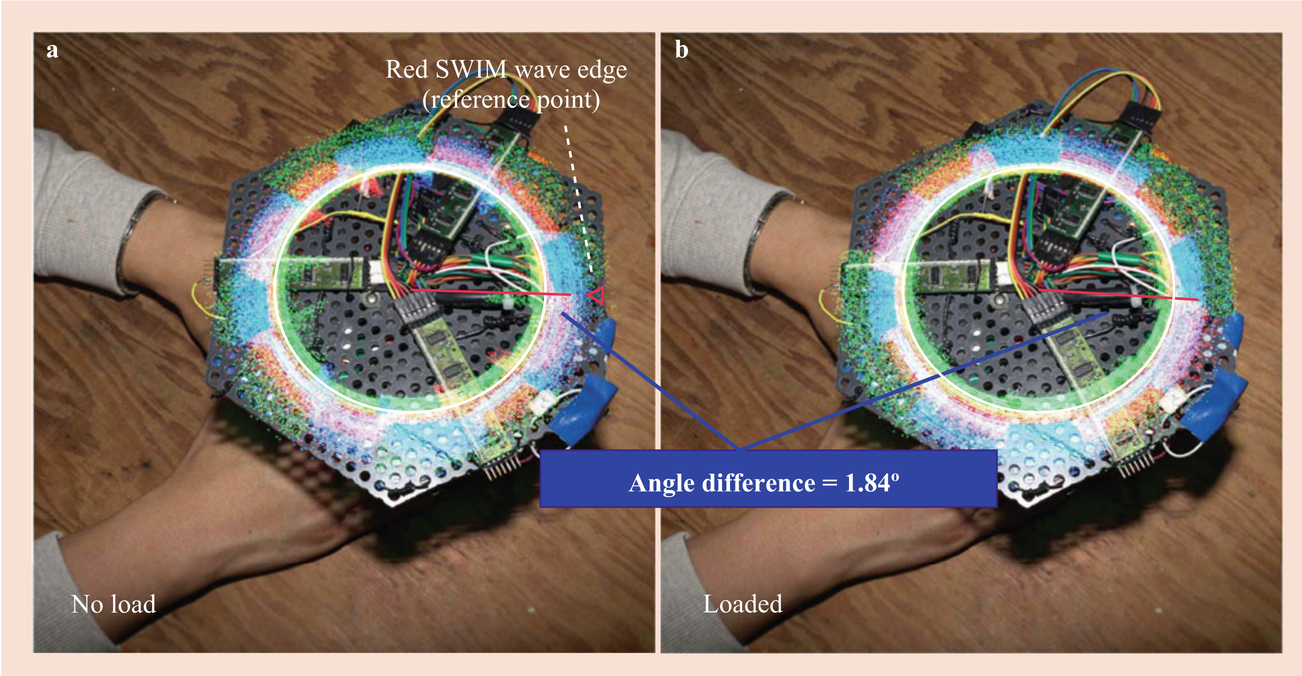


Fig. 1.34 Each phase of the three-phase motor is displayed in red, green, and blue. A difference in field waveforms can be seen when comparing between no load applied (**a**) vs. load applied (**b**) of a three-phase motor [37]

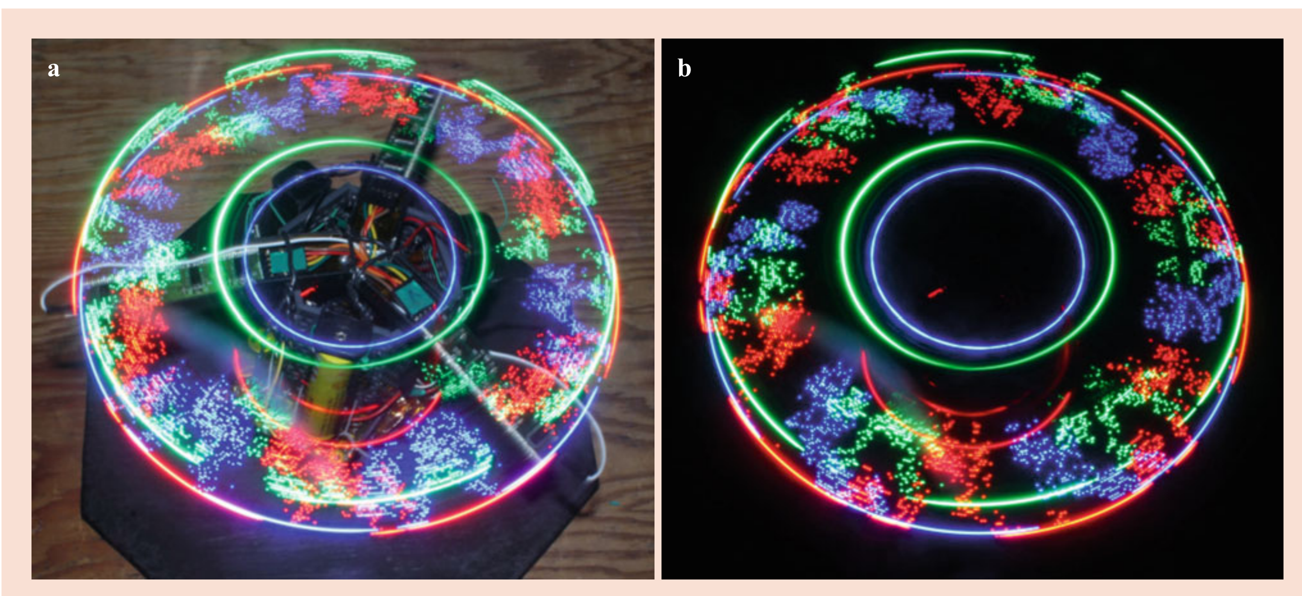


Fig. 1.35 Phenomenological augmented reality (PAR) using the Sequential Wave Imprinting Machine to visualize electromagnetic waves within a rotating motors [35]. (**a**) Photograph taken with electronic flash and room lights turned on. (**b**) Photograph taken in complete darkness

SWIM can also be used as a form of augmented reality for other kinds of electrical signals. For example, a SWIM stick was attached to the shaft of a Hammond B3 organ, to show the various musical notes as a form of augmented reality. Early prototypes also used electric motors. Electric motors typically run on 60 Hz in North America and 50 Hz in many other parts of the world.

We propose to "harmonize" the world's electricity by taking a compromise between 50 and 60 Hz, to settle on 55 Hz as an international frequency standard. The nice thing

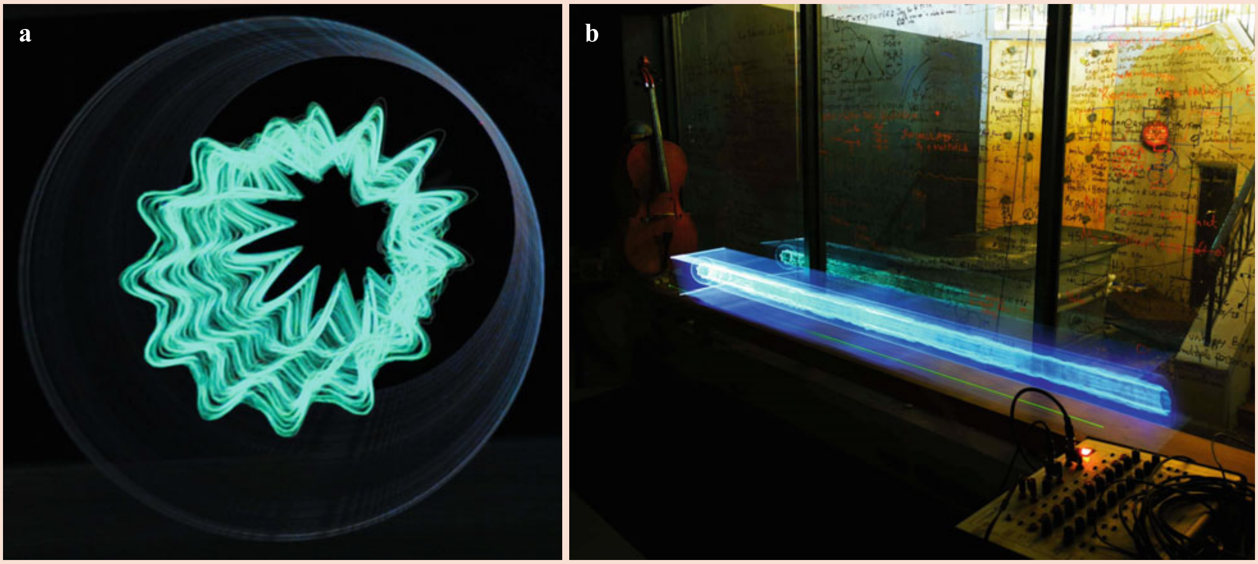


Fig. 1.36 Rotary SWIM is a simple form of augmented reality, (**a**) with which to see sound (robotically actuated cello) phase coherently. (**b**) A laptop computer running the SWIMulator is slid along a SWIM rail during a photographic exposure

about 55 Hz is that this is a musical note, "A," exactly 1/8 that of the "concert pitch" (e.g., "A 440"). The lowest note on the piano is an "A," at 27.5 Hz, so our proposed Harmonized Frequency Standard is exactly one octave above the first note of the piano.

Synchronous motors thus would turn at 3300 RPM. We have built early prototypes of augmented reality systems comprised of a linear array of light sources spun on the shaft of a 3300 RPM motor. The linear array of lights is driven by sound from a microphone located at the exact center of rotation. This results in a SWIM effect in which the mixing happens by way of the motor's rotation. The result is a pattern that is clearly visible for musical notes that are tuned to concert pitch. For example, a Pasco Fourier synthesizer connected to the device allows us to study harmonics as patterns.

More recently, we have created the SWIMulator™ which simulates the musical rotary SWIM augmented reality system. See Fig. 1.36, which shows Rotary SWIM as a simple form of augmented reality, with which to see sound, as shown in Fig. 1.36a (robotically actuated cello) phase coherently. In mechanical rotary sound SWIM, a motor is used to spin a SWIM stick at 3300 RPM, i.e., 55 rotations per second. A musical "A" note at 220 Hz then appears as a four-pedal pattern. An "A 440" concert pitch appears as an eight-pedal pattern, i.e., eight lobes, where the shape indicates the harmonic structure of the signal. Here we have a SWIMulator™ which is simulating this process. A laptop computer running the SWIMulator is slid along a SWIM rail during a photographic exposure (Fig. 1.36b).

1.7 Summary and Conclusions

We have presented multimediated reality as a proper superset of mediated (XY), mixed (X), augmented (A), and virtual (V) reality (R). Multimediated reality uses interactive multimedia in a way that is:

- Multidimensional, in a true interactive sense, i.e., direct interaction with concepts like cross-polarization, and complex-valued multidimensional mathematical quantities, e.g., Fieldary User Interfaces [10].
- Multisensory, cross-sensory (e.g., synthetic synesthesia).
- Multimodal, in many senses of the word, e.g., multimodal interaction (human-computer interaction using multiple modalities of input/output) and multimodal artifacts (artifacts that use multiple media modes, including social media, gaming, storytelling, etc) [13].
- Multidisciplinary (i.e., as a field of inquiry, involving the arts, the sciences, engineering, medicine, health, wellness, e.g., meditation in multimediated sensory attenuation tanks), etc.

A Venn diagram depicting the various realities is shown in Fig. 1.37. VR is a proper subset of AR. AR is a proper subset of mixed reality. Mixed reality and X-Reality™ are proper subsets of mediated reality (mixed, X-, and mediated/Y-reality are denoted by a dotted line in the diagram in order to indicate that either could be represented by this circle). These realities are a proper subset of multimediated reality, which

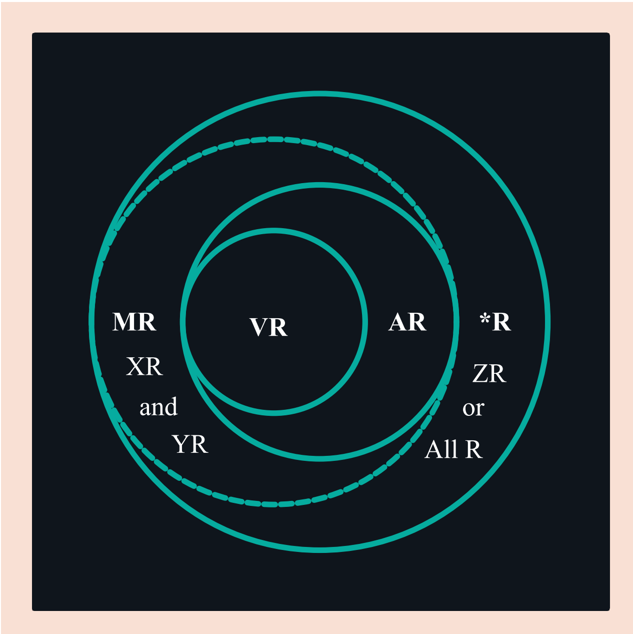


Fig. 1.37 Venn diagram showing various realities

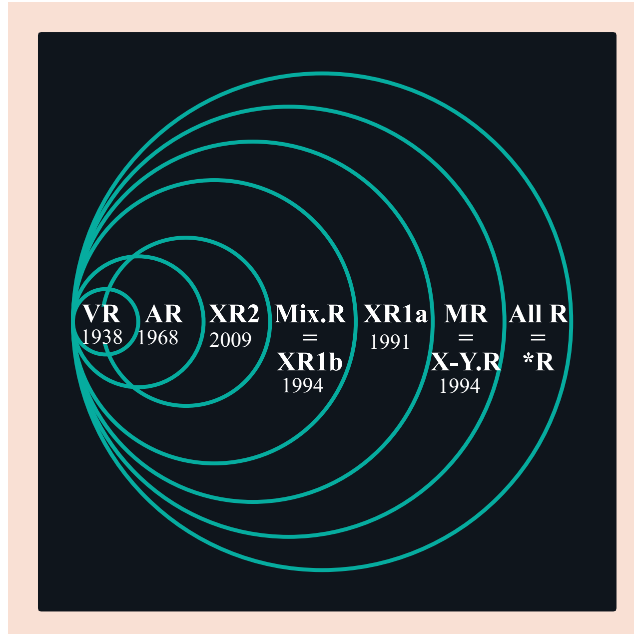


Fig. 1.38 Taxonomy of the seven realities. We hope this work can guide future research on "the realities" and help sort out the confusing mess of conflicting definitions and taxonomies that have emerged over the years. Note the multiple variations of "XR," and the possible confusion between "MR = mediated reality" and MIXed reality denoted as "MixR". In particular, we call for a return to fundamental mathematics and physics-based, and more holistic understanding of "the realities"

we proffer to call "$\star$,R" pronounced "All R," where "$\star$" denotes any of the other forms of interactive multimedia reality. Alternatively, e.g., in systems like the UNIX operating system where "$\star$" is a special reserved character, we can abbreviate "All R" as "ZR" to emphasize its multidimensional and complex-valued nature, i.e., $Z = X + iY$, where i is the imaginary unit number, $i = \sqrt{-1}$.

By combining Figs. 1.11 and 1.37, we have mostly a nested set except for XR2 which is a proper subset of mixed reality, i.e., limited to a specific kind of virtual world. See Fig. 1.38, which illustrates a taxonomy of seven realities, together with the approximate year each was first introduced, bearing in mind that some of them evolved over many years (e.g., consider that VR was introduced by Artaud in 1938, AR was introduced by Sutherland with his see-through HMD in 1968, etc.)

Amid the confusion created by a large number of different and often conflicting definitions of various kinds of reality, together with the high hype-factor of commercial interests, trademarks, "branding" and the like, this chapter aims to present a unified understanding that is both broad and deep, allowing the reader to think about "reality" while being rooted in fundamentals, deriving new ideas from first principles. This chapter will hopefully be of benefit to thought leaders and do-leaders creating fundamental advancements to the field of multimediated reality/realities.

One clarifying concept is that of a multidimensional space with nothingness at the origin, which also connects to the idea of "fluid interfaces" ("fluid user-interfaces") [21] such as immersive multimedia, VR/AR float tanks, hydraulophones, and, more generally, fluidity and "being undigital" [26].

We have presented some new forms of multisensory multidimensional interactive reality, such as multidimensional complex-valued fields and other quantities that are captured and displayed to the human senses or recorded on photographic media, video display media, or wearable media. Whereas virtual reality and other realities (e.g., augmented reality) usually show artificial graphical content (computer graphics), multimediated reality also has the capacity to show a "real reality," i.e., to make visible what is really present (but otherwise invisible) in the world around us.

Finally, we had an in-depth discussion regarding phenomenal/phenomenological augmented reality (PAR). Detailed examples of the application of the SWIM, a mechanism for PAR, were presented.

Supplementary Video Material

Examples of the applications described in this chapter can be seen in the following video(s): https://www.youtube.com/watch?v=4gfHArY_U6A

Acknowledgments The authors would like to acknowledge the many students at others at MannLab Canada, including Max, Sen, Jackson, Arkin, Alex, and Kyle. The authors appreciate the ongoing work on the Social Distancer project with Cayden Pierce, Christina Mann, and Chris Tong. Some of the work is supported, in part, by Mannlab Shenzhen.

References

1. Albright, W.F.: The Proto-Sinaitic inscriptions and their decipherment. Harvard Theological Studies (1966)
2. Anstis, S.: Visual adaptation to a negative, brightness-reversed world: some preliminary observations. In: Carpenter, G., Grossberg, S. (eds.). Neural Networks for Vision and Image Processing, pp. 1–15. MIT Press, Cambridge (1992)
3. Artaud, A.: The Theater and Its Double, 159 p (1958). First published reference to “la realite virtuell” (“virtual reality”)
4. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. **6**(4), 355–385 (1997)
5. Coleman, B.: Using sensor inputs to affect virtual and real environments. IEEE Pervasive Comput. **8**, 3 (2009)
6. Cytowic, R.E.: The Man Who Tasted Shapes: A Bizarre Medical Mystery Offers Revolutionary Insights into Emotions, Reasoning, and Consciousness. G. P. Putnam’s Sons, New York (1993)
7. Dolezal, H.: Living in a world transformed: Perceptual and performatory adaptation to visual distortion. Academic Press, Chicago (1982)
8. Euler, L.: Tentamen explicationis phaenomenorum aeris. Commentarii academiae scientiarum Petropolitanae, 347–368 (1729)
9. Ganascia, J.-G.: The generalized sousveillance society. Soc. Sci. Inf. **49**(3), 489–507 (2010)
10. Gershon, N., Kapralos, B., Mann, S.: Games, entertainment, and media for life, health, and community: IEEE GEM 2014. IEEE Consum. Electron. Mag. **4**(3), 67–72 (2015)
11. Haller, M., Landerl, F., Billinghurst, M.: A loose and sketchy approach in a mediated reality environment. In: Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 371–379. ACM, New York (2005)
12. Janzen, R., Yasrebi, S.N., Bose, A.J., Subramanian, A., Mann, S.: Walking through sight: Seeing the ability to see, in a 3-D augmediated reality environment. In: 2014 IEEE Games Media Entertainment, pp. 1–2 (2014)
13. Kress, G.R.: Multimodality: a social semiotic approach to contemporary communication. Routledge, London (2010)
14. Luce, H.R. (ed.): Special Double Issue: Photography. Life Magazine, pp. 112–113 (1966)
15. Mann, S.: ‘Mediated Reality’. TR 260. M.I.T. M.L. vismod, Cambridge (1994). http://wearcam.org/mr.htm
16. Mann, S.: Humanistic computing: “WearComp" as a new framework and application for intelligent signal processing. Proc. IEEE **86**(11), 2123–2151 (1998)
17. Mann, S.: Can humans being clerks make clerks be human? Können Menschen, die sich wie Angestellte benehmen, Angestellte zu menschlichem Verhalten bewegen? Informationstechnik und Technische Informatik **43**(2), 97–106 (2001)
18. Mann, S.: Intelligent Image Processing. John Wiley and Sons, New York (2001)
19. Mann, S.: Mediated Reality with implementations for everyday life. Presence Connect (2002). wearcam.org/presenceconnect
20. Mann, S.: Telematic tubs against terror: immersive interactive media of the post-cyborg age. Leonardo **37**(5), 372–374 (2004)
21. Mann, S.: “fl Huge UId Streams”: fountains that are keyboards with nozzle spray as keys that give rich tactile feedback and are more expressive and more fun than plastic keys. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, pp. 181–190 (2005)
22. Mann, S.: Phenomenal augmented reality: advancing technology for the future of humanity. IEEE Consum. Electron. Mag. **4**(4), 92–97 (2015)
23. Mann, S.: Surveillance (oversight), sousveillance (undersight), and metaveillance (seeing sight itself). In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1408–1417 (2016)
24. Mann, S.: Rattletale: phase-coherent telekinetic imaging to detect tattletale signs of structural defects or potential failure. In: 2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), pp. 241–248 (2017)
25. Mann, S.: Phenomenological augmented reality with the sequential wave imprinting machine (SWIM). In: 2018 IEEE Games, Entertainment, Media Conference (GEM), pp. 1–9 (2018)
26. Mann, S., Picard, R.W.: Being ‘undigital’ with digital cameras: extending dynamic range by combining differently exposed pictures. In: Proceedings of the IS&T 46th Annual Conference, pp. 422–428 (1995)
27. Mann, S., Fung, J., Moncrieff, E.: EyeTap technology for wireless electronic news gathering. SIGMOBILE Mob. Comput. Commun. Rev. **3**(4), 19–26 (1999). https://doi.org/10.1145/584039.584044
28. Mann, S., Manders, C., Fung, J.: Painting with looks: photographic images from video using quantimetric processing. In: Proceedings of the Tenth ACM International Conference on Multimedia, pp. 117–126 (2002)
29. Mann, S., Janzen, R., Aimone, C., Garten, A., Fung, J.: Performance on physiphones in each of the five states-of-matter: underwater concert performance at DGI-Byen Swim Centre (Vandkulturhuset). In: International Computer Music Conference, ICMC ’07, August 27–31, Copenhagen (2007)
30. Mann, S., Lo, R.C.H., Ovtcharov, K., Gu, S., Dai, D., Ngan, C., Ai, T.: Realtime HDR (High Dynamic Range) video for eyetap wearable computers, FPGA-based seeing aids, and glasseyes (EyeTaps). In: 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1–6 (2012)
31. Mann, S., Janzen, R., Ali, M.A., Scourboutakos, P., Guleria, N.: Integral kinematics (time-integrals of distance, energy, etc.) and integral kinesiology. In: Proceedings of the 2014 IEEE Games, Entertainment, Media Conference (GEM), Toronto. Citeseer, pp. 22–24 (2014)
32. Mann, S., Kitani, K.M., Lee, Y.J., Ryoo, M.S., Fathi, A.: An introduction to the 3rd workshop on egocentric (first-person) vision. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 827–832 (2014)
33. Mann, S., Hao, M.L., Warner, J.: Virtual reality games in sensory deprivation tanks. In: 2018 IEEE Games, Entertainment, Media Conference (GEM), pp. 1–9 (2018)
34. Mann, S., Defaz, D., Pierce, C., Lam, D., Stairs, J., Hermandez, J., Li, Q., Xiang, Y.X., Kanaan, G., Chen, H., et al.: Keynote-eye itself as a camera: sensors, integrity, and trust. In: The 5th ACM Workshop on Wearable Systems and Applications, pp. 1–2 (2019)
35. Mann, S., Do, P.V., Garcia, D.E., Hernandez, J., Khokhar, H.: Electrical engineering design with the subconscious mind. In: 2020 IEEE International Conference on Human-Machine Systems (ICHMS), pp. 1–6 (2020)
36. Mann, S., Do, P.V., Lu, Z., Lau, J.K.K.: Sequential Wave Imprinting Machine (SWIM) implementation using SDR (Software-Defined Radio). In: 2020 Seventh International Conference on Software Defined Systems (SDS), pp. 123–130 (2020)

37. Mann, S., Garcia, D.E., Do, P.V., Lam, D., Scourboutakos, P.: Moveillance: visualizing electric machines). In: 2020 22nd Symposium on Virtual and Augmented Reality (SVR), pp. 420–424 (2020)
38. Mann, S., Pierce, C., Bhargava, A., Tong, C., Desai, K., O'Shaughnessy, K.: Sensing of the self, society, and the environment. In: 2020 IEEE Sensors, pp. 1–4. IEEE, Piscataway (2020)
39. Mann, S., Wyckoff, C.: Extended reality, MIT 4–405 (1991). Available for download from http://wearcam.org/xr.htm
40. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: a class of displays on the reality-virtuality continuum. In: Telemanipulator and Telepresence Technologies, vol. 2351. International Society for Optics and Photonics, pp. 282–292 (1995)
41. Minsky, M., Kurzweil, R., Mann, S.: The society of intelligent veillance. In: 2013 IEEE International Symposium on Technology and Society (ISTAS): Social Implications of Wearable Computing and Augmediated Reality in Everyday Life, pp. 13–17 (2013)
42. Mohl, R.: Cognitive space in the interactive movie map: an investigation of spatial learning in virtual environments. Ph.D. Dissertation. Massachusetts Institute of Technology (1981)
43. Paradiso, J.A., Landay, J.A.: Guest Editors' introduction: cross-reality environments. IEEE Pervas. Comput. **8**(3), 14–15 (2009)
44. Rizzo, A.A., Galen Buckwalter, J., Neumann, U.: Virtual reality and cognitive rehabilitation: a brief review of the future. J. Head. Trauma. Rehabil. (1997)
45. Rizzo, A., Cukor, J., Gerardi, M., Alley, S., Reist, C., Roy, M., Rothbaum, B.O., Difede, J.A.: Virtual reality exposure for PTSD due to military combat and terrorist attacks. J. Contemp. Psychother. **45**(4), 255–264 (2015)
46. Robinett, W.: Synthetic experience: a proposed taxonomy. Presence Teleop. Virt. **1**, 229–247 (1992)
47. Rothbaum, B.O., "Skip" Rizzo, A., Difede, J.A.: Virtual reality exposure therapy for combat-related posttraumatic stress disorder. Ann. N. Y. Acad. Sci. **1208**(1), 126–132 (2010)
48. Scourboutakos, P., Lu, M.H., Nerker, S., Mann, S.: Phenomenologically augmented reality with new wearable LED sequential wave imprinting machines. In: Proceedings of the Eleventh ACM International Conference on Tangible, Embedded, and Embodied Interaction, pp. 751–755 (2017)
49. Stanney, K.M., Kennedy, R.S.: The psychometrics of cybersickness. Commun. ACM **40**(8), 66–68 (1997)
50. Stratton, G.M.: Some preliminary experiments on vision without inversion of the retinal image. Psychol. Rev. **3**(6), 611–617 (1896)
51. Tan, L.: Mediated reality in bearable prosthesis: a tool for surveillance and intervention. In: Proceedings of the 1st Fascinate Conference. Falmouth University, Cornwall, pp. 1–17 (2013)
52. Tang, F., Aimone, C., Fung, J., Marjan, A., Mann, S.: Seeing eye to eye: a shared mediated reality using EyeTap devices and the VideoOrbits gyroscopic head tracker. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 267–268 (2002)
53. Wiederhold, B.K., Wiederhold, M.D.: Virtual Reality Therapy for Anxiety Disorders: Advances in Evaluation and Treatment. American Psychological Association, Washington (2005)
54. Wyckoff, C.W.: An experimental extended response film. Technical report No. B-321. Edgerton, Germeshausen & Grier, Inc., Boston (1961)
55. Wyckoff, C.W.: Silver halide photographic film having increased exposure-response characteristics. US Patent 3450536A (1961)
56. Wyckoff, C.W.: An experimental extended response film. SPIE Newsletter, 16–20 (1962)

Steve Mann has been recognized as "the father of wearable computing" (IEEE ISSCC 2000) and "the father of wearable augmented reality (AR)" for his invention of "Digital Eye Glass" (EyeTap) and mediated reality (predecessor of AR). He also invented the Chirplet Transform, Comparametric Equations, and HDR (High Dynamic Range) imaging (U.S. Pat. 5828793). He received his PhD from MIT in 1997 and is a tenured full professor at the University of Toronto.

Steve received his PhD from MIT in 1997 and then returned to Toronto in 1998, creating the world's first Mobile Apps Lab (1999) as a part of his wearable computing and AR course at University of Toronto, where he is a tenured full professor in the Department of Electrical and Computer Engineering with cross-appointments to Computer Science and MIE (Mechanical and Industrial Engineering), teaching the world's first Inventrepreneurship (Invention + Entrepreneurship) course. Mann is a Visiting Full Professor at Stanford University Department of Electrical Engineering (Room 216, David Packard Building, 350 Serra Mall, Stanford, CA 94305) and is the Chair of the Silicon Valley Innovation & Entrepreneurship Forum (SVIEF).

He is also the chief scientist at the Creative Destruction Lab at Rotman's School of Management. Mann holds multiple patents and has founding or cofounded numerous companies including InteraXON, makers of Muse, "The Most Important Wearable of 2014," and Meta, a California-based start-up, bringing wearable AR glasses to a mass market (built on Mann's gesture-based wearable computing inventions [IEEE Computer, volume 30, number 2, pages 25–32, February 1997; USPTO 61/748,468, 61/916,773, and 20140184496]).

Together with Marvin Minsky, "the father of AI (artificial intelligence)," and Ray Kurzweil, Mann created the new discipline of HI (humanistic intelligence). Mann and his students cofounded numerous companies with a worth in excess of $1 billion. He also created the world's first course on "inventrepreneurship" (invention and entrepreneurship).

Mann is also the inventor of the hydraulophone, the world's first musical instrument to make sound from vibrations in liquids, giving rise to a new theory of reverse kinematics and mechanics based on the time-integral of displacement, for which Mann coined the term "absement."

Mann has authored more than 200 publications, books, and patents, and his work and inventions have been shown at the Smithsonian Institute, National Museum of American History, The Science Museum, MoMA, Stedelijk Museum (Amsterdam), and Triennale di Milano.

He has been featured in AP News, *New York Times*, *Los Angeles Times*, *Time*, *Newsweek*, *Fortune*, WiReD, NBC, ABC, CNN, David Letterman, CBC-TV, CBS, *Scientific American*, *Scientific American Frontiers*, Discovery Channel, Byte, *Reuters*, *New Scientist*, *Rolling Stone*, and BBC.

Phillip Vinh Do received his M.A.Sc. degree in Electrical and Computer Engineering from the University of Toronto in 2021, his M.Eng. degree in Electrical and Computer Engineering from McMaster University in 2019, and his B.Eng. degree in Electrical Engineering from McMaster University in 2014. During his graduate studies, he has coauthored six peer-reviewed publications with IEEE and EAI. He has served as a reviewer for ten academic papers with IEEE. His current research interests include ASICs, signal processing, augmented reality, and biomedical devices.

He has industry experience in hardware design at the system, board, and chip levels. From 2014 to 2019, he worked as a hardware design engineer at global companies, including Thales and L3Harris. Since 2020, he is working as an ASIC design engineer at a Canadian start-up company, The Six Semiconductor.

He enjoys spending quality time with friends and family, travelling, snowboarding, and exercising. He has backpacked across Western Europe and achieved a 1st-degree black belt in Kenpo Karate.

Tom Furness is an amalgam of professor, inventor, virtual reality pioneer, and entrepreneur in a professional career that spans 55 years. He has made notable contributions in photonics, electro-optics, and human interface technology and is considered the "grandfather" of virtual reality. Tom earned a BS degree in Electrical Engineering from Duke University and the Ph.D. in Engineering and Applied Science from the University of Southampton, England. Tom is currently a professor of Industrial and Systems Engineering with adjunct professorships in Electrical Engineering, Mechanical Engineering, and Human Centered Design and Engineering at the University of Washington (UW), Seattle, Washington, USA. He is the founder of the Human Interface Technology Laboratory (HIT Lab) at UW and founder and international director of the HIT Lab NZ at the University of Canterbury, Christchurch, New Zealand, and the HIT Lab Australia at the University of Tasmania, Launceston, Tasmania. He is also an Erskine Fellow and Adjunct Professor at the University of Canterbury and an Adjunct Professor at the University of Tasmania.

Prior to joining the faculty at the UW, Tom served a combined 23 years as an US Air Force officer and civilian scientist at the Armstrong Laboratory at Wright-Patterson Air Force Base, Ohio, where he developed advanced cockpits and virtual interfaces for the Department of Defense. He is the author of the Super Cockpit program and served as the Chief of Visual Display Systems and Super Cockpit Director until he joined the University of Washington in 1989. He is credited as a pioneer in developing virtual reality and augmented reality.

In addition to his academic appointments, Dr. Furness was the chairman and president of the first augmented reality company: ARToolworks Inc. He also runs his own "skunkworks" company: RATLab LLC (RAT = rockin' and thinkin') where he and his colleagues develop advanced technologies for spinoff companies. His current projects deal with advancing the state of the art in retinal scanning displays, developing pulse diagnosis as an early warning system for cardiovascular disease, the use of structured light to characterize matter, and new approaches for characterizing the human breath. He is the founder and chairman of the Virtual World Society, a nonprofit for extending virtual reality as a learning system for families and other humanitarian applications.

Since the beginning of his career Tom has continued to play an active role in virtual and augmented reality development and application. In 1998 he received the Discover Award for his invention of the virtual retinal display. In August 2014 he presented a keynote address: "Seeing Anew" at the IEEE International Symposium on Mixed and Augmented Reality in Munich, Germany, and a keynote titled "Being the Future" at the Augmented World Expo June 2015 in Santa Clara, California, where he received the first lifetime achievement award for his 50-year contribution to the VR and AR industries. In March 2016 Tom received the IEEE VR Career Award for his lifetime contributions to the fields of virtual and augmented reality and in October 2016 the Virtual Reality Foundation of Los Angeles Proto Award also for lifetime achievement. Tom is coinventor of the ChromaID technology licensed to Visualant Inc. and received the 2013 SPIE Prism Award for this work.

Tom lectures widely and has appeared in many national and international network and syndicated television science and technology documentaries and news programs. He has testified before the US Senate Commerce Committee. He is the inventor of the personal eyewear display, the virtual retinal display, and the HALO display and holds 24 patents in advanced sensor, display, and interface technologies. With his colleagues and students Dr. Furness has published over 400 papers and conference proceedings and started 27 companies, two of which are traded on NASDAQ at a market capitalization of >$12 B (USD). He is a fellow in the IEEE and a member of the Computer Society and Photonics Society of the IEEE.

Tom and his wife Linda live in Seattle, Washington, Tofino, British Columbia, and Maui, Hawaii. They have two daughters and six grandchildren.

Dr. Yu Yuan is a visionary researcher, inventor, and entrepreneur in the areas of consumer technology, transportation, Internet of Things, and digital transformation. He founded Senses Global Labs & Ventures (also known as Senses Global Corporation or 0xSenses Corporation in different countries), a multinational technology company specializing in virtual reality, augmented reality, and human augmentation. Dedicated to "creating better worlds" as its long-term vision, the company is developing technologies, infrastructures, ecosystems, and resources needed for massively multiplayer ultrarealistic virtual experiences. The company is also engaged in technology consulting, technology transfer, and system integration services to help with the digital transformation of its clients in various industries. It has been proactively supporting standards development for emerging technologies (VR/AR/MR/XR, blockchain, AI, IoT, big data, digital twin, smart lifestyle, etc.) in collaboration with many other companies and institutions. Prior to this he served as the president of CATE Global, a think tank focusing on bringing world-class expertise to clients and projects all over the world. Prior to that he worked for IBM Research as a research scientist and was a key contributor to IBM's Cell Broadband Engine, Smarter Planet, and IoT initiative. He has been a passionate volunteer in various leadership positions at IEEE and other professional communities. His outstanding service in IEEE standards activities at different levels (working groups, standards committees, and governance at higher levels) has been widely appreciated by standards developers, individual members, and corporate members. He has been serving as the chair of IEEE Consumer Technology Society Standards Committee (CTS/SC) since 2015. He is also serving as the chair of IEEE VR/AR Standards Committee (CTS/VRARSC), the vice chair of IEEE Photonics Society Standards Committee (PHO/SC), the secretary of IEEE Consumer Technology Society Blockchain Standards Committee (CTS/BSC), the chair of IEEE VR/AR Working Group (IEEE P2048.1 12 Standards), the chair of IEEE VR/AR Advisory Board, and a member of IEEE Technical Activities Board Committee on Standards. He has a PhD in Computer Science, Tsinghua University.

Jay Iorio is a technology strategist for the IEEE Standards Association, specializing in the emerging technologies of virtual worlds and 3D interfaces. In addition to being a machinimatographer, Iorio manages IEEE Island in Second Life and has done extensive building and environment creation in Second Life and OpenSimulator.

Zixin Wang received his B.Sc. degree in Physics, M.S., and Ph.D. degrees in Optical Engineering from Sun Yet-Sen University, in 1999, 2002, and 2006, respectively. In 2003, He joined AMS02 TTCS electronics group and contribute mainly to the TTCE realization and function test, TTCS ground support system.

Thanks to experience in AMS02 project, Zixin Wang started up a company to design and produce lock-in amplifier products, which are widely used in China and the United States.

In 2006, he joined the Department of Microelectronics at Sun Yat-Sen University, as an assistant professor. His specific research interests are low noise electronic techniques, lock-in amplification system, mixed-signal system, and high level synthesis. He is currently an associate professor.

"Precision Thermal Control System for Large-scale Space Experiments Based on an Active Two-Phase Fluid Loop" won the second prize of Science and Technology of Guangdong province (2016).

History of Augmented Reality

2

Raffaele Vertucci, Salvatore D'Onofrio, Stefano Ricciardi, and Maurizio De Nino

Contents

2.1 **AR Basics: Operational Definition and Enabling Technologies** . . . 35

2.2 **AR Enabling Technologies** . . . 36
2.2.1 Positional Tracking for AR . . . 36
2.2.2 Co-registration and Rendering of Visual Contents . . . 38
2.2.3 Visualization . . . 38

2.3 **The Past of AR** . . . 40

2.4 **Current AR Technologies and Applications** . . . 43
2.4.1 Main Application Fields . . . 44

2.5 **Emerging Trends and Open Challenges for AR** . . . 46

2.6 **Conclusions** . . . 48

References . . . 48

Abstract

Augmented reality (AR) is emerged in the last decade as a potentially disruptive technology capable of superimposing virtual objects generated by a computer onto the real world surrounding the user. By using see-through displays, AR systems make the user perceive both the real surrounding environment and virtual elements in a consistent way with regard to user point of view and virtual content size. The recent developments of low-cost AR technologies and mixed reality (MR) devices such as Google Glass, Microsoft Hololens, Vuzix, and many others are capturing interest of users and researchers, suggesting that AR could be the next springboard for technological innovation as also highlighted but its inclusion as enabling technology of Industry 4.0 paradigm. However, the AR is not as young as it seems. The concept of AR was formulated in the 1960s, and the first commercial AR tools appeared in the late 1980s. This chapter presents an overview of AR, the historical evolution since its first appearance, its usage in the most relevant domains including the emerging instance of Mobile/Wearable AR applications, the technical challenges in implementing AR systems, as well as main technological and applicative trends for the near future.

Keywords

Augmented reality · AR history · AR applications · AR technologies

2.1 AR Basics: Operational Definition and Enabling Technologies

Over the past 30 years, various definitions of augmented reality (AR) have been introduced. Some currents of thought, for example, inspired by [1], have defined augmented reality as a derivative of virtual reality, while others affirmed the opposite. Generally speaking, AR is a technology used to superimpose virtual information about the real environment, combining reality data with computer-generated data. AR can also be described as the representation of an altered reality, in which artificial sensory information is superimposed on the normal reality perceived through the senses. The user of AR applications will then see three-dimensional virtual objects, videos, sounds, and perceive tactile or even olfactory sensations superimposed on the real world, which have been added by means of a computer. An AR system can be implemented on various types of platforms: mobile or desktop devices, vision, listening, and manipulation devices. The aim of an AR system is to create an environment in which the user

R. Vertucci · S. D'Onofrio (✉)
Leonardo S.p.A, Giugliano in Campania, Italy
e-mail: raffaele.vertucci@leonardocompany.com; salvatore.donofrio@leonardocompany.com

S. Ricciardi
Department of Biosciences and Territory, University of Molise, Pesche, Italy
e-mail: sricciardi@unisa.it; stefano.ricciardi@unimol.it

M. De Nino
Digitalcomoedia S.r.l, Naples, Italy
e-mail: m.denino@digitalcomoedia.com

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_2

is able to live a sensory experience enriched with information and virtual elements, allowing the involvement of the user in certain applications. AR is therefore both a visual paradigm and an interface, based on a combination of information generated by a computer (static and dynamic images, sounds, tactile sensations), combined with the real environment in which the user is located. Though, according to the original definition provided by Milgram, augmented reality is a subset of mixed reality (MR), this original definition refers to an idea of AR in which the real and virtual environments appear to be complementary and merged in a continuous manner. The virtual elements integrate spatially with the physical ones in a very realistic and effective way, and the system continuously interacts with the user, even if only to update the point of view and recreate the digital elements to be placed in the environment. However, this paradigm has evolved toward a more extensive representation of the realities based on the way they are mediated; yet regardless to the terminology and the related technological and applicative implications, AR (and more in general MR) represents a transversal disruptive technology able to deeply transform the world of work, the way we interact with others, and the society in general. It is also expected to produce a huge global market, having the potential to transform almost any industrial process. The aim of this chapter is to cover the fundamental aspects of this technology as well as an insight on the near future.

The rest of this chapter is organized as follows: Sect. 2.2 describes the main software and hardware AR enabling technologies, Sect. 2.3 resumes the past of augmented reality since its first proposition, Sect. 2.4 describes some of the most representative examples of AR-related devices, development environments, and applications, while in Sect. 2.5 AR evolution, future trends and challenges are explored. Finally, Sect. 2.6 draws some conclusions.

2.2 AR Enabling Technologies

Regardless of its purpose and features, any augmented reality system performs three main tasks: *Tracking*, *Co-registration/Rendering*, and *Visualization*, and each of these tasks typically involves both software and hardware components. *Tracking* takes care of defining a user mapping in real time, tracing its position within the scene, and thus providing the position and orientation of the observer's point of view with respect to a global reference system assigned by convention to the environment in which the observer itself is located. *Co-registration* is the process of aligning the virtual objects to be inserted in the real world from the point of view of the observer on the scene itself, applying at least the appropriate geometric transformations and, ideally, virtual object/s illumination, shadow casting/receiving to/from the real world, as well as its partial/total occlusion. The *Rendering* process is directly concerned with the real-time generation of visual content to be superimposed on the real environment. Finally, *Visualization* represents the process allowing the user to perceive the result of compositing virtual and real through different types of devices and technologies.

2.2.1 Positional Tracking for AR

Positional tracking is a technology that allows a device to estimate its position relative to the surrounding environment. It is based on a combination of hardware and software to obtain the detection of its absolute position. Positional tracking is an essential technology for augmented reality (AR), which allows user to track his/her point of view with six degrees of freedom (6DOF) in order to make the insertion of virtual content visually coherent to the real world. One of the greatest difficulties in AR applications is, indeed, the real-time calculation of the user's point of view. This is essential to ensure that virtual objects are perfectly aligned with real ones, since only in case of perfect alignment can AR information be a valid aid. The position of an object in space is, in fact, defined univocally by the three coordinates (x, y, z) and the orientation with respect to a reference system, indicated by three angles: yaw, pitch, and roll (Fig. 2.1).

The most important operating parameters of a tracking system include the *work volume*, the *sampling frequency*, the *resolution* and the *latency*:

- *Work volume* defines the region of space in which the system works properly.
- *Sampling frequency* is the frequency with which the calculator detects and updates the variables.

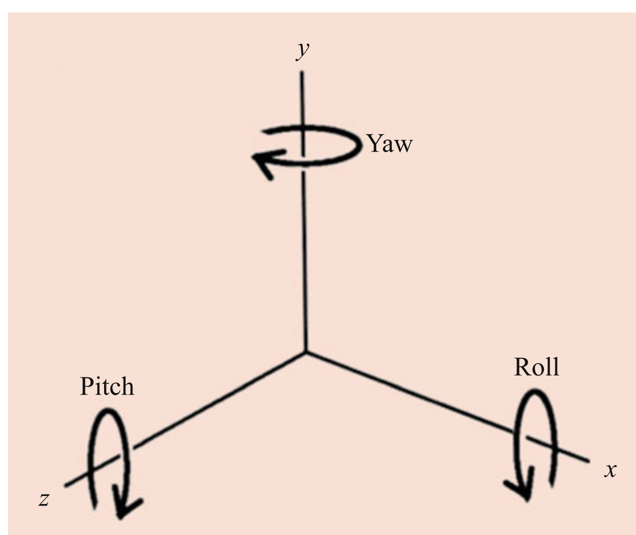


Fig. 2.1 Six degrees of freedom system (6DOF)

- *Resolution* is the smallest position variation detected.
- *Latency* is the time elapsed from the event to its recognition, or the speed of response.
- *Accuracy* is the extent of the measurement error.

Each type of application can require particular values of the above parameters, for example, a visual assessment system of the environmental impact of a building will require a very large volume of work, but not necessarily a high resolution, while the opposite could be true for an assistance system for the assembly of electronic components. 3D motion tracking devices can be classified according to the technology used for measurement: *mechanical*, *electromagnetic*, *acoustical*, *inertial*, *optical*, and *hybrid*.

Mechanical. These systems use optical potentiometers or encoders to measure the rotation of the constraint pins of an exoskeleton that the user must wear [2]. Known the angles of each joint and the lengths of the rods of the kinematic chain, it is possible to easily calculate the position and orientation of the tracked object. Another type of mechanical tracking measures the forces exerted by a strain gauge. They are characterized by constructive simplicity and the lack of transmission-reception units which reduce their cost and latency. They are also less sensitive to interference from the external environment such as, for example, electromagnetic fields. The moving parts are however subject to wear and tear; in addition, the user is bound in the movements.

Magnetic. They consist of a base station that generates a magnetic field and a receiver. This method measures the intensity of the magnetic field in different directions, with the strength of the field decreasing with increasing distance between the receiver and the base station and also allowing for orientation measurement [3, 4]. For example, if the receiver is rotated, the distribution of the magnetic field along the various axes is changed. In a controlled environment, the accuracy of magnetic tracking is of the order of a millimeter. However, it can be influenced by the interference of conductive materials near the sensor emitter, other magnetic fields generated by other devices, and ferromagnetic materials in the detection area.

Acoustical. The system has an emitter and a receiver. The first generates a sound signal, while the second (a microphone) collects it and measures the time taken by the sound to travel the path (called flight time or TOF). A computer collects data from multiple devices (three pairs are needed to measure six degrees of freedom) and calculates the position and orientation [5]. The speed of sound varies with the environmental conditions, so they are not very precise if air pressure, temperature, and humidity are not taken into account. Furthermore, these systems are unusable in environments with walls that reflect sound waves, which create ghost images of the sources. Finally, they have a limited operating volume.

Inertial. Inertial tracking is made possible by the use of accelerometers and gyroscopes. Accelerometers measure linear acceleration, which is used to calculate the speed and position of the object relative to an initial point through the mathematical relationship between position in time and speed, speed, and acceleration. The gyroscope measures angular velocity. It is a solid-state component based on microelectromechanical systems (MEMS) technology and works on the same principles as a mechanical gyroscope. From the angular velocity data provided by the gyroscope, the angular position relative to the starting point is calculated. This technology is inexpensive and can provide high update rates and low latency [6, 7]. On the other hand, calculations leading to the position of the object can result in a significant drift in the position information, considerably decreasing the accuracy of this method.

Optical. For optical tracking, various methods are available. The common element among all is the use of tracking cameras to collect positional information. If the camera is located on the device being tracked, this modality is referred as *inside-out tracking*, whereas outside-in tracking refers to the camera being located in the environment where the tracked device is within its view. Optical tracking can exploit a variety of different algorithms, such as Simultaneous Location and Mapping (SLAM) [8], Parallel Tracking and Mapping (PTAM) [9], and Dynamic Tracking and Mapping (DSLAM) [10, 11].

Marker Based. This optical tracking method uses a specific pattern of markers placed on an object. One or more cameras search for markers, using algorithms to extract the position of the object from visible markers. From the difference between what the camera is detecting and the known marker model, an algorithm calculates the position and orientation of the tracked object. The pattern of markers that are placed in the tracked object is not random. The number, position, and arrangement of the markers are carefully chosen to provide the system with all possible information so that the algorithms do not have missing data [12]. There are two types of indicators: passive and active. Passive markers reflect infrared (IR) light towards the light source. In this case, the camera provides the IR signal which is reflected by the markers for detection. Active markers are IR lights that periodically flash and are detected by cameras. The choice between the two types of markers depends on different variables such as distance, surface type, required viewing direction, and others.

Markerless. Tracking can also be performed in markerless modality. Markerless tracking uses only what sensors can observe in the environment to calculate the position and orientation of the camera. The method depends on tracked object shape and appearance rather than on markers and can use a model-based approach or perform image processing to detect features that provide data to determine position

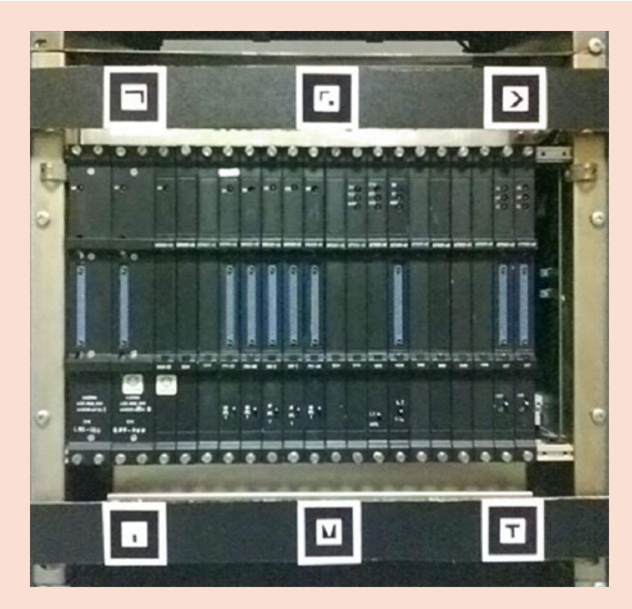


Fig. 2.2 Markers applied to an industrial rack for optical tracking. (Courtesy of Leonardo S.p.A)

and orientation [13]. Markerless tracking methods look for correspondences between the visual features present in the current frame and those observed previously. A match can only succeed if at least part of the current view has been previously acquired. These matchings are the inputs for the algorithm that calculates camera position and orientation. In contrast to marker-based tracking (Fig. 2.2), a markless approach allows the user to walk freely in a room or new environment and still receive positional feedback , expanding the applicability range. In addition to the compromise between precision and efficiency required by the complex computing tasks to be performed in real time, markerless tracking also presents several challenges, such as managing large scenarios, objects of small parts, variable lighting conditions, and materials with low roughness, reflective, and transparent properties.

Hybrid. This approach is based on the fusion of data from multiple sensors and is a method of using more than one tracking technique in order to improve the detection of the position and orientation of the tracked object. By using a combination of techniques, the disadvantage of one method can be compensated for by another. An example of this could be the combination of inertial tracking and optical tracking [14]. The former can develop a drift, and the latter is likely to be affected by markers occlusion. By combining both, if the markers are occluded, the position can be estimated by the inertial trackers, and even if the optical markers are completely visible, the inertial sensors provide updates at a higher rate, improving overall positional tracking accuracy.

2.2.2 Co-registration and Rendering of Visual Contents

Once the user point of view is known by means of the exact position and orientation in 3D space of his head provided by the tracking system, it is possible to precisely align and scale the visual contents to augment the scene observed. This is achieved by assigning the incoming tracking info (x, y, z, yaw, pitch, roll) to the rendering camera in real time, so that for any head movement a coherent virtual content transformation will be applied. This task is performed by specialized AR software libraries and SDK.

2.2.3 Visualization

Once the rendering of co-registered graphics is performed, to achieve the illusion of the coexistence of virtual and real imagery, a combination of the two is required. This can be done by means of different devices and technologies, either in an *immersive* or *non-immersive* modality. Immersive AR requires the subject to wear an *optical* or *video see-through headset*, while non-immersive AR can be achieved on a desktop by means of a simple monitor, or directly on real world by means of projection systems.

Monitor Based. A first way to visualize an increase is to observe it on a monitor. A video camera records the environment, the virtual content is superimposed on the video stream, and the result is displayed on a monitor any size, but without any immersion (Fig. 2.3).

Optical See-through. A see-through optical headset employs an optical beam divider, consisting of a translucent mirror which transmits light in one direction and simultaneously reflects light in the other. They are therefore partially transmitting technologies since, looking through the lens, you can see the virtual image superimposed on the real one (see Fig. 2.4) [15]. A characteristic of optical combiners is that they reduce the light intensity of the real scene, so that only 20–30% of real light passes through these displays to reach the eyes. In general, in optical viewers, the augmented field of view is only a fraction of the see-through field of view [16]. Furthermore, the opacity of virtual contents combined with the real images is never total, especially in overlap with very bright environments.

Video See-through. Video see-through devices work using two cameras located on the headset, one for each eye. These cameras provide the user with a view of the real world, and the video stream captured by them is combined with the virtual components created by the computer. The result is transmitted to the monitors placed in front of the user's eyes, in the display. The diagram in Fig. 2.5 illustrates the process. In a see-through video-based application, the real-world frame shown in the HMD is the same video frame used

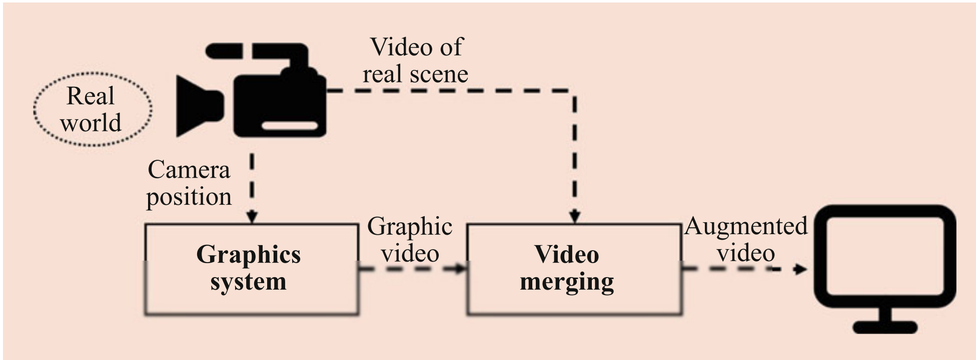


Fig. 2.3 Schematic view of monitor-based AR

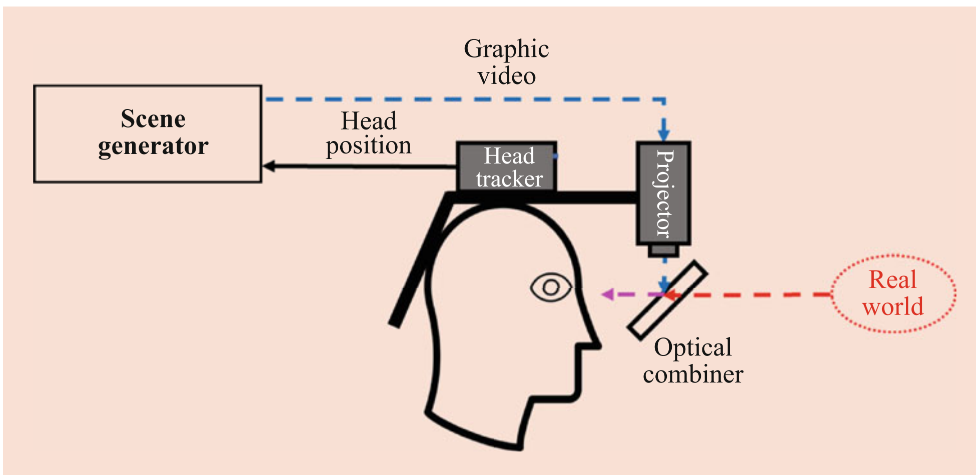


Fig. 2.4 Schematic view of an optical see-through HMD

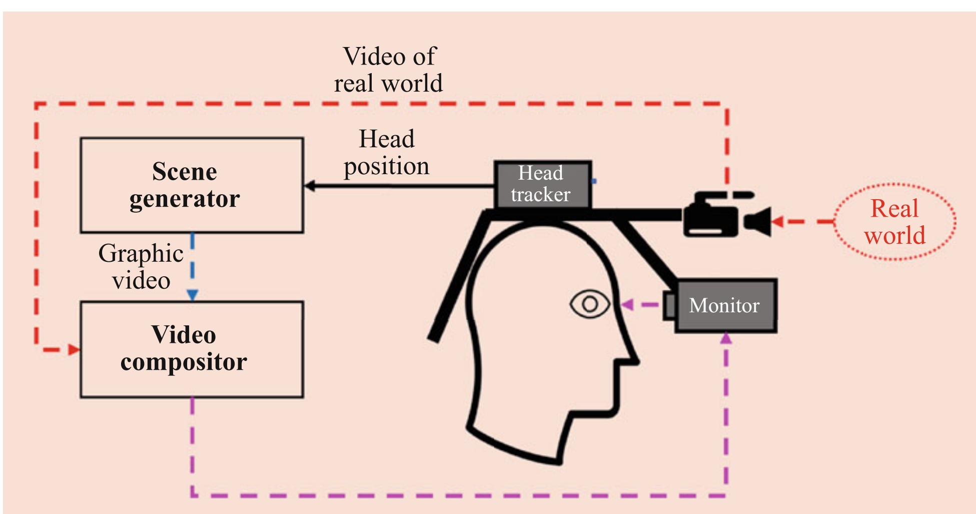


Fig. 2.5 Schematic view of a video see-through HMD

to calculate the position and orientation of the head, so that the virtual image appears exactly superimposed on the real scene [17]. This is achieved by avoiding showing the video frame on the display until the image processing is completed. Mobile smart devices, such as smartphones and tablets, can be legitimately considered video see-through devices, since they have built-in cameras and display providing a live capture of the surrounding environment and an augmentation of this environment, thanks to their real-time rendering and compositing capabilities.

Projection Based. These AR displays use projection systems to illuminate real objects with specific computer-generated images. This type of system is however sensitive to the reflectivity and physical characteristics of the illuminated objects; for example, opaque and dark bodies limit the type of information that can be displayed; also the alignment of the projectors with the projection surface is a critical factor for the generation of a credible effect [18].

2.3 The Past of AR

The history of the AR begins with a prediction of what would have been the rise of the AR in the next 100 years with the *Pokémon GO* application. It was in the year **1901** when the writer and futurist **Frank L. Baum**, author of *The Wonderful Wizard of Oz*, introduced in his novel *The Master Key* [19] the *Character Marker*, special glasses capable to display the nature of characters through a letter shown on their head.

It consists of this pair of spectacles. While you wear them every one you meet will be marked upon the forehead with a letter indicating his or her character. The good will bear the letter 'G,' the evil the letter 'E.' The wise will be marked with a 'W' and the foolish with an 'F.' The kind will show a 'K' upon their foreheads and the cruel a letter 'C.' Thus you may determine by a single look the true natures of all those you encounter.

However, the first experiments to create an augmented reality occurred only in the 1957, when an American cinematographer and visionary, called Morton Heilig, described in detail his vision of a multisensory theater that could transmit images, sounds, vibrations, and smells to spectators. In 1962, he obtained the US patent No. 3,050,870 [20] and built the first prototype of multisensory theater, together with five short films, called *Sensorama* (see Fig. 2.6), a mechanical device, very bulky, capable of projecting 3D videos on a stereoscopic color display, with stereo sounds, perfumes, wind in the hair, and vibrations, through which it simulated for the first time a motorcycle ride during which the sensory elements are activated at the right time, creating an adrenaline experience for the user, as Howard Rheingold told of his bicycle tour experience [21].

The purpose of the *Sensorama* was not only to make 3D films but also to respond to the growing demand for ways and means to teach and train civil, military, and industrial personnel without particular risks for it.

Although the *Sensorama* is still functional, the Heilig's project was interrupted, and it took a few more years to fully understand the potential of AR and to convince entrepreneurs, afraid of the high costs of making 3D films, to invest in this emerging technology.

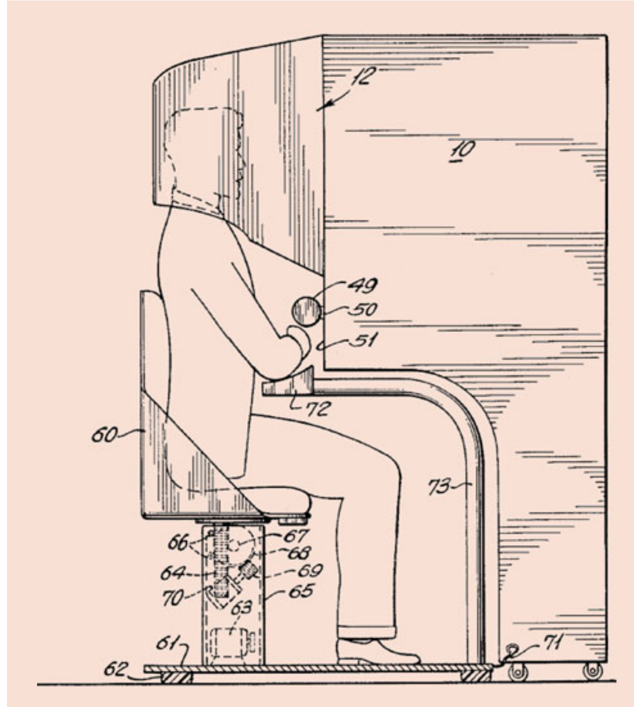


Fig. 2.6 The Sensorama Machine, from US Patent #3050870

It was the American researcher Ivan E. Sutherland already known for his *Sketchpad* application [22] who helped by his students, in 1968, to create the first head mounted display system (HMD) tracked mechanically through a ceiling-mounted hardware, called *The Ultimate Display* [23], thus opening the way to the world of augmented and virtual reality. That HMD was partially see-through and consisted of two connected cathode ray tubes that projected images directly into the user's eyes allowing the superimposition of 3D images on real objects. Sutherland's device was so heavy that it had to hang from the ceiling. Its structure was very similar to a medieval torture instrument, from which it had the nickname *The sword of Damocles*. The device was primitive, both in terms of interface and realism, but the images were already synchronized with the movements of the head. Subsequently, Sutherland went to teach at the University of Utah and here partnered with David Evans to found the Evans and Sutherland company, continuing the experiments on AR. In 1969, in parallel with the experiments of Sutherland and Evans, the research scientist Thomas A. Furness, while working at the Armstrong Laboratory at Wright-Patterson Air Force Base, developed a prototype of HMD which he called *Visually Coupled System* (VCS), and then he continued his experiments which led, in 1981, to the creation of a virtual cockpit flight simulator, *The Super Cockpit* [24], and to the foundation of the Human Interface Technology Lab, the first research laboratory for virtual reality at the University of Washington. The developments started by Sutherland and Furness were then continued by several government agencies and private companies which led to different HMD designs.

The interest in this new technology led, in the decade between 1970 and 1980, to an in-depth study on AR, and in 1975, the developments made a huge leap, thanks to the project *Videoplace* [25] by Myron W. Krueger who established a laboratory of artificial reality in which a projection system and video cameras were combined in order to produce shadows on the screen. The *Videoplace* consisted of two or more nearby or even remote environments, characterized by special hardware, projectors, and video cameras to allow users to interact with each other through this technology. User movements were recorded, processed, and represented on the screen through the use of shapes. Hence, it was possible to convey the sense of presence to users within an environment shared with other users. Through the *Videoplace* therefore, two people in different rooms, each containing a projection screen and a video camera, were able to interact and communicate through their images projected in a shared space on the screen. In the first version of 1974, no computer was used for data processing.

In the 1980s, the father of wearable computing and self-proclaimed cyborg Stephen W. G. Mann presented the prototype of *EyeTap* [26], a wearable device very similar to Google Glass, although more bulky and rough, which allows the user to superimpose images on his eyes recorded by a camera and processed by a wearable computer. Hence, what the user looked at was immediately processed by the computer which generated information to be superimposed on the real vision, thus creating an augmentation of the perceived reality. The *EyeTap* system used a splitter to send the same scene simultaneously to the camera and the eye. The acquired image was then processed by the camera and sent to a computer that after a processing sent it to a projector that presented it to the other side of the splitter and then to the eye. Hence, the image was superimposed on the real scene. The EyeTap differs from HMDs because HMDs are typically used only to provide information, or add information into what a user perceives.

Although the concept of augmented reality (AR) was explored for over two decades, the term *augmented reality* became official only in the 1990s when it appeared in a document written by two Boeing engineers and researchers, Thomas P. Caudell and David W. Mizell [27], which described an AR system for reducing errors when creating wirings on aircraft and other aerospace industry systems. Boeing was one of the first companies to use the augmented and virtual reality as part of its commercial strategies.

The years that followed led many interesting developments in AR which became a research field, and the first scientific articles, about the optimization of displays and image recognition techniques for the improvement of motion tracking systems, began to come out.

In 1992, the researcher and inventor Louis Rosenberg of the US Air Force Research Laboratory (AFRL) explained the beneficial effects of perceptual overlay at the master interface [28, 29] and how tools and fixtures in the real world could enhance human performance by guiding manual operations, providing localizing references, reducing mental workload, and increasing precision. Based on the concepts outlined, Rosenberg developed the first immersive AR system called *virtual fixtures* which allowed military personnel to virtually control and drive machines to perform tasks such as the training of their US Air Force pilots on more secure flight practices.

Since 3D graphics were still too slow, in the early 1990s to present a photorealistic augmented reality recorded in space, virtual devices used two real physical robots, controlled by an exoskeleton of the upper body, worn by the user. To create the immersive experience for the user, a unique optical configuration was used which involved a pair of binocular loupes aligned in such a way as to advance the user's view on the arms of the robot so as to appear recorded in the position exact of the user's real physical arms. Additionally, virtual fixtures could be viewed in overlapping to the real world using a composition of the real image with rendered virtual fixtures. Therefore, not only virtual fixtures were perceived through the exoskeleton, but they were also seen through the visualization system, thus increasing the feeling that the virtual fixture was actually present in the real environment.

In 1993, the *KARMA* (Knowledge-based Augmented Reality for Maintenance Assistance) project was presented by researchers Steven Feiner, Blair MacIntyre, and Doree Seligmann of Columbia University. *KARMA* was the first prototype of augmented reality system to support maintenance activities. It was a very ambitious project that used an HMD viewer and an innovative ultrasound tracking system in order to view the instructions for the maintenance of laser printers [30].

In 1994, Paul Milgram and Fumio Kishino introduced the concept of *virtual continuum* [31] which connected the real world with the virtual world by passing through augmented reality (AR) and augmented virtuality (AV) which represents the merging of real objects into the virtual world in which user interactions take place (Fig. 2.7).

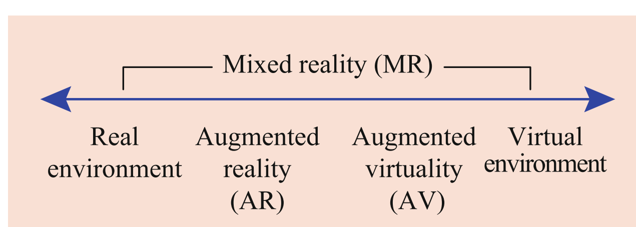


Fig. 2.7 Reality-virtuality continuum

In the same year, the AR made its debut in theater and in the medical field. Julie Martin produced the first theater show called *Dancing in Cyberspace* [32], funded by the Australia Council for the Arts, in which dancers and acrobats interacted with virtual objects projected onto the physical stage. In the medical field, Andrei State et al. from the University of North Carolina at Chapel Hill presented an AR application capable of letting physicians to observe a 3D fetus directly within a pregnant patient with see-through HMD [33]. This application showed some limitations of AR in real-time contexts, but it also demonstrated the importance of AR in medicine by paving the way for many other medical applications.

In 1996, after many years of research on AR, the Sony Computer Science Laboratories (CSL) presented the first AR system that used 2D markers called *CyberCode* [34]. It was introduced by Jun Rekimoto, professor and deputy director of the CSL, and became the model for future marker-based AR systems. *CyberCode* is a visual tagging system based on the recognition of objects and their tracking via 2D barcode using simple and inexpensive sensors. The tags used by *Cybercode* allowed to track the positioning and orientation of objects in the three dimensions thus allowing them to be used in many AR applications. In 2007, the Playstation 3 game *The Eye of Judgment* [35] used *CyberCode* tags printed on collectible cards to display a 3D model on the card image capable of adapting to changes in the movement and rotation of the card. AR officially entered in the entertainment world.

In 1997, researcher Ronald Azuma conducted the first AR survey introducing the most popular and still cited definition of AR. Azuma defined AR as a field in which 3D virtual objects are integrated into a 3D real environment in real time [1]. He put the following three characteristics that a system or experience must have to be qualified as AR: (1) *combines real and virtual* allowing real and virtual objects to coexist in the same environment; (2) *interactive in real time*, in order to eliminate from this classification animated films in which computer-generated animations are superimposed on real scenes; and (3) *registered in 3D*, requiring that the augmented contents must interact contextually with the scene. Hence, a distinction was made with applications that used HMDs that placed virtual information in the user's field of view even if this parameter excludes many applications that are normally considered to be in the domain of AR, such as Google Glass.

In 1998, AR made its debut during a game of the National Football League (NFL), thanks to the *1st & Ten* computer system [36] created by Sportvision. The AR was used to display the yellow first down line that a viewer sees during the match to make it easier for him to follow the play on the field. This type of increase, however, is in contradiction with the definition of AR proposed by R. Azuma for the lack of the second characteristic.

At this point in the history of AR, there are many scientific publications and applications used in various fields with excellent results. One of the biggest difficulties still to be overcome in the development of AR applications was to be able to track the gaze of users and understand where they were looking in order to integrate virtual images. It was then, in 1999, that the professor Hirokazu Kato from the Nara Institute of Science and Technology developed the first opensource software library called *ARToolKit*, which allowed the tracking video by calculating the exact orientation of the camera and that with respect to the physical markers in real time [37]. This library is still used today for the development of AR applications on various operating systems and desktop and mobile devices, including a plugin for the Unity3D graphics engine.

In the same year, NASA used a special AR dashboard for the first time to improve the navigation system of its X-38 spacecraft which was launched using the *Hybrid Synthetic Vision* system through which the map data was displayed directly on the screen of the pilot in the flight testing period.

In the 2000s, we saw exponential growth in mobile devices with high memory and increasingly powerful processors. Personal digital assistants (PDAs) and mobile phones, in particular, have been an attractive gateway to the virtual world since their processing capabilities have increased substantially, and an increasing number are GPS-enabled and can access to high speed data networks. This led to a significant boost in developing AR application. In the following are some of the many historical milestones on the AR.

In 2001, Gerhard Reitmayr and Dieter Schmalstieg created a mobile multi-user and collaborative augmented reality system that supported stereoscopic 3D graphics, a pen and pad interface and direct interaction with virtual objects [38]. The system exploited both advantages of mobile computing, computer supported collaborative work and AR technology, and was assembled with standardized hardware components, and it was the test bench for user interface experiments concerning the collaborative work supported by AR technology. It was composed by an advanced user interface management system that allows the user to work with different applications presenting 2D and 3D user interfaces at the same time and used *ARToolkit* [37] to process the video information in order to track the pen and pad within a personal workspace surrounding the user.

In 2002, the researcher Bruce H. Thomas presented *ARQuake* [39], the first outdoor augmented reality gaming system. With *ARQuake*, which can be considered the mobile augmented reality version of the famous shooter game Quake, the gaming experience is transferred outdoors. *ARQuake* used a low-cost, moderately accurate six-degree localization system based on GPS, a digital compass and a faithful vision-based monitoring. It is very similar to the

Pokémon GO game in which virtual creatures appear in the real world, but the user instead of capturing them must shoot them. It was an important step for the development of augmented reality which in a few years had a remarkable evolution.

In 2003, Daniel Wagner and Dieter Schmalstieg presented the first handheld AR system on a PDA with a commercial camera, thus paving the way for augmented reality on the smart devices that followed [40]. The proposed system provided the user with a three-dimensional augmented view of the environment, achieving good overlay registration accuracy by using the more and more popular marker-based tracking library *ARToolKit*, which ran directly on the PDA. The Wagner and Schmalstieg's system also introduced an optional client/server architecture that was based on wireless networking and was able to dynamically and transparently offload the tracking task in order to provide better performance in selected areas.

In 2004, Mathias Möhring introduced the first 3D marker tracking system on mobile devices, which became the first mass mobile AR application [41]. This system supported the detection and differentiation of different markers and corrected integration of rendered 3D graphics into the live video stream via a weak perspective projection camera model and an OpenGL [42] rendering pipeline.

In 2005, Anders Henrysson ported the *ARToolKit* to the Symbian operating system so allowing developers to build AR applications that run on a mobile phone. Thanks to this porting, he created *AR Tennis*, the first collaborative AR game running on a mobile phone [43]. In this application, two players sit across a table from each other with a piece of paper between them with a set of *ARToolKit* markers drawn on it.

In 2006, Nokia presented the *MARA* (Mobile Augmented Reality Application) project, a multi-sensor mobile phone AR guidance application for mobile phones which allowed to identify the objects taken from the target of the mobile phone and therefore was able to add information about them [44]. The *MARA* system exploited a GPS, an accelerometer, and a compass as hardware, and so the mobile phone thus enhanced proved to be able to identify restaurants and points of interest and to provide information such as links on the Internet, regarding what was displayed on the screen. The system could also be used to locate friends nearby, as long as they were equipped with cell phones with GPS and adequate software. When the video mode was started, *MARA* combined the various information from the three sensors to obtain the position and orientation of the phone.

In 2008, almost simultaneously to the first Android device, the Mobilizy company launched the Wikitude World Browser with AR that is an application that combines GPS and compass data with Wikipedia entries and overlays information on the real-time camera view of an Android smartphone [45]. It is the first world augmented reality browser. In 2009, SPRXmobile launched another AR browser, Layar [46], that is an advanced variant of Wikitude. Layar, which is currently available, combines GPS, camera, and compass to identify the surrounding environment and overlay information on the screen in real time.

Afterwards, there was an explosion of devices for AR and mixed reality that allowed industries to design and develop increasingly complex AR applications. Nowadays, in the era of Industry 4.0, of which AR is an enabling technology, many technological innovation projects are using AR to improve production processes. To support these developments, in 2013 Google launched a public beta test of Google Glass, glasses for enjoying AR contents through the use of the Bluetooth connection to the user's smartphone and capable of actively responding to voice commands and hand gestures. In 2016, Microsoft starts selling its version of smart glasses for the AR/MR called HoloLens, which is much more advanced and expensive than Google Glass.

2017 saw a gradual standardization of development tools and democratization of access to the creation of AR apps with the introduction of Apple's ARKIT SDK and Google's ARCore. This has resulted in the enabling of double the number of mobile devices for AR applications and triple the number of active users for these applications. The improvements made to these SDKs in 2018 and 2019 have progressively introduced new features capable of bringing the level of AR experiences possible on smartphones or tablets closer to that possible with dedicated see-through viewers. At the same time, the number of ARKit- or ARCore-enabled mobile devices has reached a significant percentage of the total number of smartphones and tablets in circulation, representing an important prerequisite for the spread of AR applications globally.

2.4 Current AR Technologies and Applications

Nowadays, AR is achieved through a variety of technological innovations that can be implemented alone or together. Today, an AR system is composed by *general hardware components* such as a processor, a display, sensors, and input devices; a *special display* such as head-mounted displays (HMD), Heads-Up Displays (HUD), virtual retinal displays, Spatial Augmented Reality (SAR), displays of mobile devices; *sensors and input devices* such as GPS, gyroscopes, accelerometers, compasses, RFID, wireless sensors, tactile and voice recognition; and finally a *software*.

The following is a short list of some of the most representative AR headsets currently available on the market.

Microsoft Hololens 2 is an optical see-through smart glass for immersive augmented reality, built around specifically developed components enabling holographic processing. It is equipped with an Intel Atom processing unit that develops a computing power comparable to that of a medium class laptop, passively cooled without fans and powered by an internal rechargeable battery that typically allows 2–3 h of active use. The coprocessor developed according to Microsoft specifications, called Holographic Processing Unit (HPU) is able to efficiently process information from all available sensors in real time. These include a three-axis inertial sensor, four high frame-rate environmental cameras, a depth camera, an ambient light sensor, and a finger recognition unit (mixed reality capture). This architecture allows for 6 DOF inside-out tracking, combined with the 3D reconstruction of the surrounding environment through proprietary shape-from-motion algorithms and interaction with the hands. It is also equipped with a micro-controller to support more complex interactions that require a tangible interface.

Magic Leap ONE is another optical see-through device, essentially consisting of three elements: one is the wearable headset including the display unit through which to look (called Lightwear), the second is a circular-shaped belt-computer (called Lightpack), and the third is a controller to be held in hand for interaction with the contents. The Lightwear includes a multiple sensors system for inside-out optical tracking and a special optic called Photonics Chip in charge of compositing virtual images onto the real world. The headset is also equipped with an eye-tracking system necessary to enable the near-far plane focusing capability, characteristic of this device. Processing is entrusted to an NVIDIA Tegra X2 chipset flanked by 8 GB of RAM, 128 GB of internal memory, and a rechargeable battery capable of ensuring up to 3 h of continuous operation.

Varjo XR-3 is an advanced video see-through headset featuring foveated rendering and display, exploiting high-precision embedded eye trackers to drive the foveal high pixel density display according to where visual focus is. The device is designed to enable switching between mixed reality and full VR scenes by means of a photorealistic pass-through capability, so that it can display simulated objects in the real world, and real objects in a simulated world. Pass-through video is powered by two 12-megapixel cameras operating at 90 Hz, and an ultra-low latency image pipeline operates at less than 20 ms. The headset has a FOV of 87 degrees and is compatible with SteamVR SDK and ART tracking systems.

About the main softwares used for AR, there are a lot of specialized AR software libraries and SDK, often also responsible for the tracking task, the most diffused of which are as follows:

AR ToolKit, a free, multi-platform, and open-source SDK that was one of the first augmented reality SDKs for mobile devices. Key features include camera position orientation and tracking, tracking of planar images, camera calibration, and square marker tracking.

ARKit, designed exclusively for iOS devices like the iPhone and iPad. Key features include space recognition using SLAM, object detection, light estimation, and a continuing AR experience through several sessions.

ARCore, open source SDK developed by Google for both the Android and iOS platforms. It is optimized to exploit existing smartphones/tablet sensors and processing capabilities. Key features include motion tracking, environmental understanding, light estimation, user interaction, oriented points, anchors and trackables, augmented images, and sharing and provide AR functionalities to Unity3D and Unreal graphic engines.

Amazon Sumerian, aimed at creating and running VR, AR, and 3D applications without needing any expertise in programming or 3D graphics. The SDK is a part of Amazon Web Services (AWS) and works with Android and iOS mobile devices as well as Oculus Go, Oculus Rift, HTC Vive, HTC Vive Pro, Google Daydream, and Lenovo Mirage.

Visual Studio is an integrated development environment for Android, iOS, Windows, web, and cloud. It can also be used with the HoloLens and Hololens 2, Microsoft's untethered mixed reality headsets.

Vuforia is one of the most diffused AR development platforms. It provides robust tracking capabilities and allows developers to orient and place virtual objects into the real world. The technology allows developers to track objects in relation to real-world images, both in marker-based and markerless modality. It supports Android, iOS, UWP, and Unity and also works with Microsoft HoloLens 2 and can be integrated in Unity3D graphic engine.

Wikitude AR SDK, targeted to smartphones, tablets, and smart glasses. It includes object and scene recognition using Simultaneous Location and Mapping, instant tracking allowing placement of a virtual object in real space, the ability to identify about 1,000 images and more than one of them at the same time, cloud recognition, and 3D model rendering.

2.4.1 Main Application Fields

The application domains of augmented reality are manifold and for the most part have been developed in recent years. The areas most affected by this new technology include many industrial sectors, entertainment, training, health, forensic, and military applications, to name a few of the most significant.

Assembling, Maintenance, and Repair. Using an AR headset, a technician repairing an engine can see images and information superimposed on its field of view. A maintenance

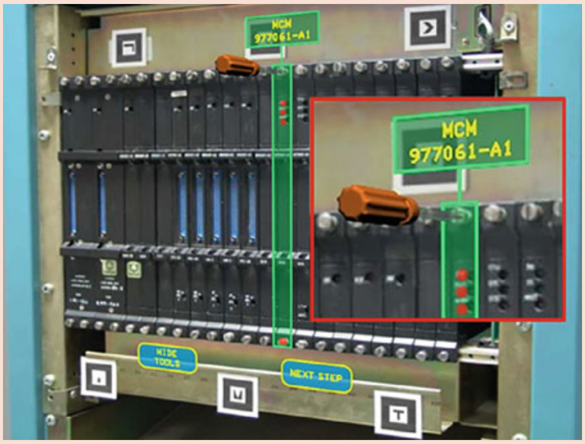


Fig. 2.8 AR assisted assembling in automotive industry. (Courtesy of Leonardo S.p.A and VRLab of University of Salerno)

Fig. 2.9 Archaelogical site augmented by virtual artifacts and characters. (Courtesy of VRLab, University of Salerno)

procedure could be presented in a corner, and an animated virtual model of the necessary tool can illustrate the exact action to be performed. The AR system can label all the important parts. Complex procedural repairs can be divided into a series of simple steps. Simulations can be used to train technicians, which can significantly reduce training costs [47] (Fig. 2.8).

Advertising. Smartphone-based AR apps can augment the surrounding environment by displaying real-time digital information in conjunction with the real world. Using the embedded cameras and GPS location feature, this kind of applications can retrieve data based on where the user is and displays this data to him on his mobile screen. Details about popular places, structures, and movies can be provided, such as street views showing the names of the restaurants and businesses, superimposed over their storefronts.

Sightseeing and Cultural Heritage Visiting. There are plenty of AR applications for the tourism and cultural heritage fields. The ability to augment a live view of exhibits in a museum with facts and graphics represents an obvious use of this technology [48]. Also sightseeing can be improved by augmented reality. Using a smartphone or a tablet, tourists can stroll through historical sites and see additional information presented as text and graphics superimposed on the live screen. These applications use GPS technology and image recognition technology to search for data from an online database. In addition to information about a historical site, there are applications that look back in history and possibly show how the position looked 10, 50, or even 100 years ago (Fig. 2.9).

Military Applications. The Heads-Up Display (HUD) represents the first example of augmented reality technology applied to the military field. A transparent display is positioned directly in the fighter pilot's view. Data typically displayed to the pilot includes altitude, airspeed, and the horizon line in addition to other critical data. The term heads-up name applies because the pilot doesn't have to look down at the aircraft's instrumentation to get the data he needs. The Head-Mounted Display (HMD) can be also used by ground troops. Critical data such as enemy location can be presented to the soldier within their line of sight. This technology is also used for simulations for training purposes.

Medical Uses. Medical students use AR technology to practice surgery in a controlled environment. Visualizations aid in explaining complex medical conditions to patients. Augmented reality can reduce the risk of an operation by giving the surgeon improved sensory perception [49]. This technology can be combined with MRI or X-ray systems and bring everything into a single view for the surgeon. Neurosurgery is at the forefront when it comes to surgical applications of augmented reality. The ability to image the brain in 3D on top of the patient's actual anatomy is powerful for the surgeon [50]. Since the brain is somewhat fixed compared to other parts of the body, the registration of exact coordinates can be achieved. Concern still exists surrounding the movement of tissue during surgery. This can affect the exact positioning required for augmented reality to work (Fig. 2.10).

Gaming. With recent advances in computing power and technology, gaming applications in augmented reality are on the upswing. Head-worn systems are affordable now, and computing power is more portable than ever. Before you can say Pokemon Go, you can jump into an AR game that works with your mobile device, superimposing mythical creatures over your everyday landscape. Popular Android and iOS

Fig. 2.10 MR hand-based interaction with medical content. (Courtesy of VRLab, University of Salerno)

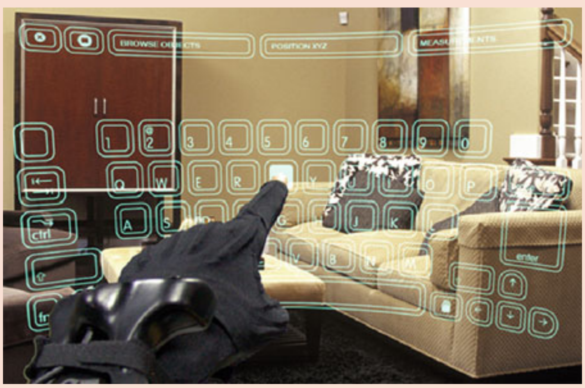


Fig. 2.11 Gesture-based interface to AR-enhanced crime scene analysis. (Courtesy of VRLab, University of Salerno)

AR apps include Ingress, SpecTrek, Temple Treasure Hunt, Ghost Snap AR, Zombies, Run!, and AR Invaders.

Forensics. Modern investigation techniques often resort to capturing the crime scene through 3D scanning and subsequent analysis of the same by means of measurement software tools that exploit conventional mouse and keyboard-based interaction techniques. In this context, the joint use of augmented reality and gestural interaction techniques allows a new investigation methodology. The idea is to make it possible for the investigator to explore the crime scene by allowing him to view the objects relevant for the formulation or verification of the hypotheses directly in the surrounding environment, interacting with them in a natural way (i.e., without using a mouse and keyboard) through hand gestures [60] (Fig. 2.11).

Navigation. Navigation applications are possibly the most natural fit of augmented reality with our everyday lives. Enhanced GPS systems use augmented reality to make it easier to get from point A to point B. Using a Heads-Up Display in combination with the GPS, users see the selected route over the windshield or an augmented view of the actual road surrounding the vehicle through a video-see through cockpit.

2.5 Emerging Trends and Open Challenges for AR

As mentioned in the introduction to this chapter, over the years, the concept of AR as a subset of MR according to the paradigm proposed by Milgram and Kishino has been rethought and extended, to encompass a much broader range of aspects than those originally considered in the monodimensional real-versus-virtual continuum. According to Steve Mann, *mediated Reality is created when virtual, or computer generated, information is mixed with what the user would otherwise normally see* [51]. A device enabling to absorb, quantify, and resynthesize the light the user would normally see is suitable for creating a "mediated reality" [52]. When the light is resynthesized under computer control, information can be added or removed from the scene [53] before it is presented to the user [54]. The (not technically trivial) possibility of removing information from the scene observed leads to the so-called Diminished Reality that allows to selectively remove elements from the visual field, possibly replacing them with synthesized objects if used in conjunction to augmented reality. Augmentation has also been extended to other senses than just vision, thus enabling haptic, olfactory, and auditory augmentation and leading to "phenomenological augmented reality" [55], that is a physics-based AR [56] on which the "natural" user-interfaces are also based [57].

The superset including the full Milgram's spectrum ranging from the "complete real" to the "complete virtual" and encompassing all human-machine interactions generated by computer technology and wearables is referred as eXtended Reality (XR), where the "X" may be seen as a variable representing any current or future spatial computing technologies [58].

However, the growth of the AR since its first appearance has always been guided by technological evolution without which the resources necessary for the realization of a usable AR system would not have been obtained. It is expected that wearable technologies, smart glasses, holograms, mobile applications, and contact lenses will become increasingly popular. However, more research is needed to explore the potential of AR in the fields of military, civil industrial application.

The interaction between people and virtual objects will become increasingly simple, thanks to innovations in the

ICT field. In the educational field, content can be enjoyed on mobile devices and enhanced, thanks to the AR that will allow students to interact with the displayed virtual objects [59]. 3D graphics, imaging computing, and human-machine interaction systems will allow the introduction of AR in robotics but also in fields such as nanotechnology and biomimetics [61]. Sensory technologies and smart devices are expected to become more widespread in the coming years, and AR will be increasingly correlated with the Internet of things [61] and artificial intelligence. Eventually, social network applications such as Facebook, Instagram, and Twitter are expected to be fully presented with AR. It is already known that Facebook has started to invest in this topic. However, the main research and applicative efforts to foster a mass diffusion of AR and MR will surely include three main goals: wearable devices enhancement, the use of Artificial Intelligence (AI) and Machine Learning (ML), and the development of a tele-presence capability for any communication scenario.

New Generation of AR Devices. AR wearable devices are currently in a phase of relative technological maturity, starting to offer sufficient performance and operational features to appreciate the potential of this technology, but will not be able to fully exploit its potential for a few more years. Nevertheless, some application proposals in the military, medical, and business fields have already begun to concretely demonstrate the contribution that AR systems can provide. One of the main factors limiting the spread of these devices and AR systems in general is represented by the autonomy guaranteed by the batteries, which can become a serious bottleneck especially for "mission critical" applications. In this regard, the new generation of AR viewers (starting with the new Microsoft Hololens 2) promises extended autonomy and, at the same time, superior processing performance combined with improved comfort of use, thanks to reduced weight and increased field of vision. These hardware improvements are fundamental to broaden and consolidate the application areas of AR, as reliability, accuracy, and consistency in the visualization of additional information associated with the real operating environment are required in many fields, failing which the full utility of the application would fail. When these critical issues are resolved, the number of users of AR systems will realistically increase exponentially. According to the most recent estimates, by 2025 over ten million workers could regularly wear smart glasses during their work to obtain context-based information and visual or textual aids for the optimal execution of their work routine. Regular use of these devices and their applications are expected to improve prototyping, testing, troubleshooting, and quality control as workers will be able to compare real-world elements with documentation and specifications on the fly available. Integration with other recent trends in workplace training, in particular, gamification, can in fact increase the effectiveness of AR and MR solutions, with advantages also on learning times and on the retention of information learned.

AI/ML for Enhanced AR/MR Experience. In the last 10 years, there has been an impressive growth of artificial intelligence methods based on machine learning and of the related enabling technologies, which has produced a very strong increase in their diffusion in a large number of application sectors. The combination of these techniques with augmented and mixed reality is particularly useful and immediate, considering that computer vision represents one of the main research topics for which AI and ML have been developed. The combination of image understanding (offered by AI) and co-registration between the surrounding environment and virtual content (offered by AR) can enable application scenarios of great technological, economic, and social impact. As an example, just consider that up to 35% of Amazon's sales derive from its recommendation engine which uses machine learning and artificial intelligence techniques to associate advertisers and their products with customers and their preferences derived from their habits: navigation on the company's website. This model transported to the real world, rather than to the intangibility of web catalogs, can generate a disruptive effect on how to purchase products of all kinds while the user is out shopping. Additional information of any kind (e.g., on prices, features, or offers in progress) may be available simply by observing an article (through pattern recognition and gaze analysis techniques) or even by formulating a request in natural language (through a dedicated chatbot based on natural language processing techniques). According to this paradigm, the answers can also be adapted to the customer's profile, resulting from the analysis of their purchasing habits, allowing greater personalization of the results.

AR-Enabled Tele-presence. The recent COVID-19 pandemic has accelerated a process of spreading remote work that has already started and which will progressively lead 25–30% of the workforce to work from home for several days a week by the end of 2022. The logistical and safety advantages that this model can guarantee and the limits of this collaboration modality related to the lack of a direct personal presence have also emerged. The role that augmented reality can play in this field, thanks to its characteristic of allowing an effective tele-presence experience in which all connected subjects can be transferred to a common environment, is potentially enormous. Global corporations such as Microsoft and CISCO Systems are both actively working on the creation of tele-presence environments based on "holographic" representations of the subjects connected remotely, and although there are still no commercial applications ripe for large-scale diffusion, the scenario of near future is clearly identified. According to this trend, shared AR environments will also be a natural choice for training sessions, conferences, and education. Finally, in the world of video games, collaborative augmented reality applications are perceived by users as extremely engaging and in the next few years will represent a growing slice of the electronic entertainment market, also crossing home borders to operate outdoors.

2.6 Conclusions

Although the concept of AR was formulated in the 1960s, only recently AR has begun to show its real potential as AR-related technologies have constantly progressed, leading to more mature solutions for a wide range of applicative fields. Low-cost solutions, exploiting the potential of mobile devices as ubiquitous AR platform products, have been introduced and are rapidly spreading across many markets. Nevertheless, a number of limitations, particularly concerning the wearable devices, are still to be addressed and currently detract from the mass diffusion of this potentially disruptive technology. According to Azuma [62], among the most relevant challenges to be faced for a full development of the AR potential, there is a need for a more accurate and robust tracking in both indoor and outdoor large environments, a much wider field of view in optical or video see-through visors, a consistent and reliable hands-free or gesture-based interface, and object recognition capability to improve the seamless incorporation of virtual and real objects within the augmented field of view. However, the AR paradigm of conveying any type of information into the user's perception field, according to a context- or spatial-based criterion, has already shown a pervasive applicability to almost any domain, from scientific visualization to big data comprehension and analysis, to advanced communications, to informed shopping, and new forms of utilization will arguably appear in the next years.

Acknowledgments This work and the research behind it would not have been possible without the non-profit support of Raffaele Vertucci and Salvatore D'Onofrio who shared their knowledge and experiences on the VR/AR technologies gained in their company Leonardo Spa, a leading company in the aerospace industry.

References

1. Azuma, R.T.: A survey of augmented reality (1997). MIT Press, In Presence: Teleoperators and Virtual Environments 6, 4, 355–385 (1997)
2. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality (2015)
3. Antigny, N., Servières, M., Renaudin, V.: Pedestrian track estimation with handheld monocular camera and inertial-magnetic sensor for urban augmented reality. In: 2017 international conference on indoor positioning and indoor navigation (IPIN), pp. 1–8. IEEE (2017)
4. Bermúdez, G.S.C., Karnaushenko, D.D., Karnaushenko, D., Lebanov, A., Bischoff, L., Kaltenbrunner, M., … Makarov, D.: Magnetosensitive e-skins with directional perception for augmented reality. Sci. Adv. **4**(1), eaao2623 (2018)
5. Ramamonjy, A., Bavu, E., Garcia, A., Hengy, S.: A distributed network of compact microphone arrays for drone detection and tracking. J. Acoust. Soc. Am. **141**(5), 3651–3651 (2017)
6. Yang, X., Guo, J., Xue, T., Cheng, K.T.T.: Robust and real-time pose tracking for augmented reality on mobile devices. Multimed. Tools Appl. **77**(6), 6607–6628 (2018)
7. Li, P., Qin, T., Hu, B., Zhu, F., Shen, S.: Monocular visual-inertial state estimation for mobile augmented reality. In: 2017 IEEE international symposium on mixed and augmented reality (ISMAR), pp. 11–21. IEEE (2017)
8. Bresson, G., Alsayed, Z., Yu, L., Glaser, S.: Simultaneous localization and mapping: a survey of current trends in autonomous driving. IEEE Trans. Intell. Veh. **2**(3), 194–220 (2017)
9. Pire, T., Fischer, T., Castro, G., De Cristóforis, P., Civera, J., Berlles, J.J.: S-PTAM: stereo parallel tracking and mapping. Robot. Auton. Syst. **93**, 27–42 (2017)
10. Xiao, L., Wang, J., Qiu, X., Rong, Z., Zou, X.: Dynamic-SLAM: semantic monocular visual localization and mapping based on deep learning in dynamic environment. Robot. Auton. Syst. **117**, 1–16 (2019)
11. Liu, Q., Wang, Z., Wang, H.: SD-VIS: a fast and accurate semi-direct monocular visual-inertial simultaneous localization and mapping (SLAM). Sensors. **20**(5), 1511 (2020)
12. Pombo, L., Marques, M.M.: Marker-based augmented reality application for mobile learning in an urban park: steps to make it real under the EduPARK project. In: 2017 international symposium on computers in education (SIIE), pp. 1–5. IEEE (2017)
13. Sadeghi-Niaraki, A., Choi, S.M.: A survey of marker-less tracking and registration techniques for Health & Environmental Applications to augmented reality and ubiquitous geospatial information systems. Sensors. **20**(10), 2997 (2020)
14. Ma, L., Zhao, Z., Zhang, B., Jiang, W., Fu, L., Zhang, X., Liao, H.: Three-dimensional augmented reality surgical navigation with hybrid optical and electromagnetic tracking for distal intramedullary nail interlocking. Int. J. Med. Rob. Comput. Assisted Surg. **14**(4), e1909 (2018)
15. Rathinavel, K., Wetzstein, G., Fuchs, H.: Varifocal occlusion-capable optical see-through augmented reality display based on focus-tunable optics. IEEE Trans. Vis. Comput. Graph. **25**(11), 3125–3134 (2019)
16. Krajancich, B., Padmanaban, N., Wetzstein, G.: Factored occlusion: single spatial light modulator occlusion-capable optical see-through augmented reality display. IEEE Trans. Vis. Comput. Graph. **26**(5), 1871–1879 (2020)
17. Cattari, N., Cutolo, F., D'amato, R., Fontana, U., Ferrari, V.: Toed-in vs parallel displays in video see-through head-mounted displays for close-up view. IEEE Access. **7**, 159698–159711 (2019)
18. Büttner, S., Prilla, M., Röcker, C.: Augmented reality training for industrial assembly work-are projection-based AR assistive systems an appropriate tool for assembly training? In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–12 (2020)
19. Baum, L.F.: The Master Key, ed. by CreateSpace Independent Publishing Platform (2016)
20. Heilig, M.: Sensorama Simulator, US Patent 3,050,870 (1962)
21. Rheingold, H.: Virtual Reality. Summit Books (1991)
22. Sutherland, I.E.: Sketchpad: A man-machine graphical communication system. In: Proceedings of the AFIPS spring joint computer conference, vol. 23, (1963)
23. Sutherland, I.E.: The ultimate display. In: Proceedings of the IFIPS congress, vol. 2, (1965)
24. Furness, T.A.: The super cockpit and its human factors challenges. In: Proceedings of the human factors society, 30th annual meeting (1986)
25. Myron, W.: Krueger: Responsive environments. In Proceedings of the national computer conference (1977)

26. EyeTap Personal Imaging Lab. http://www.eyetap.org
27. Caudell, T.P., Mizell, D.W.: Augmented reality: An application of heads-up display technology to manual manufacturing processes. In: Proceedings of 1992 IEEE Hawaii international conference on systems sciences. IEEE Press (1992)
28. Rosenberg, L.B.: The use of virtual fixtures as perceptual overlays to enhance operator performance in remote environments, Tech. Rep. (1992 DTIC Document)
29. Rosenberg, L.B.: Virtual fixtures: Perceptual tools for telerobotic manipulation. In: Virtual reality annual international symposium. IEEE (1993)
30. Feiner, S., Macintyre, B., Seligmann, D.: Knowledge-based augmented reality. In Commun. ACM. **36**(7), 53–62 (1993)
31. Milgram, P., Kishino, F.: A Taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **E77-D**(12(12)), 1321–1329 (1994)
32. Wikipedia: Augmented Reality. https://en.wikipedia.org/wiki/Augmented_reality
33. State, A., Chen, D.T., Tector, C., Brandt, A., Chen, H., Ohbuchi, R., Bajura, M., Fuchs, H.: Observing a volume rendered fetus within a pregnant patient, 364–368, CP41 (1994). https://doi.org/10.1109/VISUAL.1994.346295
34. Rekimoto, J., Ayatsuka, Y.: CyberCode: Designing augmented reality environments with visual tags. In: Proceedings of the 2000 ACM conference on designing augmented reality environments. DARE (2000)
35. Wikipedia: The Eye of Judgment. https://en.wikipedia.org/wiki/The_Eye_of_Judgment
36. Wikipedia: 1st & Ten (graphics system). https://en.wikipedia.org/wiki/1st_%26_Ten_(graphics_system)
37. Kato, H., Billinghurst, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR 99) (1999)
38. Reitmayr, G., Schmalstieg, D.: Mobile Collaborative Augmented Reality. In: Proceedings of ISAR'01, New York (2001)
39. Thomas, B.H., Close, B., Donoghue, J., Squires, J., Bondi, P., Piekarski, W.: First person indoor/outdoor augmented reality application: ARQuake. Per. Ubiquit. Comput. **6**, 75–86 (2002). https://doi.org/10.1007/s007790200007
40. Wagner, D., Schmalstieg, D.: First steps towards handheld augmented reality, 127–135 (2005). https://doi.org/10.1109/ISWC.2003.1241402
41. Moehring, M., Lessig, C., Bimber, O.: Video see-through and optical tracking with consumer cell phones. (2004). https://doi.org/10.1145/1186223.1186357
42. Wikipedia: OpenGL. https://en.wikipedia.org/wiki/OpenGL
43. Henrysson, A., Billinghurst, M., Ollila, M.: AR tennis (2006). http://doi.acm.org/10.1145/1179849.1179865. https://doi.org/10.1145/1179133.1179135
44. Greene, K.: Hyperlinking reality via phones (2006). https://www.technologyreview.com/s/406899/hyperlinking-reality-via-phones/
45. Wikitude World Browser.: https://www.wikitude.com/wikitude-world-browser-augmented-reality/
46. Layar.: https://www.layar.com/
47. Abate, A.F., Guida, M., Leoncini, M.P., Nappi, M., Ricciardi, S.: A haptic-based approach to virtual training for aerospace industry. J. Vis. Lang. Comput. **20**(5), 318–325 (2009)
48. Abate, A.F., Acampora, G., Ricciardi, S.: An interactive virtual guide for the AR based visit of archaeological sites. J. Vis. Lang. Comput. **22**(6), 415–425 (2011)
49. Abate, A.F., Acampora, G., Loia, V., Ricciardi, S., Vasilakos, A.V.: A pervasive visual–haptic framework for virtual delivery training. IEEE Trans. Inf. Technol. Biomed. **14**(2), 326–334 (2010)
50. De Marsico, M., Levialdi, S., Nappi, M., Ricciardi, S.: FIGI: floating interface for gesture-based interaction. J. Ambient. Intell. Humaniz. Comput. **5**(4), 511–524 (2014)
51. Mann, S., Niedzviecki, H.: Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer. Doubleday Canada (2001)
52. Mann, S., Barfield, W.: Introduction to mediated reality. Int. J. Hum. Comput. Interaction. **15**(2), 205–208 (2003)
53. Grasset, R., Gascuel, J.D.: Interactive mediated reality. In: The second IEEE and ACM international symposium on mixed and augmented reality, 2003. Proceedings, pp. 302–303. IEEE (2003)
54. Poelman, R., Akman, O., Lukosch, S., Jonker, P.: As if being there: mediated reality for crime scene investigation. In: Proceedings of the ACM 2012 conference on computer supported cooperative work, pp. 1267–1276 (2012)
55. Barbatsis, G., Fegan, M., Hansen, K.: The performance of cyberspace: an exploration into computer-mediated reality. J. Comput.-Mediat. Commun. **5**(1), JCMC512 (1999)
56. Scourboutakos, P., Lu, M.H., Nerker, S., Mann, S.: Phenomenologically augmented reality with new wearable led sequential wave imprinting machines. In: Proceedings of the eleventh international conference on tangible, embedded, and embodied interaction, pp. 751–755 (2017)
57. Hu, A.Z., Janzen, R., Lu, M.H., Mann, S.: Liquid jets as logic-computing fluid-user-interfaces. In: Proceedings of the on thematic workshops of ACM multimedia 2017, pp. 469–476 (2017)
58. Mann, S., Havens, J. C., Iorio, J., Yuan, Y., Furness, T.: All reality: values, taxonomy, and continuum, for virtual, augmented, eXtended/MiXed (X), Mediated (X, Y), and Multimediated reality/intelligence. Presented at the AWE 2018 (2018)
59. Bower, M., Howe, C., McCredie, N., Robinson, A., Grover, D.: Augmented reality in Education – Cases, places, and potentials. Educ. Media Int. **51** (2014). https://doi.org/10.1080/09523987.2014.889400
60. Narducci, F., Ricciardi, S., Vertucci, R.: Enabling consistent hand-based interaction in mixed reality by occlusions handling, to be published on multimedia tools and applications. Springer
61. Kipper, G., Rampolla, J.: Augmented reality: An emerging technologies guide to AR, pp. 1–158. Elsevier (2012)
62. Azuma, R.T.: The most important challenge facing augmented reality. Presence Teleop. Virt. **25**(3), 234–238 (2016)

Raffaele Vertucci Graduated with honors in Electronic Engineering, Raffaele Vertucci gained his experience working in the aerospace industries. As CTO he has been involved in research and development activities related to image processing, computer vision and virtual/augmented reality technologies, with particular focus on their application for support, maintenance and training, writing articles, managing research programs, making dissemination activities, registering a patent and developing software tools.

Salvatore D'Onofrio received the PhD in Computer Science from the University of Sannio in 2007. He works now as CTO member at Leonardo company in Italy. He is mainly involved in R&D activities focused on Virtual/Augmented Reality technologies applied to training and maintenance, Modeling & Simulation and Predictive Analysis for maintenance activities.

Stefano Ricciardi received the PhD in Sciences and Technologies of Information, Complex Systems and Environment from the University of Salerno. Main research interests include biometry, virtual and augmented/mixed reality, haptics systems and human-computer interaction. He is currently a research fellow at the University of Molise, member of IEEE and GIRPR/IAPR. and co-author of more than eighty peer reviewed research papers.

Maurizio De Nino Graduated with honors in physics with a specialization in artificial intelligence and data acquisition system, for 20 years he did his experience working for scientific centre and hitech company with a specific experience in R&D for real-time embedded system, image processing and virtual reality. In the last 10 years, he is working with Digitalcomoedia as CTO for virtual and augmented reality system development.

Part II

Principles and Fundamentals, Software Techniques, and Developments

3 Principles of Object Tracking and Mapping

Jason Rambach, Alain Pagani, and Didier Stricker

Contents

3.1 Pose Estimation and Tracking Fundamentals 53
3.1.1 Notation: Problem Formulation 54
3.1.2 Cameras 54
3.1.3 Inertial Sensors 57
3.1.4 Depth Sensors 58
3.1.5 Geospatial Navigation Sensors 59

3.2 Computer Vision Techniques 59
3.2.1 Feature Matching 60
3.2.2 Feature Tracking 60
3.2.3 Pose Estimation from Feature Correspondences 62
3.2.4 The Epipolar Constraint 63
3.2.5 Line Tracking 64
3.2.6 Direct Image Alignment 64
3.2.7 Structure from Motion 65
3.2.8 Depth Image Tracking/ICP 66
3.2.9 Deep Learning in Computer Vision 66

3.3 Model-Based Tracking 69
3.3.1 Marker Trackers 69
3.3.2 3D Model Trackers 70
3.3.3 Nonrigid and Articulated Objects 74

3.4 SLAM 74
3.4.1 Visual SLAM 75
3.4.2 Visual-Inertial SLAM 77
3.4.3 RGB-D SLAM 78
3.4.4 Deep Learning for SLAM 79

3.5 Conclusion 80

References 80

Abstract

Tracking is the main enabling technology for Augmented Reality (AR) as it allows realistic placement of virtual content in the real world. In this chapter, we discuss the most important aspects of tracking for AR while reviewing existing systems that shaped the field over the past years. Initially, we provide a notation for the description of 6 Degree of Freedom (6DoF) poses and camera models. Subsequently, we describe fundamental computer vision techniques that tracking systems frequently use such as feature matching and tracking or pose estimation. We divide the description of tracking approaches into model-based approaches and Simultaneous Localization and Mapping (SLAM) approaches. Model-based approaches use a synthetic representation of an object as a template in order to match the real object. This matching can use texture or lines as tracking features in order to establish correspondences from the models to the image, whereas machine learning approaches for direct pose estimation of an object from an input image have also been recently introduced. Currently, an upcoming challenge is the extension of tracking systems for AR from rigid objects to articulated and nonrigid objects. SLAM tracking systems do not require any models as a reference as they can simultaneously track and map their environment. We discuss keypoint-based, direct, and semi-direct purely visual SLAM system approaches. Next, we analyze the use of additional sensors that can support tracking such as visual-inertial sensor fusion techniques or depth sensing. Finally, we also look at the use of machine learning techniques and especially the use of deep neural networks in conjunction with traditional computer vision approaches for SLAM.

Keywords

Tracking · Augmented reality · Object tracking · SLAM · Computer vision · Machine learning · Cameras · Sensors · Pose estimation · Mapping

3.1 Pose Estimation and Tracking Fundamentals

In this section we introduce the notation that is used throughout the chapter and provide a description of commonly used camera models as well as camera extrinsic and intrinsic

J. Rambach (✉) · A. Pagani · D. Stricker
Augmented Vision, German Research Center for Artificial Intelligence, Kaiserslautern, Germany
e-mail: Jason.Rambach@dfki.de; alain.pagani@dfki.de; didier.stricker@dfki.de

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_3

parameters. We present the formulation that leads to the mapping of 3D information in the world to pixel locations in the camera image. In addition to red green blue (RGB) cameras, we provide an initial description of the output of other sensors that are commonly used for AR tracking, such as inertial sensors, various depth sensors, and geospatial positioning sensors.

3.1.1 Notation: Problem Formulation

The considered problem of estimating or tracking the 6DoF camera poses mainly consists of estimating a transformation between two coordinate systems. Typically, the two coordinate systems involved are the coordinate system of the camera (or the tracking device in general) and that of the tracking target, which can be a marker or specific object or any coordinate system defined for the purpose of the AR application at hand.

Coordinate System Transformations

A point $\boldsymbol{m}$ in a coordinate system A is given by a three-dimensional vector $\boldsymbol{m}_a = [x_a, y_a, z_a]^\intercal$. The 6DoF transform converting the point from coordinate system A to a different coordinate system B consists of a rotation matrix $\boldsymbol{R}_{ba} \in SO(3)$ and a translation vector $\boldsymbol{b}_a \in \mathbb{R}^3$. The rotation matrix $\boldsymbol{R}_{ba}$ describes the rotation from coordinate system A to B, whereas $\boldsymbol{b}_a$ is the origin of coordinate system B expressed in coordinate system A. The transformation of point $\boldsymbol{m}$ from coordinate system A to B is then given as

$$\boldsymbol{m}_b = \boldsymbol{R}_{ba}(\boldsymbol{m}_a - \boldsymbol{b}_a), \tag{3.1}$$

and the inverse transformation from B to A is given as

$$\boldsymbol{m}_a = \boldsymbol{R}_{ab}(\boldsymbol{m}_b - \boldsymbol{a}_b), \tag{3.2}$$

where

$$\boldsymbol{R}_{ab} = \boldsymbol{R}_{ba}^{-1} = \boldsymbol{R}_{ba}^\intercal \quad \text{and} \quad \boldsymbol{a}_b = -\boldsymbol{R}_{ba}\boldsymbol{b}_a. \tag{3.3}$$

Equations (3.3) and (3.1) can alternatively be written as

$$\boldsymbol{m}_b = \boldsymbol{R}_{ba}\boldsymbol{m}_a + \boldsymbol{a}_b. \tag{3.4}$$

6DoF Pose of Rigid Bodies

The pose of any 3D rigid object in space consists of its position and orientation and therefore has 6DoF. A coordinate system O attached to the object is defined, and the ***object pose*** is given by the transformation of this coordinate system to another well-defined coordinate system C. Thus, the pose of the object in C can be given by $\boldsymbol{R}_{oc}, \boldsymbol{o}_c$ such that for any point m

$$\boldsymbol{m}_o = \boldsymbol{R}_{oc}(\boldsymbol{m}_c - \boldsymbol{o}_c). \tag{3.5}$$

Equation (3.5) defines the pose of a rigid object as the relative transformation between two coordinate systems, namely, the object coordinate system and a reference coordinate system. Equation (3.1) shows that the canonical representation of the relative pose between coordinate systems A and B consists of the rotation $\boldsymbol{R}_{ba}$ and the translation $\boldsymbol{b}_a$. However, using Eq. (3.3) we can derive three additional equivalent transformation equations using $\boldsymbol{R}_{ba}$ or $\boldsymbol{R}_{ab}$ and $\boldsymbol{b}_a$ or $\boldsymbol{a}_b$. This shows the importance of a clear notation of pose such as the one given here instead of simply defining a pose as $\boldsymbol{R}, \boldsymbol{t}$ as is often encountered.

3.1.2 Cameras

Red green blue or monochrome cameras are the most common devices used for tracking for AR as they offer low cost and wide availability and can provide highly accurate tracking. Therefore, we will describe the camera geometry models focusing on the most commonly used perspective camera model and discuss intrinsic and extrinsic camera parameters and lens distortion models. Combining this information then leads to the formulation of the equations that are used to establish a correspondence between 3D points in the real world and their projected pixel locations on the camera image, which is the cornerstone of camera-based pose tracking.

Perspective Camera Model

A graph representation of the ***perspective*** or "pinhole" ***camera model*** is given in Fig. 3.1. In this model, the camera is represented by a 2D image plane and the center of projection C. All rays of light are gathered on a central point and then projected on the image plane. The image plane is perpendicular to the z-axis of C also referred to as the optical axis or principal axis. The principal point P is located at the intersection of the optical axis and the image plane and is the origin of the image coordinate system P. The camera focal distance or focal length f is the distance from the camera center to the image plane. A 3D point $\boldsymbol{m}_c = [x, y, z]^\intercal$ is mapped to a 3D point $\boldsymbol{m}_p = [fx/z, fy/z, f]^\intercal$ in the image plane. The corresponding 2D point is then $\boldsymbol{n}_p = [fx/z, fy/z]^\intercal$.

Intrinsic Parameters

Using a homogeneous representation of 2D points $\boldsymbol{n}_p = [x_p, y_p, 1]^\intercal$, the mapping of 3D points to 2D image points can be expressed in a linear set of equations so that

$$\boldsymbol{n}_p = \boldsymbol{K}\boldsymbol{m}_c, \tag{3.6}$$

where K is the ***camera intrinsic parameters*** matrix. For the perspective camera model, $\boldsymbol{K}$ is a diagonal matrix with $diag(\boldsymbol{K}) = [f, f, 1]$. In practice, the following form of the intrinsic matrix is used:

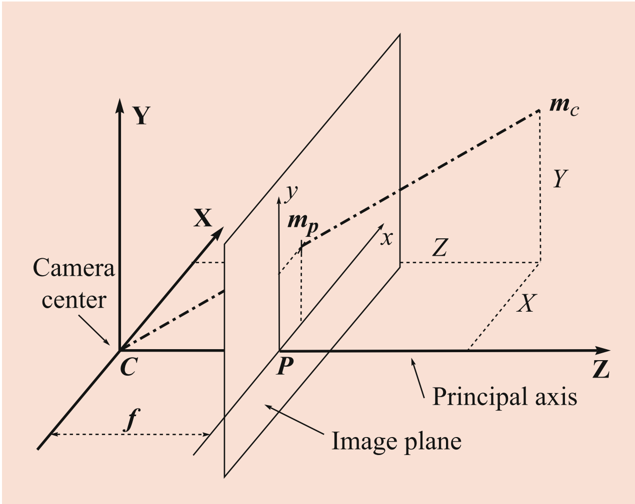


Fig. 3.1 Perspective camera geometry

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.7}$$

where f_x, f_y is the focal length in pixels in the x and y directions that are obtained as $f_x = f/d_x, f_y = f/d_y$, where d_x, d_y are the width and height of the camera pixels in metric units. Similarly c_x, c_y is the principal point P in pixels. The skew factor s is only nonzero to compensate for cases when the image plane is not orthogonal to the optical axis.

Image Distortion

The projective camera model presented above is an ideal camera model that does not consider the camera lens and its effects. Most camera lenses, especially wide angle, induce distortion. Thus, after the perspective projection, an additional 2D to 2D mapping is required to account for the ***distortion***.

Several distortion models can be found in the literature [104]. Radial distortion appears as a deficiency in straight lines transmission and is stronger on the edges of the image. In barrel radial distortion points are shifted toward the center of the image, whereas in radial pincushion distortion points are shifted outwards. Radial distortion is modeled using a polynomial approximation. A distorted 2D point $\boldsymbol{n}_{p_d} = [x_{p_d}, y_{p_d}]$ is mapped to the equivalent undistorted point $\boldsymbol{n}_{p_u} = [x_{p_u}, y_{p_u}]$ by the following equation:

$$\begin{bmatrix} x_{p_d} \\ y_{p_d} \end{bmatrix} = (1+\kappa_1 r^2+\kappa_2 r^4+\kappa_3 r^6+\cdots+\kappa_N r^{2N}) \begin{bmatrix} x_{p_u} \\ y_{p_u} \end{bmatrix}, \tag{3.8}$$

where κ_i are distortion parameters and $r^2 = x_{p_u}^2 + y_{p_u}^2$.

Tangential distortion occurs when the lens is not exactly parallel to the image plane. The mapping equation is in this case

$$\begin{bmatrix} x_{p_d} \\ y_{p_d} \end{bmatrix} = \begin{bmatrix} 2\sigma_1 x_{p_u} y_{p_u} + \sigma_2 \left(r^2 + 2x_{p_u}^2\right) \\ \sigma_1 \left(r^2 + 2x_{p_u}^2\right) + 2\sigma_2 x_{p_u} y_{p_u} \end{bmatrix}, \tag{3.9}$$

where σ_1, σ_2 are the tangential distortion parameters. Distortion models such as the ones from Eqs. (3.8) and (3.9), can be combined in a single mapping equation [50]. In practice, often only two or three radial distortion parameters are considered. An example of a distorted image and its undistorted version is given in Fig. 3.2. In the following we refer to the vector of all considered distortion parameters as $\boldsymbol{\kappa}$.

Extrinsic Parameters

The camera intrinsic parameters describe the mapping of 3D points in the camera coordinate system to 2D points in the image plane. Additionally, the mapping of 3D points expressed in an external coordinate system (e.g., a world coordinate system W or an object coordinate system O) to the camera coordinate system C is of interest. This 6DoF camera pose or ***camera extrinsic parameters*** is graphically depicted in Fig. 3.3. The mapping of a 3D point $\boldsymbol{m}_w$ in the world coordinate system to a 3D point in the camera coordinate system $\boldsymbol{m}_c$ can be given by $\boldsymbol{R}_{cw}$, $\boldsymbol{c}_w$ as

$$\boldsymbol{m}_c = \boldsymbol{R}_{cw}(\boldsymbol{m}_w - \boldsymbol{c}_w), \tag{3.10a}$$

or alternatively by $\boldsymbol{R}_{cw}$, $\boldsymbol{w}_c$ as

$$\boldsymbol{m}_c = \boldsymbol{R}_{cw}(\boldsymbol{m}_w - \boldsymbol{c}_w) = \boldsymbol{R}_{cw}\boldsymbol{m}_w - \boldsymbol{R}_{cw}\boldsymbol{c}_w = \boldsymbol{R}_{cw}\boldsymbol{m}_w + \boldsymbol{w}_c. \tag{3.10b}$$

Using a homogeneous representation of 3D points $\boldsymbol{m}_c = [x_c, y_c, z_c, 1]^\intercal$, $\boldsymbol{m}_w = [x_w, y_w, z_w, 1]^\intercal$, we can express the pose transformation with a matrix multiplication as

$$\boldsymbol{m}_c = \begin{bmatrix} \boldsymbol{R}_{cw} & \boldsymbol{w}_c \\ 0 & 1 \end{bmatrix} \boldsymbol{m}_w, \tag{3.10c}$$

where $\mathcal{M}_{cw}$ is the ***modelview matrix***.

From World to Pixel Coordinates

Camera intrinsics and camera pose (extrinsics) can be combined to provide a mapping of 3D points in any external coordinate system to 2D pixel coordinates in the camera image. This is linearly expressed in a formulation combining previously given Eqs. (3.6) and (3.10c) and using homogeneous representations of 2D and 3D points. A 3D point in a world or object coordinate system $\boldsymbol{m}_w$ is then projected to the 2D point $\boldsymbol{n}_p$ in the image plane using

$$z_c\boldsymbol{n}_p = \boldsymbol{K}\left[\boldsymbol{I}_{3\times 3} \;\; \boldsymbol{0}_{3\times 1}\right]\mathcal{M}_{cw}\boldsymbol{m}_w = \boldsymbol{P}\boldsymbol{m}_w, \tag{3.11}$$

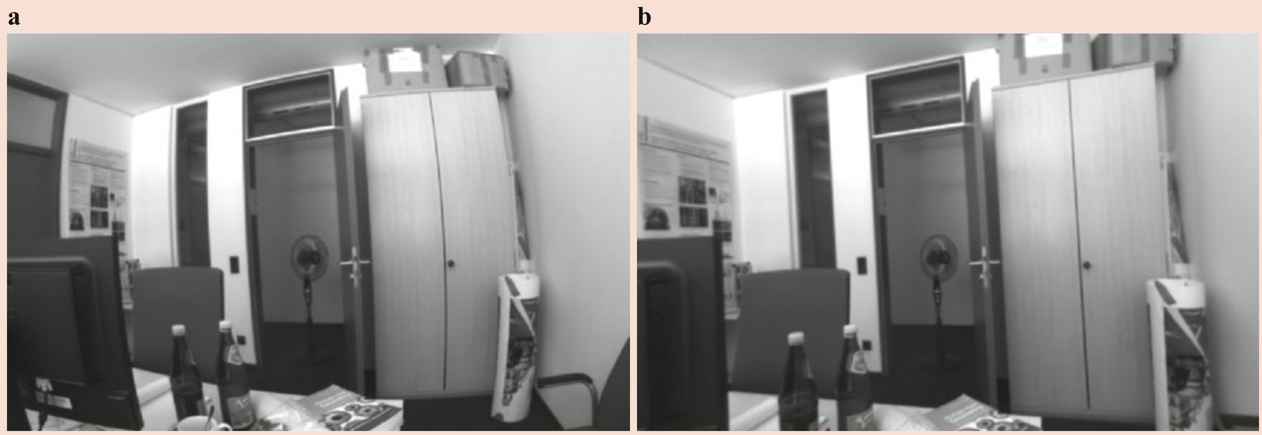


Fig. 3.2 (**a**) Distorted image from a wide angle lens camera and (**b**) its undistorted version

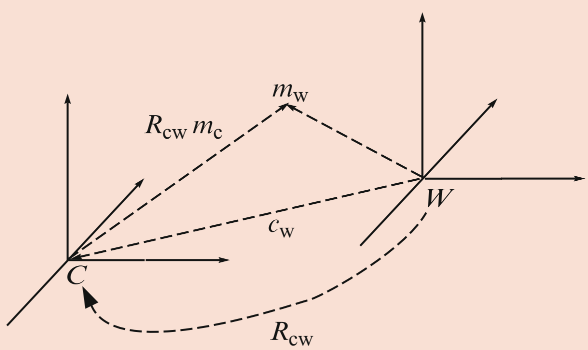


Fig. 3.3 Transformation between camera coordinate system C and world coordinate system W

where $\boldsymbol{I}$ is the identity matrix and $\mathbf{0}$ a vector of zeros. Note that the projected 2D point still needs to be normalized by dividing by the depth z_c in the camera coordinate system to obtain its image pixel coordinates. The matrix $\boldsymbol{P} = \boldsymbol{K}\left[\boldsymbol{I}_{3\times3} \ \ \mathbf{0}_{3\times1}\right]\mathcal{M}_{cw}$ is also named ***projection matrix***. In the following we denote all the projection operations from 3D world coordinates to 2D pixel coordinates including the undistortion mapping using the projection function Π as

$$\boldsymbol{n}_p = \Pi(\boldsymbol{K}, \boldsymbol{\kappa}, \boldsymbol{R}_{cw}, \boldsymbol{w}_c, \boldsymbol{m}_w). \tag{3.12}$$

Establishing this correspondence between the camera image and the real world is of fundamental importance for AR as it allows virtual overlays to be placed in the image that correspond to 3D positions in the real world, thus achieving a seamless integration of the virtual and real world. The quality of visual tracking systems relies on an accurate camera calibration (intrinsic parameters estimation) and an accurate estimation of the camera pose (extrinsics) on any frame under challenging conditions.

Camera Calibration

Camera ***calibration*** is the process of estimating the distortion parameters and intrinsic parameters of a camera. Typically, calibration is an offline procedure performed before the use of cameras in a tracking system. Calibration approaches use several images of a calibration pattern (e.g., a chessboard pattern) acquired from different poses (see Fig. 3.4). The chessboard corners provide reliable correspondences of 2D image points to 3D world coordinates $(\boldsymbol{n}_p, \boldsymbol{m}_w)_{i,j}$ where i is the image number and j the chessboard corner number. Commonly used calibration techniques make an initial estimation of the intrinsic and extrinsic camera parameters ignoring the distortion [135]. Because all the chessboard points lie on a plane, it is possible to use a homography pose estimation for this (further explained in the section "Homography Pose Estimation"). The initial estimation of the intrinsics and pose and the distortion coefficients can then be refined by solving the nonlinear least squares optimization problem to minimize the reprojection error of all 2D points:

$$\min_{(\boldsymbol{K},\boldsymbol{\kappa},\boldsymbol{R}_{cw},\boldsymbol{w}_c)} \sum_i \sum_j \|\boldsymbol{n}_p^{(i,j)} - \Pi(\boldsymbol{K}, \boldsymbol{\kappa}, \boldsymbol{R}_{cw}^{(i)}, \boldsymbol{w}_c^{(i)}, \boldsymbol{m}_w^{(i,j)})\|^2. \tag{3.13}$$

Standard optimization techniques such as the Levenberg-Marquardt algorithm can be used for this minimization [92]. Several software solutions for camera calibration are freely available, for example, Bradski [16].

Although such offline calibration techniques using known patterns are usually sufficient in practice, autocalibration of cameras observing an unknown scene is also a widely researched topic. Approaches using information from supplementary sensors such as inertial measurement units, specific

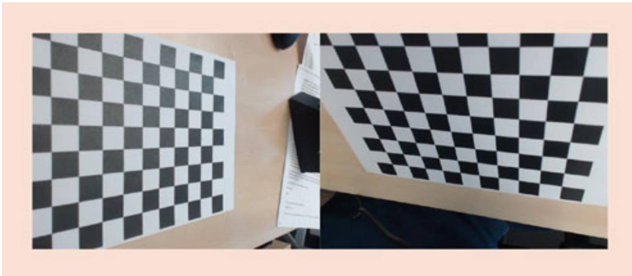


Fig. 3.4 Images of a chessboard pattern for camera calibration

constraints from the scene observation and approaches including the calibration within a Structure from Motion (SfM) framework can be encountered [102, 136].

Spherical Camera Model

The perspective camera model (section "Perspective Camera Model") cannot accurately describe all types of cameras. The radial distortion models of the section "Image Distortion" can compensate for small distortions of wide-angle lenses but are not sufficient to describe omnidirectional cameras such as catadioptric cameras using lenses and curved mirrors or fisheye cameras. However, all these cameras can be described by a central projection of light rays on a specific imaging surface. Estimating the relation between a 2D point on that surface and the originating ray is the task of camera calibration. The 3D rays can be modeled in a generic way using a spherical distribution of rays around the camera center. It is therefore useful for a large class of central cameras to work directly in a ***spherical camera model***.

Rolling Shutter Effects

To this point, in our description of camera models we have assumed that the cameras use a global shutter, meaning that they capture the entire image simultaneously. However, cameras used in practice very often have rolling shutters meaning that the images are captured by horizontally or vertically scanning the scene. ***Rolling shutter*** cameras use inexpensive CMOS sensors and are thus widely used (e.g., cameras on smartphones); however, the captured images can be affected by distortion. The distortion in rolling shutter cameras is caused by the fact that different parts of the images are captured with small time differences. Thus, when the camera is moving fast or fast-moving objects are present in the scene, the distortion becomes clearly visible.

This distortion induces systematic errors in the projection equations, especially under fast motion, and can cause failures of tracking systems. Undistorting the rolling shutter effects on images is a challenging problem that requires knowledge of the camera motion relative to all parts of the scene and the specific sensor scanline. Determining the delay of each scanline of the sensor is a calibration problem [87]. Existing approaches to mitigate the rolling shutter effects are based on simplifications of the problem, such as using a motion model or assuming a planar scene [8, 71].

3.1.3 Inertial Sensors

Inertial sensors are commonly used in tracking systems for AR, usually as complementary sensors to the visual ones. The term Inertial Measurement Unit (IMU) refers to the combination of an accelerometer, gyroscope, and a magnetometer. This sensor package can measure translational and rotational kinematics of a moving body without need for external references. Typical IMU sensor systems provide calibrated measurements in physical units as well as 3D body orientation. All sensor readings, namely, the 3D angular velocity, 3D linear acceleration, and 3D earth-magnetic field, are expressed in the sensor coordinate system S, which is normally aligned to the external housing of the sensor. The individual sensors are usually bundled with a microprocessor for the synchronization of measurements and the communication modules to the recipient of the sensor readings.

Magnetometers are known to be very sensitive to magnetic disturbances owing to the presence of magnetic fields other than the earth's soft iron effect and temperature change effects [103]. Therefore, in many applications the use of the magnetometer readings is excluded, especially in industrial or in-vehicle applications.

3D orientation estimation of the sensor with respect to a coordinate system aligned to the earth's gravitational force direction can be done using the accelerometer readings (assuming no translational movement of the sensor) and/or the magnetometer readings as absolute measurements and the integrated angular velocity gyroscope readings for relative updates of the angles. By convention, the coordinate frame with respect to which the orientation is estimated has the x-axis pointing to the local magnetic north and the z-axis pointing opposite to gravity (i.e., away from the earth). We will refer to this coordinate system as the global coordinate system G. The global and sensor coordinate systems are also illustrated in Fig. 3.5.

Gyroscopes

Ideally gyroscopes measure the angular velocity $\boldsymbol{\omega}_s^{gs} = [\omega_x, \omega_y, \omega_z]^\intercal$ in the sensor coordinate frame S. This angular velocity can be assumed relative to the global coordinate frame G. In practice, the calibrated gyroscope signal $\boldsymbol{\gamma}_s$ is not noise-free but can be modeled as

$$\boldsymbol{\gamma}_s = \boldsymbol{\omega}_s^{gs} + \boldsymbol{b}_s^{\omega} + \boldsymbol{e}_s^{\omega}, \tag{3.14}$$

where $\boldsymbol{b}_s^{\omega}$ is a bias that can be partially orientation dependent and $\boldsymbol{e}_s^{\omega}$ is zero-mean Gaussian noise.

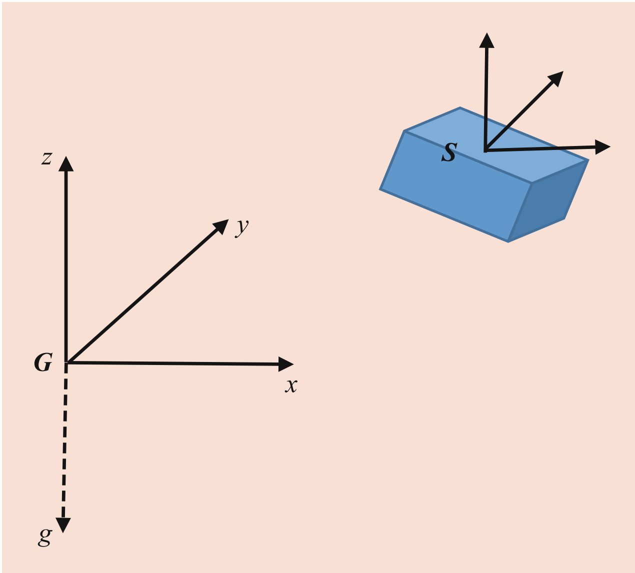


Fig. 3.5 The global coordinate system G has the z-axis pointing opposite to gravity g. The sensor provides measurements on the sensor coordinate system S, which is attached to its housing

Accelerometers

The 3D accelerometer measures all nongravitational accelerations affecting the IMU, which consist in the body acceleration $\ddot{\boldsymbol{s}}_s$ (free acceleration), plus the perceived acceleration induced by gravity ($-\boldsymbol{g}_s$), both expressed in the sensor frame S. Note that the perceived acceleration induced by gravity has a direction opposed to the gravity vector $\boldsymbol{g}$, because the accelerometer does not measure gravity directly but interprets a force due to mass attraction as a force due to a movement induced by an acceleration in the opposite direction. Similar to the gyroscope signal, the calibrated accelerometer signal $\boldsymbol{\alpha}_s$ can also be modeled with a bias:

$$\boldsymbol{\alpha}_s = \ddot{\boldsymbol{s}}_s - \boldsymbol{g}_s + \boldsymbol{b}_s^{\alpha} + \boldsymbol{e}_s^{\alpha}. \tag{3.15a}$$

In the global frame G, the gravity vector is a constant $\boldsymbol{g}_g = [0, 0, -g]^\intercal$. Therefore, by expressing the linear acceleration in the global frame, we can rewrite Eq. (3.15a) as

$$\boldsymbol{\alpha}_s = \boldsymbol{R}_{sg}(\ddot{\boldsymbol{s}}_g - \boldsymbol{g}_g) + \boldsymbol{b}_s^{\alpha} + \boldsymbol{e}_s^{\alpha}, \tag{3.15b}$$

meaning that if the orientation of the sensor in the global frame is known, it is possible to remove the part related to gravity from the accelerometer signal to extract a noisy measurement of the free acceleration given by

$$\ddot{\boldsymbol{s}}_g = \boldsymbol{R}_{sg}^\intercal \left(\boldsymbol{\alpha}_s - \boldsymbol{b}_s^{\alpha} - \boldsymbol{e}_s^{\alpha}\right) + \boldsymbol{g}_g. \tag{3.15c}$$

Inertial Navigation

It is possible to track the orientation and position relative to a known starting point of a moving body that has an IMU attached to it using only the gyroscope and accelerometer readings. This process is called ***strapdown inertial navigation*** [123]. On each time step, which corresponds to the IMU period, the procedure consists of first updating the orientation by integrating the angular velocities measurements from the gyroscope. Subsequently, the newly estimated orientation can be used to subtract gravity from the accelerometer measurements. The resulting body acceleration is double integrated in order to update the position.

As discussed previously, inertial measurements from IMU devices suffer from noise and biases, the range of which is dependent on the sensor quality. The accumulation of these errors causes the overall estimation to drift over time. Most critically, a small error in orientation causes an error in gravity removal, which leads to nonexistent body accelerations, which in turn leads to rapid increase in the position error due to the double integration. For this reason, inertial navigation is often done within a statistical filtering scheme or is corrected through fusion with another tracking source.

3.1.4 Depth Sensors

The acquisition of 3D information of the environment is one of the most crucial tasks for vision-based tracking systems. For systems that operate in unknown environments, 3D scene estimation has to be done in real time parallel to tracking. 3D information can be obtained from detection of features in different camera views with known poses using multi-view geometry techniques. This is one of the most challenging tasks in computer vision. Sensors that can provide direct depth measurements from a scene without requiring multiple views can be of assistance in this task.

Depth cameras provide depth images where each pixel stores a depth value (see Fig. 3.6). This value is the orthogonal distance between the visible corresponding 3D point and the optical camera center. Given the 2D location of a feature in the image $\boldsymbol{n}_p$ and a depth measurement $\hat{z}_c$, the 3D location of the corresponding point $\boldsymbol{m}_c$ in camera coordinates is computed as

$$\boldsymbol{m}_c = \boldsymbol{K}^{-1} \begin{bmatrix} \boldsymbol{n}_p^\intercal \\ 1 \end{bmatrix} \hat{z}_c. \tag{3.16}$$

If the extrinsic camera parameters are known, the point in world coordinates can also be computed. The provided input of depth cameras can be very useful in tracking systems; however, the limitations of current depth sensing technologies have to be considered as well. Depth sensors tend to have high noise levels, a limited range, and issues in outdoor functionality.

Several types of depth-sensing technologies have been used in depth cameras. Although magnetic-based and sound-based systems exist and are used in many applications, their accuracy is not sufficient for AR tracking applications. Thus,

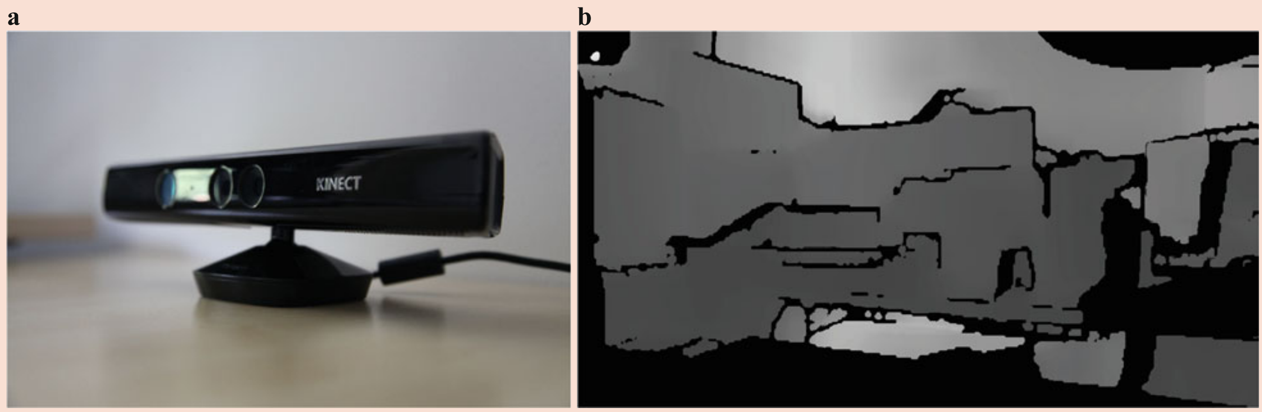


Fig. 3.6 (**a**) The Microsoft Kinect device. (**b**) Output of a depth camera

light-based techniques are of greater interest in the AR tracking context. These can be classified in laser, multi-view, pattern projection, and time-of-flight sensors. Laser scanners are highly accurate but with a prohibitive cost for many applications. Multi-view approaches (e.g., stereo cameras) use multiple regular cameras in a fixed setting and estimate depth from feature matching and are thus still relying on visual texture features. Pattern projection depth cameras consist of an infrared laser emitter, infrared camera, and RGB camera. They also use multiple view geometry for depth estimation but enhance its robustness by observing the projected pattern from the laser. The first version of the Microsoft Kinect is such a sensor that is being widely used in AR tracking applications. Finally, ***Time-of-Flight (ToF)*** sensors estimate depth from the time required for the light signal to reach the scene and be received back by the camera. The concept was used in the second version of the Microsoft Kinect device achieving improved quality of image depth [129]. For a detailed analysis of the topic, we refer to Sansoni et al. [112].

3.1.5 Geospatial Navigation Sensors

Geospatial navigation using Global Navigation Satellite Systems (GNSS) has been used in several AR applications to provide a coarse localization. GNSS are currently available or being implemented in many regions: The American Global Positioning System (GPS) [43], the Russian counterpart constellation GLONASS, and the European Union Galileo. GNSS has an accuracy of about 10–15 m, which would not be sufficient for precise position tracking, but several techniques to increase the precision are already available. These techniques are referred to as GNSS augmentation. GNSS augmentation is a method of improving the navigation system's attributes, such as accuracy, reliability, and availability, through the integration of external information into the calculation process. One can distinguish between satellite-based augmentation systems (SBAS) that use additional satellite-broadcast messages and ground-based augmentation systems (GBAS), where the precision is increased by the use of terrestrial radio messages. An example of SBAS is the European Geostationary Navigation Overly Service (EGNOS), which supplements the GPS, GLONASS, and Galileo systems by reporting on the reliability and accuracy of the positioning data. This increases the horizontal position accuracy to around 1 m. A similar system used in North America is the Wide Area Augmentation System (WAAS), which provides an accuracy of about 3–4 m. In any case, the accuracy levels provided by GNSS systems are far from the required accuracy of most AR systems. However, GNSS are often fused with other types of localization sensors, such as inertial sensors or cameras. The benefit of GNSS is that they can provide an absolute global position in any outdoor environment. This position can be used by other AR trackers to load local data for tracking (e.g., images for visual matching) or for loading local virtual content (augmentations). Thus, GNSS localization is an important tool for large-scale outdoor AR applications especially considering the wide availability of GNSS receivers on consumer devices such as smartphones.

3.2 Computer Vision Techniques

Visual tracking using camera images is the most commonly used tracking method for AR. Therefore, in this section we analyze the main computer vision techniques involved in visual tracking, starting from the matching of visual features between images and continuing with techniques for robust pose estimation and structure from motion techniques used in SLAM systems.

3.2.1 Feature Matching

Feature matching refers to the problem of finding corresponding landmarks between images on a wide baseline. In other words visual features are searched in a target image without any prior knowledge of their location, i.e., a full image search is required. It is possible to match features by means of an exhaustive search over the image using a correlation metric on the pixel intensities; however, such solutions are highly expensive computationally while being sensitive to illumination variation and viewpoint changes. Therefore, the use of ***feature descriptors*** is widely used for this task.

Descriptor-based approaches perform an encoding of image patches. The process is divided into feature detection and matching. Detection is aimed at finding areas in the images that have strong edge characteristics. This is often done using the autocorrelation matrix of the image intensity as proposed in the Harris corner detector [47]. Features from Accelerated Segment Test (FAST) [105] poses a trade-off between computational complexity and performance by adding fast elimination of non-corner points using machine learning classification on the candidate points. Matching is done by comparing the encoded features, making it much more efficient.

A large number of approaches to feature detection and matching have been proposed over the years, a recent overview and evaluation of which can be found in Mukherjee et al. [79]. The appearance of features between images can vary significantly; therefore, feature descriptors are targeted toward invariance to scale, illumination, orientation, and affine or perspective transformations. The scale-invariant descriptor Scale-Invariant Feature Transform (SIFT) [74] is a position-dependent histogram of local image gradient directions around the corner point. Oriented FAST and rotated BRIEF (ORB) [106] features show stable performance on many applications while requiring moderate computational effort, which makes them suitable for real-time systems [79]. Feature orientation is computed as a vector from the intensity centroid of an image patch around the corner point. A 32-byte binary Binary Robust Independent Elementary Features (BRIEF) [20] descriptor is computed from the patch using intensity comparisons on 256 pixel pairs. ORB orients these descriptors based on its angle vector to obtain an orientation invariant descriptor. The matching can then be done by computing Hamming distances between the descriptors. An example of wide baseline feature matching with ORB features is given in Fig. 3.7.

Feature matching using descriptors does not typically provide the required accuracy needed in AR tracking systems. Therefore, it is mostly used for estimating a coarse pose that is further refined using other techniques or is used for initialization of trackers or relocalization.

3.2.2 Feature Tracking

Although feature matching is efficient for wide baseline matching of images, there are cases when prior knowledge of the visual features positions is available and should be taken into account. In particular, when the camera pose is being tracked from frame to frame in an image sequence, the assumption holds that visual features can be found in the same neighborhood between consecutive frames. Therefore, a local search for features around their previous position is more accurate and efficient than a global feature search.

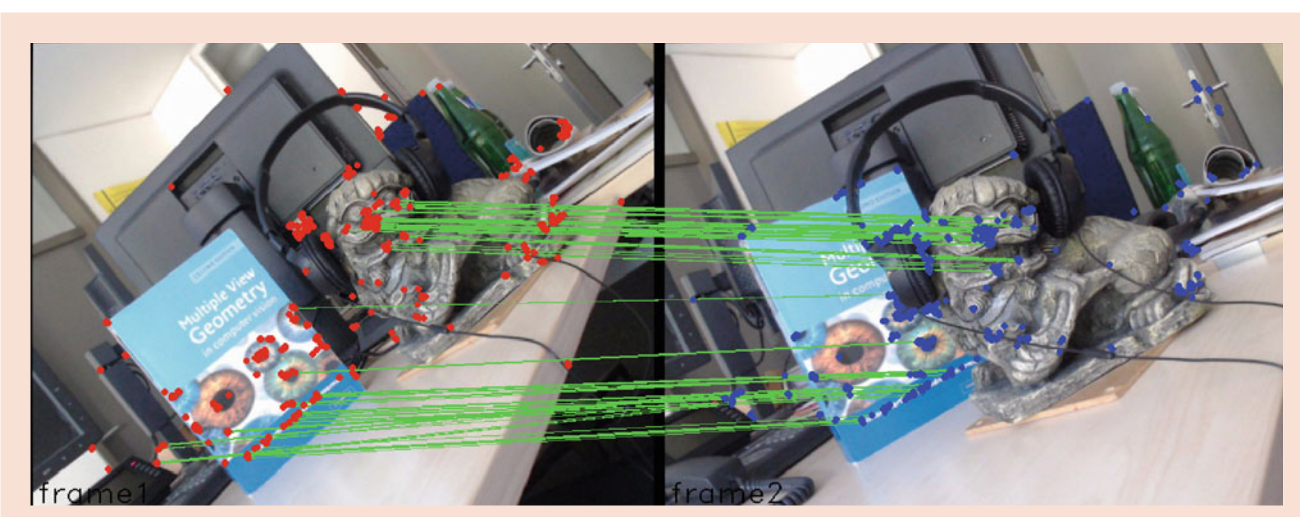


Fig. 3.7 Matching of ORB points between two images with a large baseline and rotation

Optical Flow Estimation

The problem of tracking the motion of image features between adjacent frames is also encountered in the literature as ***2D optical flow*** [36]. Existing solutions are based on the assumptions that pixel intensities do not change between consecutive frames and that neighboring pixels have similar motion. Given a pixel $\mathcal{I}(x, y, t)$ in the initial image, the ***intensity constraint*** suggests that in the next frame a pixel with the same intensity can be found at a displacement dx, dy after time dt so that

$$\mathcal{I}(x, y, t) \approx \mathcal{I}(x + dx, y + dy, t + dt). \qquad (3.17a)$$

Expanding Eq. (3.17a) in a Taylor series up to first-order term gives

$$\mathcal{I}(x, y, t) = \mathcal{I}(x, y, t) + \frac{\partial \mathcal{I}}{\partial x} dx + \frac{\partial \mathcal{I}}{\partial y} dy + \frac{\partial \mathcal{I}}{\partial t} dt + \cdots . \qquad (3.17b)$$

Dividing Eq. (3.17b) by dt gives

$$\frac{\partial \mathcal{I}}{\partial x}\frac{dx}{dt} + \frac{\partial \mathcal{I}}{\partial y}\frac{dy}{dt} + \frac{\partial \mathcal{I}}{\partial t} = 0, \qquad (3.17c)$$

which can be written as

$$\nabla \mathcal{I} \boldsymbol{v} + \mathcal{I}_t = 0, \qquad (3.17d)$$

where $\nabla \mathcal{I} = (\mathcal{I}_x, \mathcal{I}_y)$ is the spatial gradient of the image and $\boldsymbol{v} = (v_x, v_y)$ is the image velocity. This is a linear equation of two unknowns that cannot be solved as such, and additional conditions need to be added. Several methods for solving the optical flow problem have been proposed, an overview and evaluation of which are available in Baker et al. [9].

One of the most frequently used optical flow methods in AR tracking systems is the Kanade-Lucas-Tomasi (KLT) method [75]. KLT uses the additional assumption that neighboring pixels have the same motion. Thus, the intensity constraint from Eq. (3.17d) must hold for all N pixels $\boldsymbol{n}_i, i = 1, \cdots, N$ in a small area around the tracked corner point. The resulting equations for each pixel can be written together in the matrix form $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{bmatrix} \mathcal{I}_x^{\boldsymbol{n}_1} & \mathcal{I}_y^{\boldsymbol{n}_1} \\ \mathcal{I}_x^{\boldsymbol{n}_2} & \mathcal{I}_y^{\boldsymbol{n}_2} \\ \vdots & \vdots \\ \mathcal{I}_x^{\boldsymbol{n}_N} & \mathcal{I}_y^{\boldsymbol{n}_N} \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} -\mathcal{I}_t^{\boldsymbol{n}_1} \\ -\mathcal{I}_t^{\boldsymbol{n}_2} \\ \vdots \\ -\mathcal{I}_t^{\boldsymbol{n}_N} \end{bmatrix}. \qquad (3.18)$$

The system of equations from Eq. (3.18) is now overdetermined, and a solution can be found using least squares. The solution is then given by

$$\boldsymbol{A}^\intercal \boldsymbol{v} = \boldsymbol{A}^\intercal \boldsymbol{b} \qquad (3.19a)$$

or

$$\boldsymbol{v} = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal \boldsymbol{b}. \qquad (3.19b)$$

Equation (3.19b) can be solved if $\boldsymbol{A}^\intercal \boldsymbol{A}$ is invertible which is ensured if its eigenvalues satisfy $\lambda_1 \geq \lambda_2 > 0$. Additionally, λ_1 being much larger than λ_2 indicates that there is a strong feature gradient only on one direction, making the feature an edge rather than a corner, which is not suitable for tracking. Thus, a good condition for a feature to be tracked is that the two eigenvalues have similarly high values.

We presented the formulation of a basic solution for the optical flow problem (see also Fig. 3.8). Algorithms used in practice use various techniques to improve the overall performance. For example, motion models can be integrated in the estimation, a limit on the feature displacement between two frames can be set to remove outliers, or reverse optical flow can be used to increase robustness [45, 54].

For optical flow to perform well, limited motion between the consecutive images can exist so that the intensity assumption holds. This can be enhanced using a ***coarse-to-fine*** strategy with a pyramidal structure implementation [14]. In this coarse-to-fine strategy, estimation starts with subsampled coarse images and gradually moves to higher resolutions in order to refine the results. The image pyramids are obtained by repeatedly subsampling and lowpass filtering the image. ***Image pyramids*** and coarse-to-fine techniques are widely used in feature matching and tracking implementations.

In contrast to gradient-based feature tracking approaches such as KLT, ***block-based matching*** methods (used for example in Davison and Murray [27]) follow a different principle. Using an estimate of the expected position of an image patch, such techniques perform an exhaustive local search in order to find the position that maximizes the correlation between the original patch and the target image.

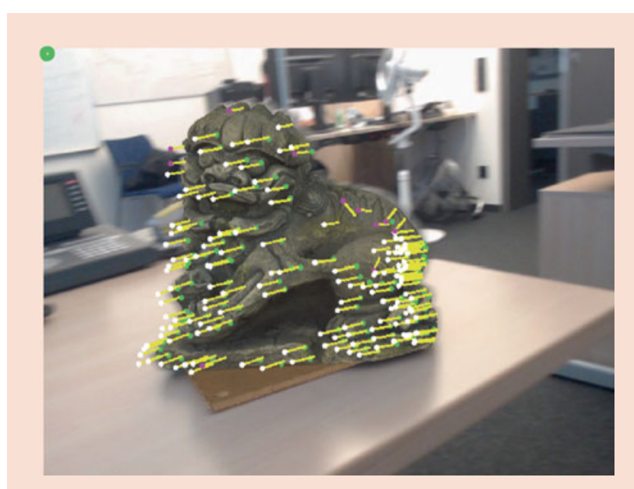


Fig. 3.8 Optical flow vectors on a tracked object under fast camera movement

Finally, another important aspect that influences the success of optical flow tracking is the frame rate of the camera. A low frame rate signifies larger displacements of features between consecutive frames, which can make the tracker less efficient. On the other hand, a high frame rate could be challenging for some devices of limited computational power as frames have to be processed much faster to ensure real-time functionality.

3.2.3 Pose Estimation from Feature Correspondences

Feature matching or tracking provides correspondences between 2D features from a source and a destination image. If the pose and the 3D locations of the features are known in the source image, then correspondences between 3D points seen in the source image and 2D points in the destination image are established. These correspondences can be used to obtain a solution for the pose matrix $\mathcal{M}_{cw}$ of Eq. (3.11). Many approaches exist in the literature to solving this problem, also commonly referred to as the ***Perspective-n-Point (PnP)*** problem [29, 39, 51, 67].

The DLT Algorithm

The Direct Linear Transformation (DLT) algorithm is a direct method for estimating the projection matrix $\boldsymbol{P}$ of Eq. (3.11) when a number of N correspondences $\boldsymbol{m}_w^i \leftrightarrow \boldsymbol{n}_p^i, i = 1, \cdots, N$ exist between 3D points $\boldsymbol{m}_w^i$ and 2D image points $\boldsymbol{n}_p^i$. A solution for the projection matrix $\boldsymbol{P}$ is required so that $\boldsymbol{n}_p^i = \boldsymbol{P}\boldsymbol{m}_w^i$ is satisfied for every correspondence i, which also translates to $\boldsymbol{n}_p^i \otimes \boldsymbol{P}\boldsymbol{m}_w^i = 0$, where $\otimes$ denotes the vector cross product.

With $\boldsymbol{n}_p^i = [x_p^i, y_p^i, z_p^i]^\top$, this can be written as

$$\boldsymbol{n}_p^i \otimes \boldsymbol{P}\boldsymbol{m}_w^i = \begin{bmatrix} y_p^i \boldsymbol{P}^{3\top}\boldsymbol{m}_w^i - z_p^i \boldsymbol{P}^{2\top}\boldsymbol{m}_w^i \\ z_p^i \boldsymbol{P}^{1\top}\boldsymbol{m}_w^i - x_p^i \boldsymbol{P}^{3\top}\boldsymbol{m}_w^i \\ x_p^i \boldsymbol{P}^{2\top}\boldsymbol{m}_w^i - y_p^i \boldsymbol{P}^{1\top}\boldsymbol{m}_w^i \end{bmatrix} = 0 \tag{3.20a}$$

with $\boldsymbol{P}^{j\top}$ being the jth row vector of $\boldsymbol{P}$. From Eq. (3.20a) we can obtain the following set of equations:

$$\begin{bmatrix} \boldsymbol{0}_{1\times4} & -z_p^i \boldsymbol{m}_w^{i\,\top} & y_p^i \boldsymbol{m}_w^{i\,\top} \\ z_p^i \boldsymbol{m}_w^{i\,\top} & \boldsymbol{0}_{1\times4} & -x_p^i \boldsymbol{m}_w^{i\,\top} \\ -y_p^i \boldsymbol{m}_w^{i\,\top} & x_p^i \boldsymbol{m}_w^{i\,\top} & \boldsymbol{0}_{1\times4} \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_1 \\ \boldsymbol{P}_2 \\ \boldsymbol{P}_3 \end{bmatrix} = 0 \tag{3.20b}$$

The equations of Eq. (3.20b) are not linearly independent, so that only the two first can be used. By stacking the two equations from Eq. (3.20b) for N correspondences, we obtain a $2N \times 12$ matrix $\boldsymbol{A}$. Therefore, a minimum of 6 correspondences are needed to solve the equation $\boldsymbol{A}[\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_3]^\top = 0$. The system of equations can be solved by applying Singular Value Decomposition (SVD) [92] on the matrix $\boldsymbol{A}$. For a calibrated camera with known intrinsic matrix $\boldsymbol{K}$, the model view matrix containing the pose can be computed by multiplication with the inverse of $\boldsymbol{K}$, $\mathcal{M}_{cw} = \boldsymbol{K}^{-1}\boldsymbol{P}$.

Homography Pose Estimation

When all the 3D points $\boldsymbol{m}_w^i$ used for pose estimation lie on the same 3D plane, many existing PnP algorithms fail. Points on a plane are a very common case, encountered, for example, when tracking visual markers or 2D tracking patterns. In fact, considering the planarity constraint leads to a simplification of the PnP algorithms.

Without loss of generality, we consider that all points lie on the $z_w = 0$ plane. The projection equation (Eq. (3.11)) for a point on that plane will then be

$$\begin{aligned} z_c\boldsymbol{n}_p &= \boldsymbol{P}\boldsymbol{m}_w \\ &= \boldsymbol{K}\left[\boldsymbol{r}_{cw}^{(1)}\boldsymbol{r}_{cw}^{(2)}\boldsymbol{r}_{cw}^{(3)}\boldsymbol{w}_c\right]\begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} \\ &= \boldsymbol{K}\left[\boldsymbol{r}_{cw}^{(1)}\boldsymbol{r}_{cw}^{(2)}\boldsymbol{w}_c\right]\begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \\ &= \boldsymbol{H}_{cw}\begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \end{aligned} \tag{3.21}$$

where $\boldsymbol{r}_{cw}^i$ are the columns of the rotation matrix $\boldsymbol{R}_{cw}$ and $\boldsymbol{H}_{cw}$ is the 3×3 ***homography matrix***. The matrix $\boldsymbol{H}_{cw}$ can be recovered using correspondences between 2D image points and 3D planar points with a method similar to the DLT (section "The DLT Algorithm") requiring 4 correspondences only. Using $\boldsymbol{H}_{cw}$, we can compute

$$\left[\boldsymbol{r}_{cw}^{(1)}\boldsymbol{r}_{cw}^{(2)}\boldsymbol{w}_c\right] = \boldsymbol{K}^{-1}\boldsymbol{H} \tag{3.22}$$

and using the vector cross product $\boldsymbol{r}_{cw}^{(3)} = \boldsymbol{r}_{cw}^{(1)} \times \boldsymbol{r}_{cw}^{(2)}$. The recovered rotation matrix $\boldsymbol{R}_{cw}$ is not necessarily orthogonal. Using the singular decomposition $\boldsymbol{R}_{cw} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ allows us to recover the nearest rotation matrix in the sense of the Frobenius norm as $\boldsymbol{R}_{cw} = \boldsymbol{U}\boldsymbol{V}^\top$.

Nonlinear Optimization

Methods of pose estimation such as DLT are in practice used to obtain an initial estimate of the pose, which is then refined by an optimization method. The error function that has to be minimized is the ***reprojection error*** of the points on the image using the estimated pose. Thus, for a set of N correspondences $\boldsymbol{m}_w^i \leftrightarrow \boldsymbol{n}_p^i$, a pose matrix $\mathcal{M}_{cw}$ has to be estimated that minimizes the error

$$\mathcal{E}_{\text{repr}} = \sum_{i=1}^{N} \|\boldsymbol{n}_p^i - \Pi(\boldsymbol{K}, \boldsymbol{\kappa}, \mathcal{M}_{cw}, \boldsymbol{m}_w^i)\|^2. \tag{3.23}$$

Equation (3.23) describes a nonlinear least squares problem that can be solved using nonlinear optimization techniques such as Levenberg-Marquardt optimization. Feature matching and tracking methods generally produce a large number of erroneous correspondences commonly referred to as ***outliers***. Considering these in an optimization scheme will for certain affect the estimated pose and the convergence of the optimization. Robust estimators such as the M-estimator [53] that can reduce the influence of outliers have been proposed for the pose estimation problem. Even these, however, cannot deal with high ratios of outliers in the 3D/2D correspondences. In these cases, pose estimation within a RANSAC framework is more effective.

Random Sample Consensus

Random Sample Consensus (RANSAC) is a general iterative technique for the removal of outliers from a set of observations [35]. In the case of pose estimation, using the abovementioned algorithms within a RANSAC framework is capable of producing robust poses even at high outlier ratios. This, however, comes at a significant computational cost owing to the iterative nature of the algorithm. A basic RANSAC procedure for pose estimation from a set C of 3D/2D correspondences is given in Algorithm 1.

Algorithm 1 RANSAC pose estimation

procedure RANSAC POSE ESTIMATION(C)
  Set of best inliers $\hat{C}_{\text{in}} = \emptyset$
  for Number of RANSAC iterations **do**
    Create a random subset C_s of 6 correspondences from C
    Estimate a camera pose $\mathcal{M}_{cw}{}^s$ from C_s using, e.g., DLT and optimization.
    for each correspondence i in C **do**
      Compute reprojection error $\mathcal{E}_{\text{repr}_i}$ for i as in Eq. (3.23)
      if Reprojection error $\mathcal{E}_{\text{repr}_i}$ is small **then**
        Add i to set of inliers C_{in}
    if Total reprojection error of all inliers $\mathcal{E}_{\text{repr}}^{C_{\text{in}}}$ is less than the error for the best set of inliers so far $\mathcal{E}_{\text{repr}}^{\hat{C}_{\text{in}}}$ **then**
      $\hat{C}_{in} = C_{in}$
  Estimate a final camera pose $\mathcal{M}_{cw}$ from the best inliers set $\hat{C}_{in}$

It is clear that the outcome of the RANSAC pose estimation is dependent on the number of iterations as well as on the threshold of the reprojection error for considering a point to be an inlier. It is also a good practice to set a threshold on the ratio of inliers to outliers for a solution to be considered valid. Modifications of the basic RANSAC approach using improved sampling or cost functions other than the reprojection error have been proposed in Nistér and Subbarao and Meer [86, 116].

3.2.4 The Epipolar Constraint

The ***epipolar constraint*** is defined by the geometry of two camera views, observing a common visual feature $\boldsymbol{m}_w$ in the world. The two views can be acquired with a time difference by a single camera moving in the world or by a multi-camera device (e.g., stereo camera) simultaneously. The two views are coupled by a relative pose $\boldsymbol{R}, \boldsymbol{t}$. The epipolar constraint states that the projection $\boldsymbol{n}_{p,2}$ of a point $\boldsymbol{m}_w$ on one image defines an epipolar line $\boldsymbol{l}_{p,1}$ on the other image on which the corresponding projection $\boldsymbol{n}_{p,1}$ of the same point $\boldsymbol{m}_w$ is located. This is a necessary condition for any feature correspondence between images given the camera pose. The epipolar geometry is graphically depicted in Fig. 3.9.

The relative pose $\boldsymbol{R}, \boldsymbol{t}$ can be computed from the individual camera poses as

$$\boldsymbol{R} = \boldsymbol{R}_{cw,2}\boldsymbol{R}_{cw,1}^{\mathsf{T}} \tag{3.24a}$$

$$\boldsymbol{t} = \boldsymbol{R}\boldsymbol{w}_{c,1} - \boldsymbol{w}_{c,2}. \tag{3.24b}$$

The epipolar constraint is then formally expressed as

$$\left[\boldsymbol{n}_{p,2}^{\mathsf{T}}\ 1\right] \boldsymbol{F} \left[\boldsymbol{n}_{p,1}^{\mathsf{T}}\ 1\right]^{\mathsf{T}} = 0. \tag{3.25}$$

where $\boldsymbol{F}$ is the ***fundamental matrix*** $\boldsymbol{F} = \boldsymbol{K}^{-1\,\mathsf{T}}\boldsymbol{E}\boldsymbol{K}^{-1}$ with $\boldsymbol{E}$ being the ***essential matrix*** that can be defined by the relative rotation $\boldsymbol{R}, \boldsymbol{t}$ as

$$\boldsymbol{E} = \boldsymbol{R} \begin{bmatrix} 0 & -t_z & t_z \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \tag{3.26}$$

The epipolar lines are given by $\boldsymbol{l}_{p,1} = \boldsymbol{F}\left[\boldsymbol{n}_{p,2}^{\mathsf{T}}\ 1\right]^{\mathsf{T}}$ and $\boldsymbol{l}_{p,2} = \boldsymbol{F}^{\mathsf{T}}\left[\boldsymbol{n}_{p,1}^{\mathsf{T}}\ 1\right]^{\mathsf{T}}$, and the points $\boldsymbol{e}_{p,1}, \boldsymbol{e}_{p,2}$ are the epipoles.

Using the epipolar constraint, it is possible to estimate the fundamental matrix from a set of 2D/2D $(n_{p,1}^{(i)}, n_{p,2}^{(i)})$ correspondences between two camera images. A solution requiring 8 correspondences performs similarly to the DLT

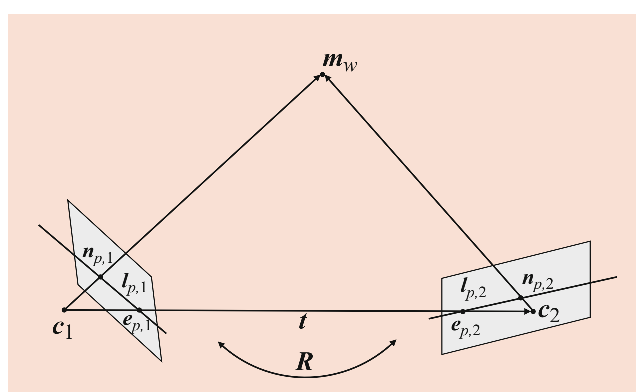


Fig. 3.9 The epipolar geometry between two camera views

algorithm, forming a linear equation system and solving it with SVD [73]. Another solution, estimating the essential matrix from only 5 correspondences, was presented in Nistér [85]. An initial estimate can be refined by nonlinear optimization minimizing the distance of points to epipolar lines. From the estimated fundamental or essential matrix, the relative pose can be extracted as well as the camera poses and 3*D* points, however, with a scaling unknown. The epipolar constraint is also often used for the validation of pose estimation in monocular systems when reference views are available.

3.2.5 Line Tracking

Line tracking techniques are used in many AR systems, especially in model-based trackers for textureless objects that do not have sufficient edge features. When a line model of the tracked object or environment is established, frame-to-frame tracking of lines can be performed. Assuming an image $\mathcal{I}_k$ for which the camera pose $\boldsymbol{R}_{cw}, \boldsymbol{w}_c$ with respect to the line model is known and its subsequent image $\mathcal{I}_{k+1}$ for which we wish to estimate the pose, line tracking approaches first project the line model onto the new image $\mathcal{I}_{k+1}$ using the previous known pose. After sampling a number of control points on the lines, a 1D search for strong gradients along the normal vector of the line is done to search for correspondences (see Fig. 3.10). An edge enhancement approach is often applied on the target images to improve the line matching.

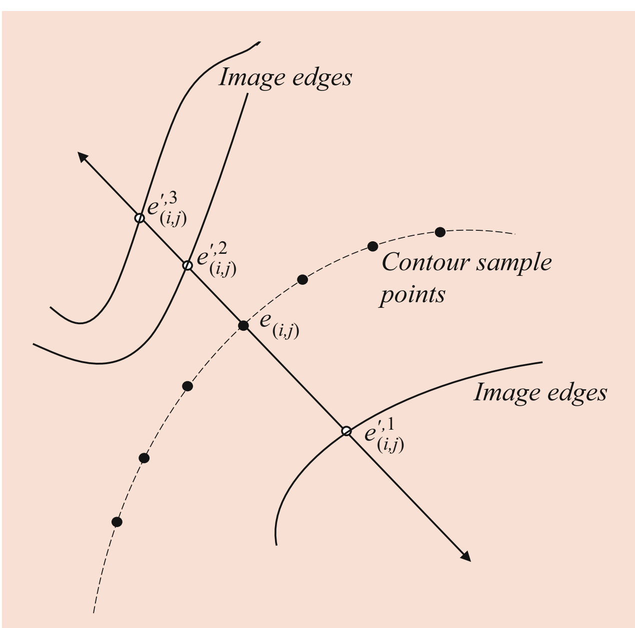


Fig. 3.10 Sample points and image edges as possible correspondences. Considering multiple hypotheses along the normal of each sample point provides a significant improvement

Given the line E_i, the sample point $e_{(i,j)}$ and its match $e'_{(i,j)}$, and a distance function Δ, a cumulative $\mathcal{E}$ error over all lines and sample points can be computed as

$$\mathcal{E} = \sum_i \sum_j \Delta(e_{(i,j)}, e'_{(i,j)}) \tag{3.27}$$

This error formulation can be used as a loss function for an iterative optimization process in order to find the camera pose estimate $\boldsymbol{R}'_{cw}, \boldsymbol{w}'_c$ that minimizes the line error.

Important improvements to the basic line tracking technique of [46] were the use of robust estimators such as the Tukey estimator, to reduce the influence of outliers on the error function. The Tukey estimator can be computed as follows:

$$\rho_{\mathrm{Tuk}}(x) = \begin{cases} \frac{c^2}{6}\left[1 - \left(1 - \left(\frac{x}{c}\right)^2\right)^3\right], & \text{if}|x| \leq c \\ \frac{c^2}{6}, & |x| > c \end{cases}, \tag{3.28}$$

where $c = k\sigma$ is a threshold depending on the standard deviation σ of the estimation error. Another important improvement of line tracking approaches was the incorporation of multiple hypotheses for each control point in the loss function [125] (see Fig. 3.10). Line tracking approaches were sensitive to strong gradients in cluttered scenes that can create wrong correspondences. By modifying the Tukey estimator, several hypotheses for each control point can be considered so that the error function to be minimized becomes

$$\mathcal{E} = \sum_i \sum_j \rho_{\mathrm{Tuk}}\left(\Delta\left(e_{(i,j)}, \left\{e'^{,1}_{(i,j)}, \cdots, e'^{,N}_{(i,j)}\right\}\right)\right). \tag{3.29}$$

3.2.6 Direct Image Alignment

In contrast to feature-based approaches, ***direct image alignment*** methods perform frame-to-frame tracking by aligning consecutive images using information of the entire image. Considering an image $\mathcal{I}_k$ and a subsequent image $\mathcal{I}_{k+1}$, such approaches estimate a geometric transform $\mathcal{T}(\mathcal{M}_{cw}, \boldsymbol{n}_p)$ that transforms all pixels of the initial image to the target image. Finding the parameters of the transform (camera pose) that minimizes the image intensity error (***photometric error***)

$$\mathcal{E}(\mathcal{M}_{cw}) = \sum_{\boldsymbol{n}_p} (\mathcal{I}_{k+1}(\boldsymbol{n}_p) - \mathcal{I}_k(\mathcal{T}(\mathcal{M}_{cw}, \boldsymbol{n}_p))^2 \tag{3.30}$$

gives the pose update for the tracker. Applying the geometric warping model to the images requires maintaining a dense scene model. Solving for the parameters that minimize the intensity error can be done using iterative Gauss-Newton

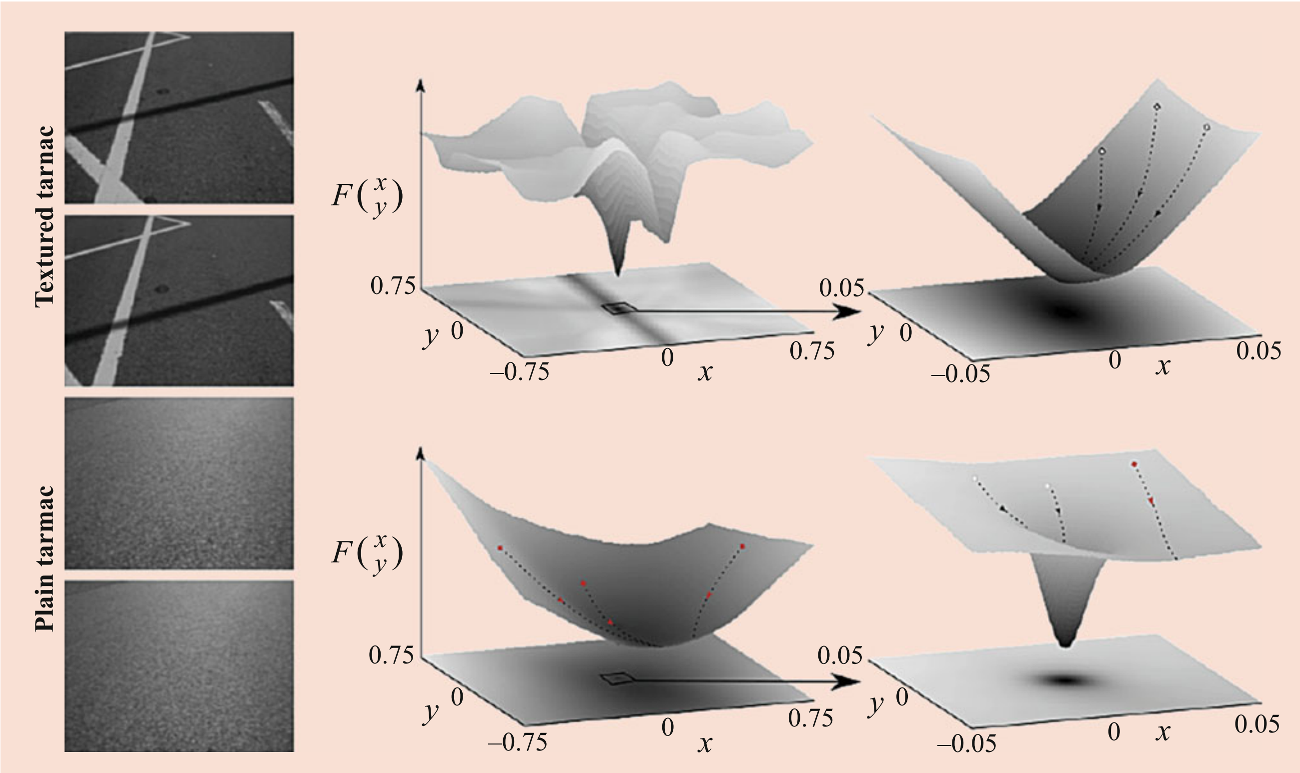


Fig. 3.11 Examples of the direct alignment loss function for two image pairs, textured and plain. Source [1]

minimization. An example of the intensity loss function is shown in Fig. 3.11. Full image direct tracking approaches were introduced in Engel et al. and Newcombe et al. [33, 84].

Direct methods are considered advantageous for using information from the entire image. They are generally computationally more expensive but easier to parallelize and more sensitive to image effects such as strong illumination changes and model inconsistencies such as rolling shutter effects. Additionally, direct methods perform better on images with strong similarity (i.e., adjacent images from high fps cameras).

3.2.7 Structure from Motion

SfM refers to computer vision techniques that simultaneously estimate the camera pose and scene structure from a captured image sequence. SfM is the core technology of visual SLAM and 3D reconstruction and one of the most challenging and extensively researched topics of computer vision. A fundamental SfM approach consists of three modules, namely, a feature matching or tracking approach to establish correspondences between images, a pose estimation module that estimates the current camera pose using the available 3D structure information, and a mapping module that estimates this structure utilizing multiple views of features. Because the structural information is required for the pose estimation and vice versa, the main difficulty of SfM approaches is the accumulation of ***drift***. This comes from the fact that small errors in the estimated poses lead to larger errors in the 3D structure estimation, etc. This calls for additional measures apart from robust pose estimation techniques. One method is the fusion of the visual SfM with other sensor outputs, for example, in visual-inertial fusion. For visual-only systems, bundle adjustment optimization techniques are capable of reducing the drift.

Triangulation

Triangulation techniques are used in SfM in order to recover the 3D position $\boldsymbol{m}_w$ of points whose 2D image reprojection is observed in multiple camera views. The collinearity constraint defines that a 3D point, its projection on the image and the camera center are collinear. Using this constraint, two equations can be defined from each view of a 3D point. With a minimum of two views of the same point in two different images with known pose, a system of equations can be defined to estimate the position of the point in 3D. Several approaches to solving this problem were proposed, an overview and evaluation of which are available in Hartley and Sturm [48].

Bundle Adjustment

Bundle Adjustment (BA) is often used by SfM systems in order to reduce the accumulated drift over several frames and refine the estimated 3D structure [124]. It is an optimization procedure aiming to minimize an error indicator over the entire processed data (global BA) or at least over a batch of frames (local BA). In the case where the 3D structure is represented as a 3D point cloud, it is a minimization of the reprojection error of all 3D points, by adjusting the estimated camera poses and/or the estimated 3D positions of points, i.e.,

$$\min\,(\boldsymbol{R}_{cw}^{(j)}, \boldsymbol{w}_c^{(j)}\{\boldsymbol{m}_w^{(i)}\}) \sum_i \sum_j \|\boldsymbol{n}_p^{(i,j)} - \Pi(\boldsymbol{K}, \boldsymbol{\kappa}, \boldsymbol{R}_{cw}^{(j)}, \boldsymbol{w}_c^{(j)}, \boldsymbol{m}_w^{(i)})\|^2, \tag{3.31}$$

where i is the 3D point index and j the image index. As BA can be too computationally expensive for real-time systems, it is often done in parallel to the main system on a background thread [60].

3.2.8 Depth Image Tracking/ICP

When scene depth information is provided, by using, for example, depth sensors as described in Sect. 3.1.4, the information can be used for alignment with an existing scene model and estimation of the pose directly in 3D space. The ***Iterative Closest Point (ICP)*** algorithm is often used for the alignment of 3D point clouds [134]. Assuming two 3D point clouds, the algorithm attempts to iteratively find a transform to apply on one point cloud in order to minimize its alignment error with the other point cloud. The procedure involves finding closest point correspondences for each point, estimating a transform from them, and applying it on the points. Several improvements and methods for speeding up this procedure that tends to be computationally highly demanding were proposed [109].

Depth tracking can also be performed using dense image information. When two depth images or a depth image and dense reconstructed scene model are available, tracking techniques can perform alignment using the entire depth images. Similar to direct image alignment methods (see Sect. 3.2.6), the optimization procedure searches for a pose transform that minimizes the geometric error between the depth images. In the case of RGB-D systems combining depth and normal cameras, it is possible to optimize both the geometric and the photometric error jointly for improved results.

3.2.9 Deep Learning in Computer Vision

Deep learning is a field of machine learning that has seen a continuously increasing interest in recent years, mainly because of the success of deep learning methods in solving problems that appear intuitive to humans but are extremely challenging for computers as they cannot be easily described in a formal way [44]. Such examples of computer vision problems are object detection and classification or scene segmentation.

Deep learning techniques revolve around the idea of creating structures that learn complex representations as combinations of simpler concepts. The depth of such multilayered representations is also related to the origin of the name deep learning. The main concepts of deep learning are not new; however, the increase in computational capabilities combined with an increased availability of training data led to a resurgence in the use of neural networks in the beginning of the twenty-first century and to the conceptualization of the term deep learning [10, 91]. In the following, we discuss CNNs as one of the most successful and commonly used neural network types and fundamental network training approaches from the selection of loss functions to optimization algorithms.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs), introduced in LeCun et al. [65], are, as their name suggests, neural networks that perform convolution instead of matrix multiplication in at least one of their layers. These networks have shown great success in processing data with a grid-structured topology, especially 2D image data. The convolution operation of practical interest is the 2D convolution of an input image $\mathcal{I}$ with a 2D kernel $\mathcal{K}$, which is given by

$$\mathcal{S}_{(i,j)} = (\mathcal{I} * \mathcal{K})_{(i,j)} = \sum_m \sum_n \mathcal{I}_{(m,n)} \mathcal{K}_{(i-m,j-n)}. \tag{3.32}$$

The convolution output is often also referred to as the feature map. Note that the discrete convolution operation can also be represented as a specifically constrained sparse matrix. For the 2D convolution case that is a doubly block circulant matrix.

There are several arguments for the use of convolutional layers in deep neural networks. Compared with fully connected layers where every input unit interacts with each output, in convolutional layers ***sparse connectivity*** is achieved when the size of the kernel is smaller than that of the input. This allows detecting and focusing on small features, which is important when processing images and increases the efficiency of the network. Additionally, because of the ***equivariance*** of convolution to translation, when processing images, convolution can create a 2D map of the location where certain features appear in the input. Thus, the feature map representation follows the position of features in the input.

Typically in Convolutional Neural Networks (CNNss), convolution operations are combined with a detector stage and a pooling stage to form a convolutional layer. The detector stage typically contains a nonlinear activation function, for example, sigmoid, tanh, or rectified linear unit (ReLu) [82] to introduce nonlinearity properties to the network and the learned model. The pooling stage is meant to statistically summarize different nearby outputs of the previous stage to a single output. Examples thereof are max pooling (i.e., selecting the maximum value of a neighborhood) or a norm or weighted average of the output neighborhood. Pooling is meant to make the learned representation approximately invariant to small translations of the input and is also useful to reduce the size of the layer output.

In practice, several variants of the convolution function of Eq. (3.32) are used. For example, convolution with a stride is used when a downsampling of the input is required. Other variants are tiled convolution [66] or atrous convolution [24].

An example of a widely used CNNs is AlexNet [63]. It was presented in 2012 and was the first CNNs to win the ImageNet object recognition challenge. The network uses a series of convolutional layers, pooling and fully connected layers to encode the input image to a one-dimensional feature vector that is used for classification. The performance of AlexNet was quickly surpassed by more complex neural networks such as ZFNet [133], GoogleNet [119], VGGNet [114], and ResNet [49].

Different CNNs architectures targeting problems different than classification were also developed in recent years. Encoder-decoder networks first reduce the input to a compact feature level representation and then map expand this through a series of decoding layers back to an image-level output representation, to perform, for example, image segmentation [7] or depth estimation [64]. Siamese networks consist of two parallel branches that receive two images as input and are used for tasks such as tracking [11] or feature matching [76].

Network Training

In this section we discuss the entire procedure for ***training*** a neural network. In the general case, the purpose of training is to approximate a function $f^\star$ that maps the input data $\boldsymbol{x}$ to an output $\boldsymbol{y}$ so that $\boldsymbol{y} = f^\star(\boldsymbol{x})$. Assuming a network with multiple hidden layers each applying a function $f^{(i)}, i = 1, \cdots, N$ to its input, the function applied by the entire network will be $f(x) = f^{(N)}(f^{(\cdots)}(f^{(1)})))$. Training aims to learn a set of internal weighting parameters $\boldsymbol{\theta}$ that define the function $f(x)$ to best approximate the function $f^\star(x)$, using the available approximate examples of $f^\star(x)$ as training data. Thus, the main components in the process of training a neural network are the selection or generation of training data, the definition of a loss function defining the training accuracy, and the optimization procedure that is followed.

Training Data Ideally, it is desirable for available training data to sufficiently cover the entire range of the function $f^\star(x)$ that the neural network is expected to learn. This is however not always the case in practical situations and becomes harder as the dimension of the input increases to 2D or 3D data. Several techniques exist in order to cope with this, often going by the name ***dataset augmentation***. These can be, for example, alterations to an input image such as cropping and scaling or rotations and other transformations.

The requirement for larger training sets can be fulfilled by the use of ***synthetic data***, an approach that has seen increased usage, especially for computer vision applications. Synthetic datasets can be made indefinitely large and can be manufactured to cover the entire desired function range. However, training on synthetic data often leads to adaptation issues when the learned function is applied in the real world. This problem is commonly referred to as ***domain adaptation*** and has been researched over recent years [72, 100, 117]. The addition of artificial noise and other effects to the synthetic data in order to increase the similarity to real data is a straightforward solution to mitigate the domain adaptation accuracy loss.

A common requirement for trained deep learning networks is the ability to generalize or else to be able to make good predictions, even for input data that do not resemble the training data. Generalization ability would be the opposite to ***overfitting***, where the learned function perfectly fits the training set but does not respond well, even to small perturbations from it. The ability to generalize has been one of the main reasons for the recent success of deep learning. Techniques aiming to improve the generalization of the trained network are named regularization techniques. An overview on the large range of such techniques can be found in Goodfellow et al. [44].

Loss Function An important aspect of the training of deep learning networks is the choice of an appropriate loss function. The cost or loss function defines the error between the predictions done on the training set to the actual data and is used to guide the optimization of the network parameters during training. Therefore, it is important that the gradient of the cost function remains high during learning, or else that the function does not saturate.

Normally, the model will define a distribution $p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$ so that the principle of maximum likelihood can be used. Thus, the cross-entropy between the expected output on the training data and the prediction of the network can be used as a cost function. This cost function would be the negative log-likelihood.

A regularization term is often added to the cost functions, to assist the model in learning to generalize and avoid overfitting. This regularization term can be in its simplest form a function of the norm of the learned parameters $\boldsymbol{\theta}$ in order to contain them to small values during training.

Optimization The core procedure of training neural networks involves the minimization of a cost function through an optimization algorithm. Most of the algorithms still follow the principles of ***gradient descent*** optimization. The main idea of gradient descent is that the derivative of a function with respect to its input gives the slope at that point which can be used to find the direction in which the input can be shifted slightly in order to reduce the function output. The procedure can be followed iteratively so that a minimum of the function is found.

For a function $f(\boldsymbol{x})$ where the input is a vector $\boldsymbol{x}$, the gradient is the partial derivative with respect to all parameters given by $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$. Every step of gradient descent then suggests a new input $\boldsymbol{x}'$ that follows the negative direction of the gradient so that

$$\boldsymbol{x}' = \boldsymbol{x} - \epsilon \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) \tag{3.33}$$

where ϵ is the learning rate that should in general be small enough not to miss a minimal point and large enough to ensure fast convergence of the algorithm.

There are important differences in the practical training of a deep network compared with gradient descent optimization of a function. Computation of the gradient needs to be done with respect to the entire set of network parameters $\boldsymbol{\theta}$ whereas the amount of training data is typically too large to allow the evaluation of the gradient over the entire dataset at each step of the optimization. Stochastic Gradient Descent (SGD) is an extension of gradient descent that provides a solution for handling large datasets. Assuming a dataset of M training samples $\mathbb{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M\}$, a cost function $\mathcal{L}(\boldsymbol{\theta})$ can often be decomposed to a sum of costs over all training samples as

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^{M} L(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{\theta}), \tag{3.34}$$

where $L(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{\theta})$ is the cost for a specific sample. In order to apply gradient descent, a computation of the gradient with respect to the parameters $\boldsymbol{\theta}$ of the function for each sample is required to obtain

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{\theta}). \tag{3.35}$$

For a deep network with a large number of parameters and a large training dataset, the computational complexity for the calculation of the gradient to perform a single step of gradient descent becomes prohibitive. SGD is based on the observation that the gradient is an expectation that can be approximated using only a small, randomly drawn subset M' of the total number of training samples M termed batch or minibatch. This allows a gradient estimate to be computed

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})^{\star} = \frac{1}{M'} \nabla_{\boldsymbol{\theta}} \sum_{i=1}^{M'} L(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{\theta}), \tag{3.36}$$

which is used for the parameter update in SGD as

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - \epsilon \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})^{\star}. \tag{3.37}$$

Minibatches are typically only small fractions of the entire available training datasets, and it is common to train neural networks over multiple ***epochs*** (i.e., over several passes through the entire dataset).

The nonlinearity added in deep neural networks results in loss functions that are usually nonconvex, which means that gradient descent methods cannot guarantee global convergence of the optimization. Rather, decreasing the loss sufficiently in an acceptable training time is a realistic target of the training procedure.

The observation was made that it is beneficial for the training procedure to gradually decrease the learning rate in order to cope with noise induced by the random sampling of SGD as the true gradient of the cost is approaching 0. Therefore, algorithms commonly used in the training of deep neural networks are based on the SGD concept with some additional logic to control the learning rate. The AdaGrad algorithm [32] adapts the learning rates of all parameters by scaling them inversely proportionally to the square root of the sum of all historical values of the gradient. The RMSProp algorithm is a modification to the AdaGrad that changes the gradient summation into an exponentially weighted moving average in order to discard older values. The Adam [59] (named after Adaptive Moments) is another frequently used optimization algorithm that combines RMSProp with derivatives of the gradient (momentum of the gradient). The use of momentum is meant to accelerate the learning, especially in areas of high curvature by smoothing the descent.

Previously, we used the gradient of the cost function with respect to the network parameters $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ on every step of an optimization process based on SGD. However, the computation thereof is not trivial and involves a significant computational load for a deep neural network. To achieve this in an efficient manner so that the network training time remains acceptable, the ***back-propagation*** algorithm is used [107] that traverses through the network and iteratively computes the gradient.

The computation is based on the chain rule of calculus. Assume a function $y = g(x)$ and $z = f(g(x)) = f(y)$. Then the derivative of the final output z with respect to the input x can be given by

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}. \tag{3.38}$$

This can be extended to a vector input as

$$\nabla_x z = \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right)^{\mathsf{T}} \nabla_y z, \tag{3.39}$$

where $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$ is the Jacobian matrix of function $\boldsymbol{y} = g(\boldsymbol{x})$. The loss function of a network is thus computed by the back-propagation algorithm by concatenating the gradients of all layers. When a layer has multiple inputs, the gradients from both are simply added. As the same gradients can appear on different paths through the network, back-propagation algorithms need to manage this efficiently in order not to repeat the same computations. Additionally, deep networks contain different types of layers that require a specific computation or approximation of their gradient. In practice, layer implementations in common deep learning libraries provide their own gradient computation so that they can be seen as a black box by the back propagation algorithm.

Deep Learning in Tracking and Mapping

Although initially deep learning in computer vision was mostly concentrated on classification or detection problems, recently deep learning approaches expanded to other estimation and regression problems, providing alternatives to geometric approaches. For example, approaches that learn to estimate an object pose from a single image [100] or that learn to track by estimating the relative camera pose between two subsequent images have been introduced [128]. Deep learning-based feature detectors and matchers or generators of camera views or depth images are also widely researched [30, 64].

Generally, two directions in the use of deep learning in tracking and mapping are seen. On one side, there are ***end-to-end*** learned approaches that advocate the exclusive use of deep learning to solve a problem. Observing the limitations that occur in the ability of such systems to generalize to a universal scale, many new approaches turn to solutions that combine the output of deep learning approaches with traditional computer vision techniques in order to refine or increase the robustness of their results.

3.3 Model-Based Tracking

Tracking systems for AR can be coarsely divided into model-based approaches and SLAM approaches. Model-based approaches use preexisting 3D information of the environment, whereas SLAM approaches gradually reconstruct their environment while tracking it. In this section we will discuss model-based tracking systems that can rely either on 2D fiducial markers or some 3D representation (model) of specific objects used as tracking targets.

3.3.1 Marker Trackers

Markers are artificial visual patterns constructed with the purpose of being easily distinguishable by visual tracking systems. Marker-based trackers were widely used in AR applications, especially in the early years. Nowadays, markers are still used in AR as a simple solution for establishing a camera pose with minimum preparation overhead (see Fig. 3.12). Simplicity rises from the planar nature of markers and the fact that they can simply be printed on a paper and experimented with. However, markers are now considered obtrusive to the environment and confining to the AR application as the area of coverage is limited by the marker visibility. We will provide a brief overview of marker tracking techniques and existing systems.

Markers used in computer vision are typically square or circular 2D patterns. These geometric primitives are well-detectable in images and usually serve as an initial hint for the detection of a marker. Combining black and white textures provides strong gradients that facilitate detection and decoding. Square markers have the advantage that after detection the pose computation can be easily done using the four corners of the marker as point correspondences in a homography pose computation (see the section "Homography Pose Estimation"), assuming that the camera intrinsic parameters are known. The markers of ARToolkit [55] and ARTag [34] are examples of square markers used frequently in AR applications.

Using circular markers, the pose can be computed from the whole contour around the marker, which can be detected on an image using an ellipse detector. This makes circular markers more robust to occlusions but with a pose estimation

Fig. 3.12 A circular marker of Pagani [88] used for placing an AR augmentation

ambiguity in the single marker case [62]. In Naimark and Foxlin [81], at least four circular markers are required to estimate the camera pose. In Pagani [88], circular markers are discussed in detail, and an approach to single marker pose estimation with resilience to occlusions is presented.

It is often required for AR applications to be able to detect and uniquely identify multiple markers. Therefore, an identification pattern is also often added in many marker trackers. For example, in Kato and Billinghurst [55] a random binary pattern is added that can be matched using correlation matching techniques. Other approaches use structured patterns by dividing the marker area into a grid. In Wagner and Schmalstieg [127], the identification was made more robust by using error correction with checksums and forward error correction when reading the binary code.

Similar to the classical markers, it is also common to use a simple 2D image with rich texture as a marker. Visual features are extracted from the target image and their 3D locations computed considering a reference coordinate system on the image. In this case it is possible to compute the pose by applying feature matching techniques as in Sect. 3.2.1 and then track the pose using frame-to-frame feature trackers. This is an approach followed for example in the markers provided by Vuforia [5].

3.3.2 3D Model Trackers

An evolution over marker tracking approaches is trackers that use a 3D model representation of an object as their reference for estimating the pose of the real object. Such models can be textured 3D reconstructed or Computer-Aided Design (CAD) representations of the objects or more abstract models such as line and contour models depending on the tracker requirements. Compared with marker-based approaches, this has the advantage of avoiding the obstruction of the scene while also creating a direct correspondence to the object for the placement of AR content. The 6DoF transformation between the camera coordinate system and a coordinate system relative to the rigid object has to be estimated.

We first discuss frame-to-frame tracking approaches where the pose for a camera frame is estimated given the pose in a previous, highly similar frame. We analyze different alignment approaches that have been proposed over the years by tracking systems. However, pose estimation of an object without previous knowledge of its pose is an often overlooked but equally important problem, appearing, for example, in the initialization of tracking systems. Therefore, we also present existing approaches to solving this challenging problem. Of course, a single image pose estimation technique can also be used in a tracking by detection approach that does not use information of the previous frame. This typically negatively affects computation time and accuracy but prevents pose drift due to the accumulation of error over multiple frames.

Object Model Acquisition

The acquisition of object models is a very extensive topic on its own. Models of objects at different levels of accuracy and abstraction level can be created, depending on the requirements of the tracking system used. For example, high-quality 3D reconstructed models require dedicated setups and devices, whereas lower-quality models can be obtained by handheld ***scanning*** using inexpensive devices such as the Microsoft Kinect. In industrial manufacturing, CAD models are often available in different forms. For some trackers, simple object contour models are required. Some examples of object models can be seen in Fig. 3.13.

Feature Trackers

Feature-based object trackers extract natural features with strong edge characteristics from the object models and perform pose tracking on these, applying feature matching and tracking techniques and pose estimation from 3D/2D correspondences. This can be computationally efficient and robust in general; however, such approaches are only applicable on objects with an overall strong texture, thus limiting the range of possible applications. This also indicates that textured 3D models of the objects are required. A significant advantage of these methods is that they do not operate over the entire object, making it easier to handle occluded or partial object views.

The tracking system of Vacchetti et al. [126] uses a set of ***keyframes*** of the object provided by the user prior to the operation of the tracker. Using these keyframes as reference, matches to the current frame can be established resulting in 2D-3D correspondences. Although it is possible to estimate the pose relying only on these matches, this would not take into account previous frames, resulting in pose jitter. Therefore, the approach also takes into account neighboring frames and adds the tracked correspondences from the previous frame to an optimization problem to minimize the reprojection error. Thus, the approach combines the benefits of frame-to-frame tracking and tracking by detection.

In Park et al. [89], the same concept is followed using only a subset of all available object keyframes for matching. The frame-to-frame tracking is done by a local search and cross-correlation maximization. Using a thread for object detection and one dedicated to the tracking of each detected object allows for online handling of multiple objects.

In the tracking framework of Rambach et al. [99], preparatory steps are not required by the user. An automated procedure uses the object 3D model to discover feature locations suitable for tracking and to create a set of synthetic keyframes. The keyframes in this work are only used for ini-

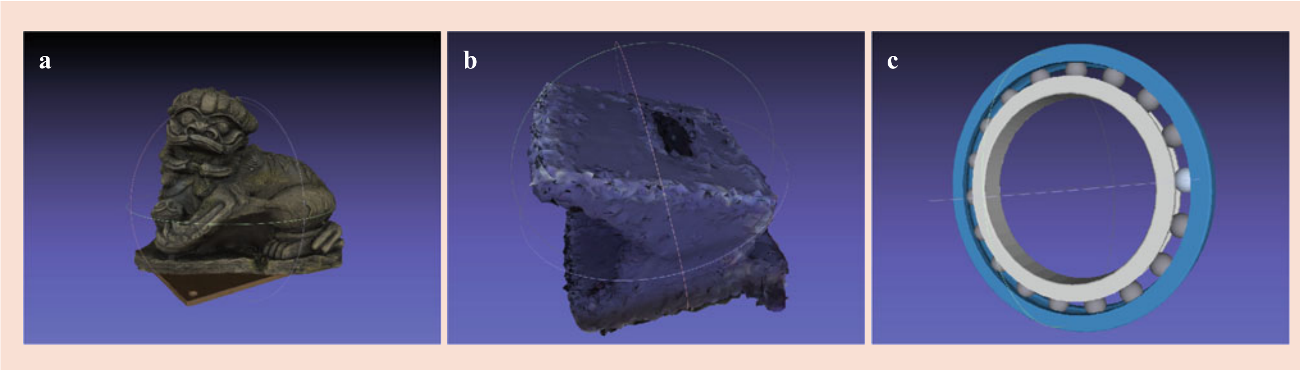


Fig. 3.13 (**a**) High-quality reconstruction with structured light approach. (**b**) Handheld Kinect scan object reconstruction. (**c**) CAD model

tialization and re-initialization of the frame-to-frame tracker. The tracking uses a two-step KLT from the previous real image to the current and then to a rendered image of the object model to eliminate drift. To increase the resilience of the tracking to illumination changes and to facilitate the matching between synthetic and real images, all images are preprocessed with pencil filter.

Edge Trackers

Edge tracking systems are based on the general line tracking concept presented in Sect. 3.2.5. The object model is rendered over the real object using the previous pose or a pose estimate, and correspondences for edges are searched for in a 1D search on the direction of the edge normal. The pose is then optimized iteratively using these correspondences. The main advantage of edge tracking methods over feature-based ones is the fact that objects without texture can be tracked given only their line model. On the other hand, these approaches often have difficulties when tracking is done in a cluttered environment, as wrong correspondences become more frequent. The number of iterations required to optimize the pose can also lead to an increased computational overhead. Additionally, pose initialization for these approaches was traditionally a problem because feature matching is not possible. Therefore, pose initialization had to be done manually by the user moving to an initialization pose. Recently, deep learning-based pose estimation has given a solution to this problem.

The initial edge tracking approach was introduced in Harris and Stennett [46]. This tracker, although being the first real-time object tracking system, was not entirely robust. Problems were encountered in cluttered environments, under occlusions and large displacements between frames. Several improvements were proposed in Drummond and Cipolla [31]. Handling of self-occlusions of the line model was done by rendering with a binary space partition. The use of robust M-estimators was used to decrease the influence of wrong edge correspondences on the cost function.

In Vacchetti et al. [125], the authors introduced the use of ***multiple hypotheses*** for edge correspondences. By modifying the robust estimator for each edge allowing multiple minimums, it is possible to consider more than one correspondence for each edge in the optimization. In Wuest et al. [132], the visual properties of the edges were examined for similarity to their matches instead of simply searching for strong gradients. Additionally, the control points along the contours were selected adaptively on the image projection instead of the 3D model in order to distribute them more evenly.

Finally, edge tracking systems are used in widely adopted commercial object tracking systems, for example, in tracking software provided by Vuforia [5] and VisionLib [4].

Direct Trackers

In contrast to feature or edge-based methods, region-based and direct trackers use information from the entire object surface. Thus, ***direct methods*** do not require strong texture features and are less affected by cluttered environments than edge-based trackers. However, direct tracking is less robust to illumination changes as the ***brightness constancy*** assumption between frames is often violated. Simultaneous object segmentation and pose tracking was proposed in [93]. The approach uses optimization to find the object pose that maximizes the posterior per-pixel probability of foreground and background membership. Because only the object contour is considered, the tracking is resilient to motion blur and noise. The approach is made real-time capable through a parallelized GPU implementation. Direct photometric error minimization for object pose tracking is applied [21, 113]. The main concern of such approaches is finding a suitable technique for dealing with the violations of photometric constancy between images, which is a common occurrence under illumination variations, especially for objects with reflective surfaces. Adjusting the photometric error minimization formulation (see Eq. (3.30)) with a multiplicative or additive term is a possible approach to allow specific intensity changes; however, Seo and Wuest [113] use a more sophisticated approach by assuming Lambertian reflectance properties on the object surface normals and therewith performing a prediction of the image intensity to assist the optimization.

Deep Learning-Based Trackers

The use of deep learning techniques for object tracking has also been proposed recently. Initially, in Garon and Lalonde [40] a Siamese deep neural network receiving the previous and current RGB-D images of the object as input and regressing the relative pose was presented. The network is mainly trained on synthetic rendered images of the objects; however, a refinement of the training using real data is required. Object-specific training of the network was needed in this version. In a later version, a multi-object and a generic variant was shown, with a loss in tracking accuracy [41]. In Li et al. [70], a network architecture that can be used to iteratively estimate the pose of an object by matching its 3D model to an object image is given. The approach is computationally demanding owing to the iterative nature. Finally, there are several approaches that tackle the problem in the form of ***pose refinement*** of an initial object pose on a single image [57]; however, the same concepts can be applied to frame-to-frame tracking using the previous pose as a coarse initial estimate as long as they real-time capable.

Hybrid Trackers

As seen so far, each approach for object tracking has its own advantages and disadvantages; therefore, combining different object tracking strategies for achieving improved results is a viable approach. For example, Vacchetti et al. [125] proposed the combination of feature and edge tracking. This is done by adding reprojection error terms for texture feature matches in the edge-based loss formulation described in Sect. 3.2.5. In Brox et al. [17], feature matching and optical flow tracking are combined with region-based correspondences for refinement of the estimated pose.

In Kehl et al. [58] pre-rendering of the 3D model from different poses and extraction of contour and interior information are performed to reduce the time needed for tracking each frame. The optimization of the contour segmentation is done by a sparse approximation to further reduce the overhead. In the case of the availability of depth data, optimization by point cloud alignment is done jointly with the RGB contour tracking approach. The RGB-D tracker of Tan et al. [120] combines learning using random forests to predict the pose change between two frames with a Gauss-Newton region-based optimization technique to align the model with the scene for refining the estimated pose. This combines robustness to large frame-by-frame movements with the increased accuracy of optimization model fitting.

Tracking by Detection/Pose Initialization

The ***object pose estimation*** problem consists of estimating the pose of an object without any prior knowledge of it. Such approaches are of importance mainly for the ***initialization*** of tracking systems. However, if such an approach is computationally efficient to allow real-time operation, it can also be used in a tracking-by-detection system where poses for every frame are computed independently without using prior information on the pose. In this case, the accumulation of pose drift from tracking over several frames is avoided but higher jitter is typically present in the computed poses.

The work of Hinterstoisser et al. [52] was focused on the detection and pose estimation of textureless objects in RGB-D images by employing learning on ***templates*** created using the object 3D model. Templates are created by a sampling of viewpoints around the object. The pose is computed by matching to a template based on color gradients and surface normals and then improving the pose using the depth images in an ICP method. The released dataset, often encountered as the LINEMOD dataset, is one of the most commonly used benchmark datasets for object pose estimation (see also Fig. 3.14).

Brachmann et al. [15] use uncertainty modeling of the estimated pose from RGB images in a multistage approach with a classification-regression forest, RANSAC pose estimation from 2D-3D correspondences, and subsequent refinement of the pose. Their approach is computationally demanding and necessitates training on real images of the objects, which requires a cumbersome collection of real data with ground truth pose.

Several approaches using deep neural networks have been introduced for object pose estimation. SSD-6D [57] is a CNNs-based approach that uses the network to place bounding boxes in images around the objects and then use their size to estimate depth. Rotation is obtained by viewpoint classification. The approach is highly dependent on a 2D ICP pose refinement technique following the initial estimate. Furthermore, the approach uses only synthetic data for training but is vulnerable to domain adaptation issues when color dissimilarity between the 3D model and the real object exists. A solution to bridge the domain gap between synthetic image training and real image network deployment was given in [100]. There it was proposed to use the pencil filter image preprocessing on both synthetic training images and real evaluation images in order to improve adaptation by allowing the network to focus on edge features rather than color, which is more volatile (see Fig. 3.15).

In BB8 [96] a first CNNs finds the bounding boxes, and a second CNNs estimates the 3D positions of the edges of the bounding box that are then used to estimate a 6DoF pose by solving a PnP problem. In Tekin et al. [122], a similar approach is used using however only one CNNs to directly

Fig. 3.14 Object 3D models of the LINEMOD dataset [52]

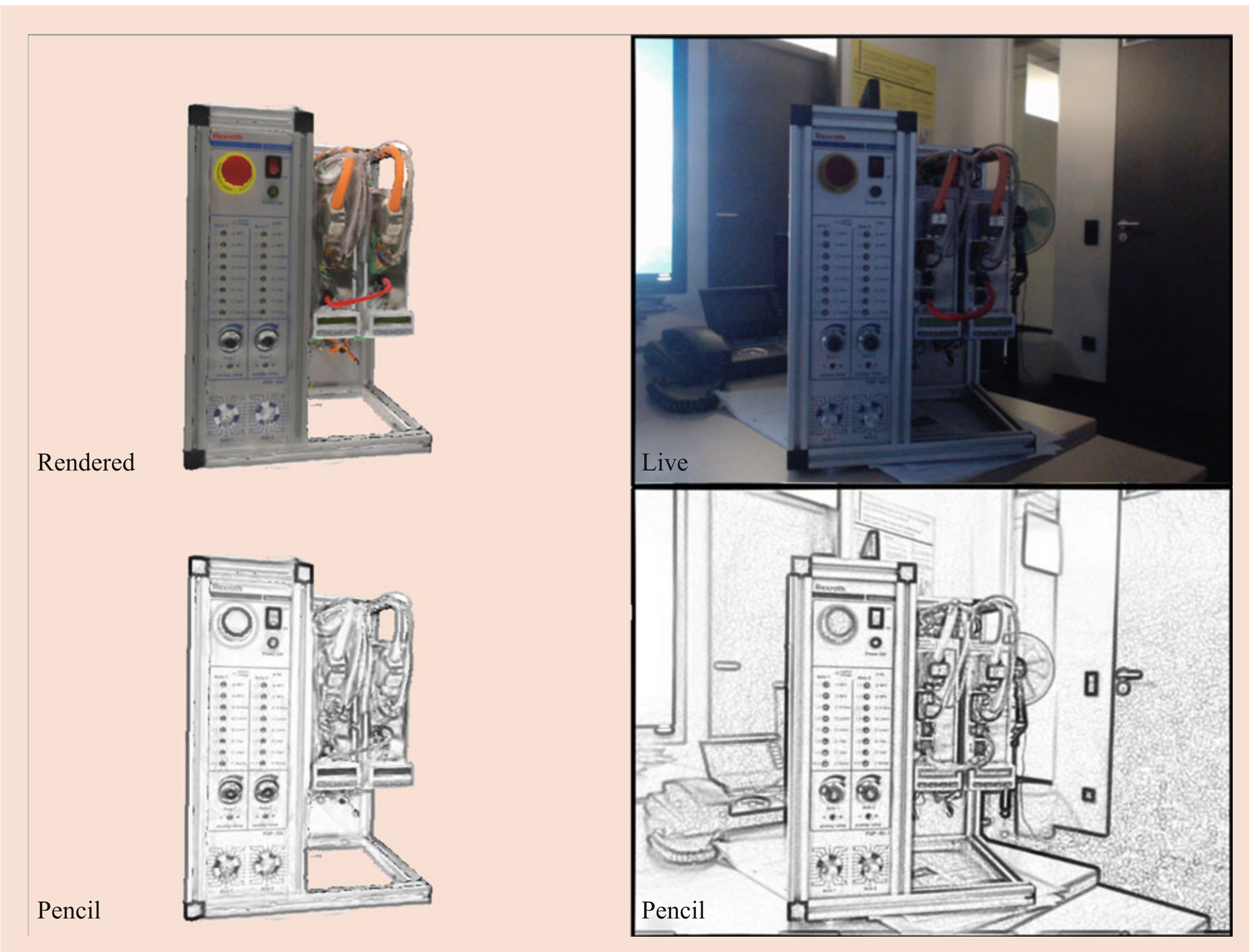


Fig. 3.15 The pencil filter of [99] can be used as a domain adaptation technique between synthetic images of object models and real images

estimate the 3D positions of the object bounding box corners. Both systems use real training images for their networks, which significantly simplifies the problem as the challenge of transferring training on synthetic images to the real world is avoided. Recently, Sundermeyer et al. [118] introduced an approach that uses synthetic training without requiring explicit mapping of images to object poses by training an autoencoder network to encode object orientation.

3.3.3 Nonrigid and Articulated Objects

Up to this point, the case of rigid object tracking was covered. However, there are cases where AR applications are required to track objects that are not rigid (i.e., their structure changes over time in a sequence) or objects that consist of multiple components. This adds further challenges to tracking systems that require different handling than the approaches discussed so far. Although this topic extends far from the scope of this chapter, we will mention a few cases where ***nonrigid*** tracking is of relevance for AR.

For industrial applications, AR-guided assembly of objects is a use case that has attracted significant interest over the years. Such applications require a tracker that can detect and track all the components of an object, as well as their combinations, which result in different object assembly states. Several approaches focused on a specific object and using markers on each component for tracking have been shown [97]. In Su et al. [115], a deep learning approach using a CNNs that can directly estimate the object assembly state and its 6DoF pose was proposed, relying only on synthetic data for training (Fig. 3.16). This means that the training data collection procedure can be automated given the object models of all components and a state graph showing their possible combinations. In Brox et al. [17], the authors show that their object tracking approach using contours and points can also extend to articulated objects.

Augmentation involving humans is another situation where nonrigid object tracking is performed. Many AR applications that detect ***faces*** and apply different augmentations on them have been introduced, several of them available in commercial products. Such approaches usually do not perform full 6DoF pose tracking, instead they use face alignment [19, 56], detecting facial landmarks (e.g., eyes, nose), and place their augmentations using these as a reference (see Fig. 3.16).

Even more challenging scenarios are encountered in medical AR applications, especially in AR-assisted surgery, where tracking targets such as internal organs are often highly nonrigid and textureless [22]. For example, the silhouette-based approach for pose estimation of organs for surgical AR given in Adagolodjo et al. [6] uses a combination of a biomechanical model of the organ with a set of deformation constraints. Local feature matching-based registration accepting multiple affine transformations for object regions was proposed in Puerto-Souza and Mariottini [94]. By adding neighboring features constraints, Paulus et al. [90] showed how cutting and tearing of objects can be correctly detected and displayed in AR.

3.4 SLAM

In contrast to model-based approaches, SLAM systems operate in previously unknown environments in which they are able to estimate the camera pose and simultaneously create a map of the environment. For AR, SLAM systems are of interest as there is no requirement for predefined tracking

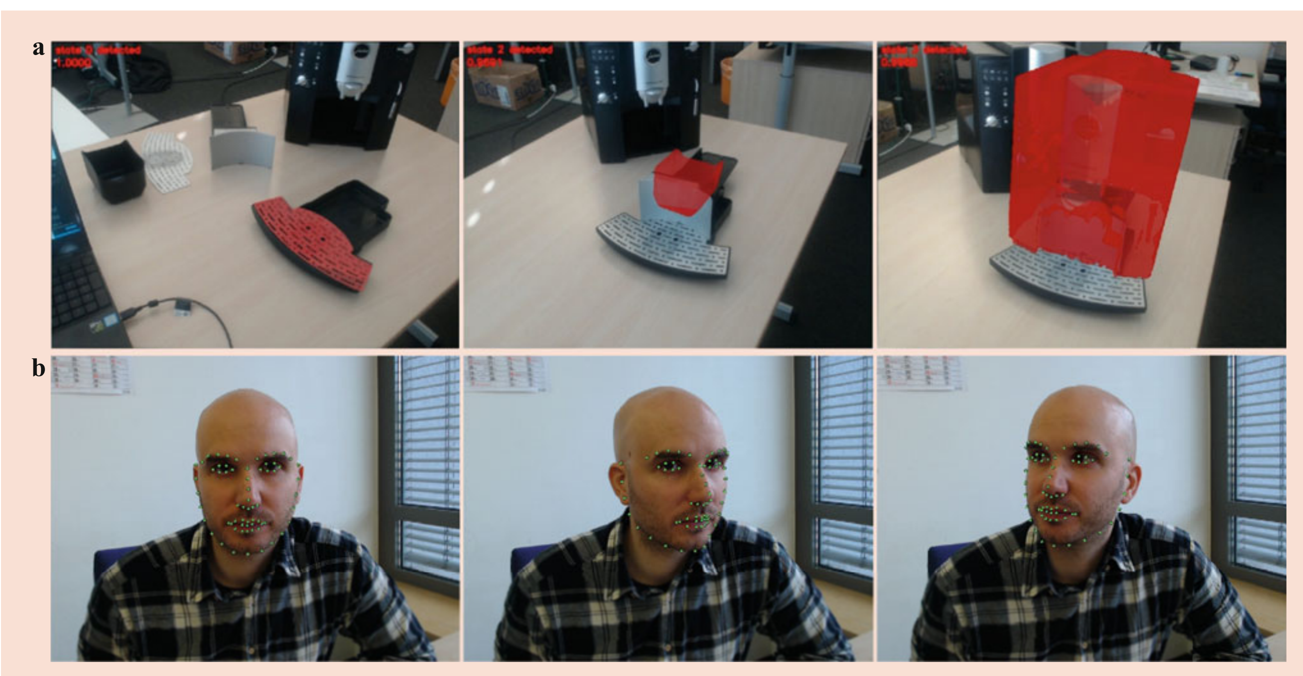


Fig. 3.16 (**a**) Multistate object tracking and AR assembly instructions [115]. (**b**) Face alignment output from Kazemi and Sullivan [56]

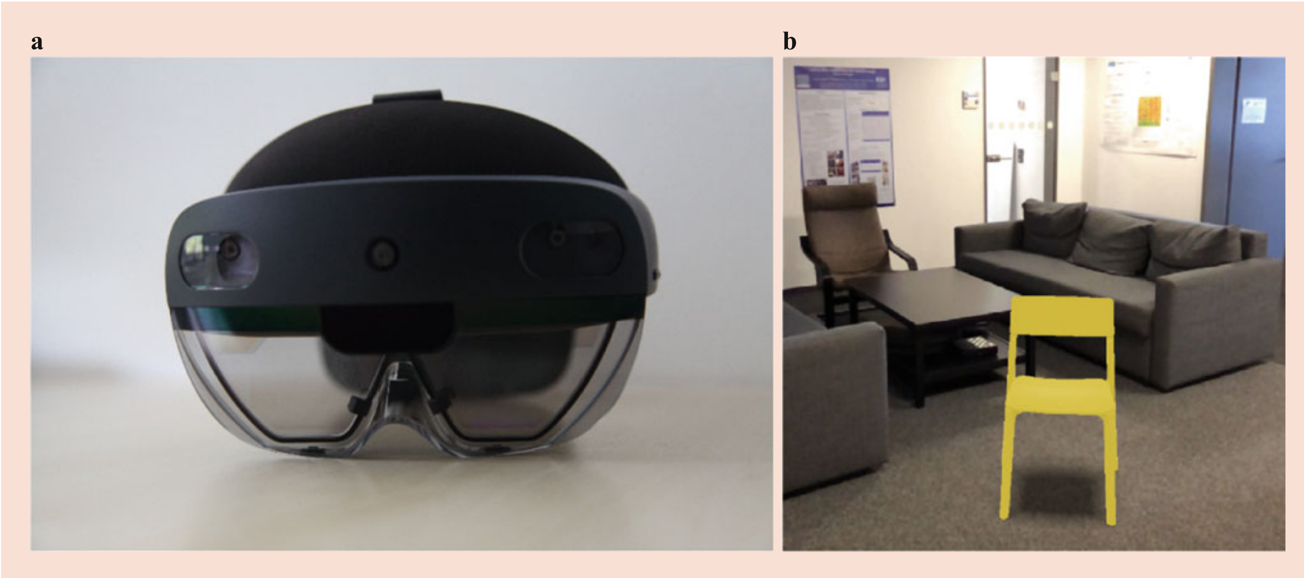


Fig. 3.17 Several commercial AR tracking solutions are based on SLAM systems. (**a**) Microsoft Hololens HMD [3], (**b**) Apple ARKit for smartphones [18]

targets; however, their use creates additional complications making them dependent on the type of planned application. In particular, apart from accurate localization in an unknown environment, a detailed ***mapping*** of this environment in a denser format than a 3D point cloud or at least partial knowledge of planar surfaces is required for the registration of AR visualizations in the environment and for interactions with it. Additionally, reconstruction and semantic understanding of objects are needed for applications that provide object-centered visualizations. We discuss different existing approaches for purely visual SLAM systems relying only on single or multiple camera inputs. As the SLAM problem is considered to be one of the most challenging vision problems, often the use of additional complementary sensors is preferred to assist the vision-based localization. For example, many commercial AR tracking SLAM solutions use sensor fusion (see Fig. 3.17). Thus, we also discuss the integration of inertial and depth sensors into SLAM-based tracking approaches. Finally, we look at the current main challenges in the SLAM domain and provide a future outlook on their importance for AR.

3.4.1 Visual SLAM

Visual SLAM using monocular or binocular camera images as input is a fundamental computer vision problem that has received significant attention. Many successful approaches have been introduced, each with its own strengths and weaknesses. It is generally accepted that the problem is inherently challenging. Apart from the common issues of vision-based systems that appear when the image quality decreases, the nature of SLAM requiring tracked features for 3D mapping while relying on mapping for the pose tracking leads to rapid accumulation of small errors that contribute to the drift of the computed pose. This is especially problematic when a monocular camera is used so that feature depth computation has to be performed using multi-view SfM techniques and when no additional sensors are available for correcting errors through sensor fusion. Additionally, high computational effort is involved in many techniques used in SLAM, making its deployment challenging in less powerful mobile end-user devices. Retrieving the real-world scale of the scene is not possible in visual-only SLAM unless a reference object is present in the scene or a machine learning technique for scale estimation is deployed. Based on the tracking technique used, a common classification of SLAM systems is into keypoint-based and direct SLAM, whereas it is also possible to characterize them by the type of mapping performed into sparse and dense systems.

SLAM Architecture/Main Components

SLAM systems typically consist of a group of cooperating modules, each being responsible for a specific task. The basic modules encountered in any system are modules for initialization, tracking, and mapping.

Initialization is required for creating an initial map as a starting point for the system. Using a small selection of camera frames with sufficient baseline and deploying a structure-from-motion technique using the relative camera poses from a fundamental matrix computation (see Sect. 3.2.4) is a frequently utilized technique to create an initial map of arbitrary scale [60]. As this technique does not perform well in planar scenes, it can be used in parallel with a homography-based computation assuming planarity [80]. Multi-camera systems can directly create an initial map without need for camera motion.

The tracking module is responsible for computing the camera pose with respect to the SLAM map on every frame given the pose of the previous frame. Keypoint-based [80] or direct techniques [84] can be used for this. Maintaining a correct pose at all times is highly important because the tracked poses are used by the mapping module. Tracking is a module with strong real-time constraints.

The mapping module attempts to map the tracked environment using multi-view geometry and optimization techniques. Mapping can operate at lower rates than the tracking; however, it should still be fast enough to prevent starvation of the tracker (i.e., failure of the tracker to go into fully non-mapped areas). Keypoint-based approaches use triangulation techniques between multiple views to reconstruct 3D points, whereas direct methods optimize a per-pixel depth estimate map. The ***keyframe*** concept is of importance in mapping. As it is very costly to perform mapping operations on all camera frames, frames with important information (keyframes) such as views of newly visited areas have to be selected. Selection of keyframes is another crucial task of SLAM systems as too many will lead to computational overhead and too few will lead to tracker starvation. Commonly systems rely on heuristic measures for the selection of keyframes, for example, the number of frames since the last keyframe or the amount of common features with other frames [80]. Some approaches suggest a restrained keyframe selection [60], whereas others prefer to select many keyframes and later reduce their number using a culling mechanism [80].

Apart from these basic modules, most visual SLAM systems also have a number of supporting modules that are however equally essential for the reliability and robustness of the system. A ***relocalization*** module is required for two cases. First, if the tracking is lost, it can be used to recover the camera pose within the known map. Second, when a mapped place is revisited but the pose has drifted, it allows for closing of loops. Relocalization can be done, for example, by matching against a visual bag of words vocabulary and then by feature matching [80] or using methods such as tree-based classifiers [60]. Mapping optimization is another important module to ensure global consistency of the map. Optimization can be done by bundle adjustment techniques optimizing the total reprojection error of several keyframes, or it can be initiated by the detection of a loop. An illustrative example of the main modules encountered in a visual SLAM system is given in Fig. 3.18.

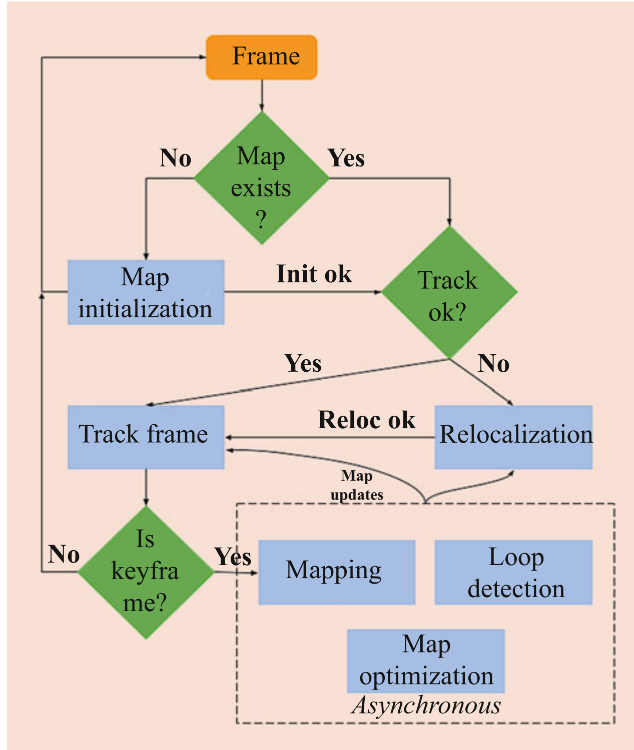


Fig. 3.18 Flow chart and main components of a visual SLAM system

Keypoint-Based SLAM

Initial SLAM systems were based on keypoints for tracking the camera pose from frame to frame and also for creating a map by triangulating points from different views. ***Keypoint-based*** SLAM is advantageous in the sense that keypoint processing is in general less computationally demanding and more robust to image degradation effects such as illumination variation, noise, and blurring than approaches utilizing the whole image. However, such approaches have the limitation of requiring sufficient texture in the tracked environment. Furthermore, keypoint-based approaches generally only reconstruct a sparse 3D point cloud map of their environment. Although this reduces complexity, making such approaches attractive for mobile device implementations, it also limits the ability for interaction with the environment for AR applications. To deal with this, additional techniques that restore parts of the map densely using the 3D point cloud as a starting point need to be employed [26, 101].

One of the first monocular SLAM systems proposed was MonoSLAM [28]. The approach used an Extended Kalman Filter (EKF) that estimates camera motion and 3D structure of the environment jointly. A motion model is used for the filter prediction step, and tracked feature positions are used for the correction step of the filter. An issue with this

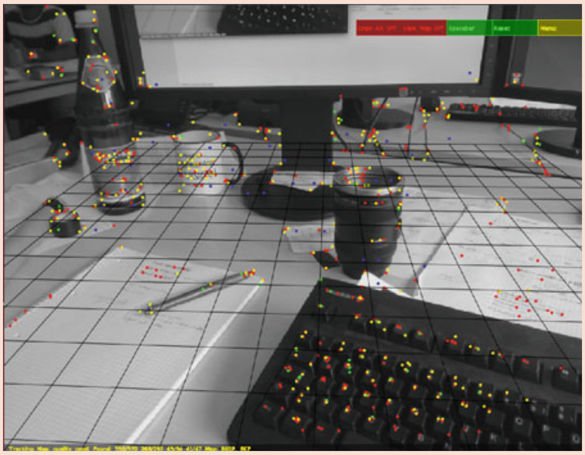


Fig. 3.19 PTAM SLAM system [60] creates a 3D point cloud map and estimates a dominant plane for AR augmentations

approach and similar filter-based approaches was ***scalability***. The increase in the map through the addition of new points leads to the increase in the filter state size and thus a rapid increase of computational complexity. Therefore, filter-based methods are only viable in limited areas.

An important development in visual SLAM was PTAM [60]. The tracking and mapping were decoupled and executed by different threads. This ensured that tracking could be done in real time with minimum overhead, whereas the mapping operations would be done in a background thread and only on selected keyframes. The mapping was optimized using bundle adjustment techniques for the first time both locally and also over the entire map (global bundle adjustment). The resulting SLAM system was able to track over a larger area (the target was functionality over small AR workspaces as shown in Fig. 3.19), and its paradigm of multi-threaded SLAM was followed by many subsequent approaches. A downsized mobile version of PTAM was presented in Klein and Murray [61]. The system was also used as a base for the visual tracking part of the multisensory Microsoft Hololens [3].

ORB-SLAM [80] later expanded on these concepts further to propose a monocular system able to perform on larger areas including long outdoor sequences. A ***loop closing*** optimization thread was added along with improved management of map keyframes and points using a pose graph. An example of the SLAM output is shown in Fig. 3.20.

Direct SLAM

Direct approaches do not match handcrafted features (i.e., keypoints) but instead perform matching directly on images using approaches that assume photometric consistency either over entire images or selected image regions. These methods are advantageous in that they allow for increased frame-to-frame tracking accuracy and higher density mapping. The downside is in general the increased computational requirements, especially for approaches that apply full image alignment and the sensitivity to image effects, especially lighting variations. DTAM [84] was one of the first fully direct SLAM approaches. The proposed system reconstructs a fully ***dense map*** and tracks by using an optimization-based full image alignment between the 3D model of the map and the current image. The system was able to run in real time on a parallelized GPU implementation. Still, this comes at an increased computational load and limitation to the size of the area that can be mapped. However, this approach demonstrated the value of dense mapping for AR by showing how this can be used for interaction of virtual content with the environment.

LSD-SLAM [33] proposed the idea of semi-dense SLAM. A direct approach to tracking was used here as well; however, instead of aligning the entire image, the approach focuses only on high-gradient image regions. This reduces the computational effort and improves the map depth quality by not considering the problematic textureless areas. Pose-graph optimization was deployed to ensure longer range functionality of the system. Although this results in higher accuracy, if fully dense mapping is needed, additional techniques have to be used, for example, filling of textureless regions assuming planarity and color homogeneity [26] or machine learning depth estimation [121].

A ***semi-direct*** approach that combines ideas from direct and keyframe-based SLAM was proposed in SVO [37]. This system uses direct photometric matching on image patches around reconstructed 3D points. Thus, tracking is done by a direct method, whereas mapping creates 3D points from multiple views using an uncertainty-aware depth filter that makes use of the entire tracked path of a feature. This results in a fast and accurate SLAM system that only creates a sparse map. Therefore, its main application area is autonomous navigation of micro air vehicles.

3.4.2 Visual-Inertial SLAM

The combination of visual tracking systems with inertial sensors is a very popular technique. The reasons for this are the complementary properties of cameras and inertial sensors. Camera images can provide very accurate measurements by matching of visual features and multi-view geometry.

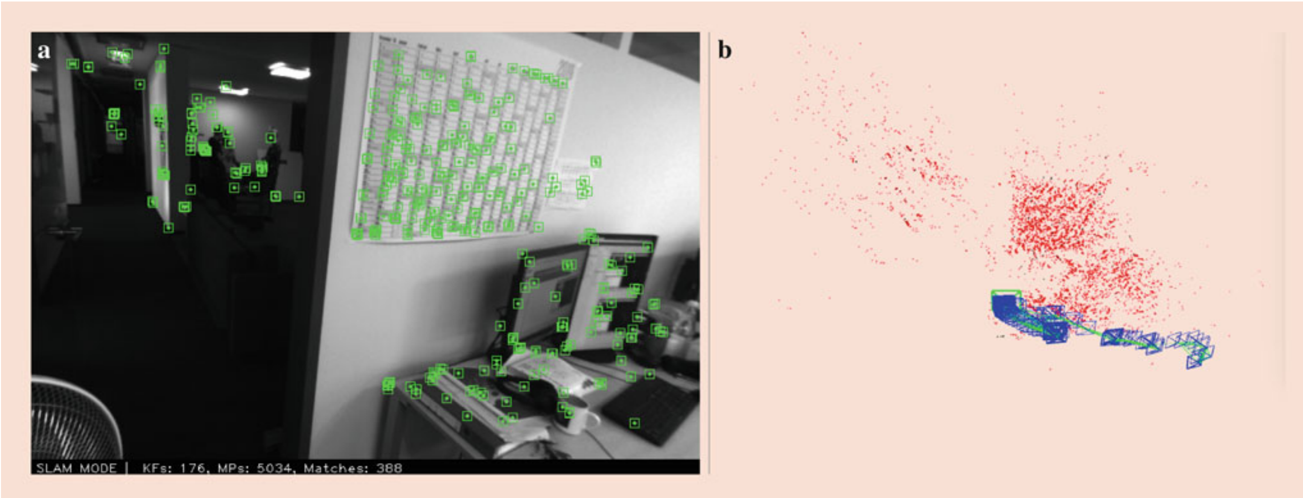


Fig. 3.20 ORB-SLAM tracking and mapped 3D point cloud with keyframe poses

However, when the quality of the image is compromised, owing to blurring caused by fast motion or sudden illumination changes, purely visual tracking systems tend to fail. Inertial sensors on the other hand are not affected by the image quality and are especially capable of tracking fast motion at high frequencies. The measurements of an IMU suffer nevertheless from high noise levels and drift over time. Challenging aspects of ***visual-inertial fusion*** include the requirement of calibrating the camera and inertial sensor in order to align the measurements in the same coordinate system (hand-eye calibration). Sensor synchronization up to a millisecond accuracy is needed; therefore, visual-inertial approaches function optimally in integrated camera-inertial sensor devices built for the specific purpose of tracking.

Existing approaches can be coarsely classified into filtering-based and optimization-based approaches. ***Filtering-based*** fusion employs statistical filtering techniques such as the extended or unscented Kalman filter (EKF, UKF) or particle filter (PF) [12, 42]. Filtering approaches typically use IMU measurements for performing multiple time updates of the filter and then use visual measurements to correct the filter state and compensate for the drift of inertial navigation. The EKF especially has been widely used in this problem in many forms [12, 69]. Different models exist for the integration of inertial measurements in the filter state, for example, assuming no motion and using accelerometer readings only for estimation of gravity direction to stabilize the orientation estimation or using a motion model and accelerometer and gyroscope readings as control inputs to the filter or including them in the filter state [12].

Optimization-based approaches define a cost function including reprojection residuals and an inertial measurements term and typically optimize that over a number of consecutive frames. ***Pre-integration*** of inertial measurements, summarizing multiple IMU measurements to an equivalent one, significantly aided optimization- based approaches [38]. This led to the development of SLAM systems such as OKVIS [68] and VINS-Mono [95].

Another classification of visual-inertial approaches is into tightly and ***loosely coupled*** systems. Loosely coupled systems perform inertial and visual tracking in parallel and then apply a fusion scheme that combines poses from both trackers. On the other hand, tightly coupled systems directly fuse output measurements of the sensors such as accelerometer and gyroscope readings from an IMU and 2D/3D feature correspondences from visual sensors. Loosely coupled systems are better at overcoming complete failures of one of the sensors, whereas tightly coupled systems use all the available information from the sensors and keep track of errors and covariances of the measurements to achieve superior results. An in-depth analysis and evaluation of many existing visual-inertial systems can be found in Chen et al. [23].

Visual-inertial fusion techniques for tracking have been applied in several successful commercial AR systems such as the Microsoft Hololens Head-Mounted Display (HMD) [3] and the ARCore [2] and ARKit [18] tracking frameworks for smartphones.

3.4.3 RGB-D SLAM

The use of depth cameras for SLAM offers significant advantages compared with the monocular approaches. Direct depth estimates are provided, which alleviates the need for visual feature matching and multi-view techniques. Dense mapping is therewith simplified even for untextured areas and under challenging illumination conditions. Additionally, dense cameras solve the scale ambiguity problem as they provide measurements in metric space. Consumer depth cameras are available at a relatively low cost; however, their functionality is limited to a specific range of depth and to indoor usage.

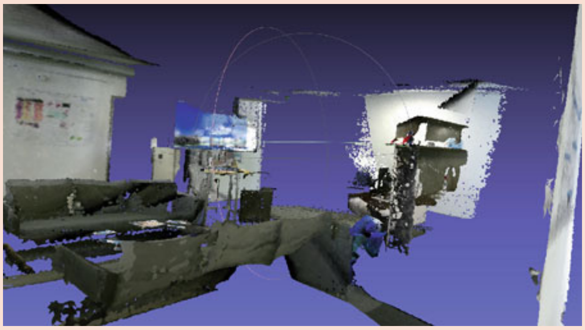


Fig. 3.21 Indoor mapping created using the RGB-D SLAM system of Whelan et al. [131]

RGB-D SLAM systems use depth images to support visual SLAM. Using the Microsoft Kinect, one of the first commercially available low-cost cameras, the KinectFusion RGB-D SLAM approach was presented in Newcombe et al. [83]. Scenes are reconstructed using the depth maps obtained from different poses whereas tracking is performed using an ICP algorithm that aligns the current depth map to the existing scene model. The RGB images can be used to obtain an initial estimate of the camera pose from feature matching that is then refined by the ICP depth alignment. The large volume of data for processing and storing is in general an issue of RGB-D systems. Therefore, dense map representations that reduce this overhead have been introduced. In Salas-Moreno et al. [111], the map complexity is reduced by merging points to planes when possible. Plane representations are also advantageous for AR, allowing alignment of virtual content. Subsequent RGB-D systems focused on the optimization of the mapping and the extension of the range of the SLAM system [130, 131] (see Fig. 3.21).

3.4.4 Deep Learning for SLAM

Deep learning approaches have been very influential in SLAM research over the previous years. Initially, some end-to-end learning solutions for SLAM were proposed; however, recently, it seems that more effort is being placed into the combination of deep learning modules with traditional geometric constraint reasoning.

A widely explored idea is that of ***learned depth*** estimation. This consists of training an encoder-decoder neural network to generate a depth image using an RGB image as input, thus basically simulating a depth camera. Several approaches have been proposed, achieving impressive results although the remaining question is the ability of such systems to generalize well in all possible environments. A representative approach was presented in Tateno et al. [121]. The approach fuses the depth map of LSD-SLAM [33] with the depth images generated by the network. This allows the semi-dense map to be completed with depth estimates of textureless regions but also provides a solution to the scale problem, which is essential for AR. Monocular visual SLAM is a map-scale agnostic, but the deep network estimate can be up to scale if the training data are as well. Additionally, in contrast to monocular SLAM, the approach survives under pure rotational motion using only the estimated depth from the CNNs. Managing large dense maps for SLAM still remains an issue for the long-term operation of the system.

In Bloesch et al. [13] the use of a trained auto-encoder was proposed to learn a compact representation of a scene in order to reduce the overhead of storing dense maps. In this approach it is possible to perform an efficient joint optimization at a level of encoded keyframes. In Rambach et al. [101], efficient piece-wise dense mapping is achieved using a network that predicts planar areas in images and fusing this with the 3D point cloud of a keypoint-based SLAM system.

Approaches that treat tracking or mapping as an end-to-end problem that can be learned by a neural network have also been proposed. In Wang et al. [128], a deep recurrent convolutional network (RNN) is trained to estimate camera poses of an image sequence, thus learning the odometry part of a SLAM system. Even though the results were encouraging, training and evaluation were done on sequences from the same vehicle dataset (KITTI) with limitations to the motion degrees of freedom, so that the ability to generalize still remains an open question. In Zhou et al. [138], using neural networks for both tracking and mapping is proposed. The tracking network performs small-step incremental frame-to-frame tracking to align the current image to a synthetic viewpoint, leading to better generalization ability. The mapping network is trained to optimize the depth estimates from several keyframes.

Deep learning has been applied in the visual-inertial fusion problem as well. A long short-term memory (LSTM) network that updates the camera pose using inertial measurements was presented in Rambach et al. [98]. The network prediction is used for the time update step of a Kalman filter and fused with the vision-based pose estimate. The approach is able to encompass sensor and hand-eye calibration and integration, but training is device specific. In Clark et al. [25], an end-to-end learned fusion approach is presented. Frame-to-frame tracking is performed by a Siamese network and inertial tracking by an LSTM. Each produces a feature vector that is concatenated and a pose update is regressed.

Towards Semantic SLAM

Apart from dense mapping that enables interaction in AR, ***semantic understanding*** of scenes can enable truly intelligent AR applications that can provide object-specific augmentations in unknown environments. It is envisioned that future semantic SLAM systems will be able to perform this type of dense mapping with semantic labeling. Already, significant effort and steps in that direction were made. Deep learning has played an important role in that because CNNs-based approaches have been very successful at detection, classification, and segmentation problems.

Existing semantic SLAM approaches typically add semantic labels produced by a segmentation network to every element of a dense map. These labels are refined in a probabilistic framework as new keyframes are added to the SLAM system. Examples thereof are the learned depth system of Tateno et al. [121] and the RGB-D approach of McCormac et al. [77]. However, such techniques add an additional overhead to the dense map and do not take advantage of the semantic cues for the improvement of the tracking and mapping modules. The work of Zhi et al. [137] showed promising results by adding semantic labels to their learned encoded scene representation, making it possible to jointly optimize motion, geometry, and semantics.

A different direction toward semantic mapping is given by approaches that attempt to directly map the environment directly at an ***object level*** instead of performing dense reconstruction and then adding semantic labels. This was initially proposed in Salas-Moreno et al. [110], where objects were detected and their 3D models were used to populate the SLAM map. This however requires that 3D models of objects exist in advance. Some recent RGB-D methods such as in McCormac et al. and Runz et al. [78, 108] were able to create object maps without previous knowledge of object models. Their work is based on Mask RCNN, a deep network that is able to detect, segment, and classify objects. Its output is combined and iteratively fused with the dense map of an RGB-D SLAM in order to isolate the 3D shapes of objects. Although encouraging results were achieved and some initial concepts showing the application of object-level semantic SLAM in AR have been described [108], a complete ***high-level representation*** used in the entire SLAM system, which would alleviate the need for pixel-level matching techniques and increase the range and robustness of SLAM, has not yet been developed.

3.5 Conclusion

In this chapter, we first analyzed basic concepts of computer vision, sensor fusion, and machine learning related to AR tracking systems. Subsequently, we provided an overview of established tracking systems, starting from object tracking methods using 3D models and moving to SLAM approaches without prior knowledge of the environment. We covered approaches using diverse sensor systems, from monocular and stereo cameras to depth cameras and inertial and geospatial sensors. We considered approaches based on traditional computer vision and statistical filtering techniques but also newer approaches using deep neural networks as well as combinations of both. Currently, the focus of research efforts is shifting to the enrichment of information provided by the SLAM system. In particular, the importance of dense scene reconstruction and semantic labeling is continuously stressed, together with the increase in the range of SLAM systems and ability of collaborative mapping. Object pose estimation and tracking remain equally important for many AR applications, and issues such as occlusion and object symmetry as well as estimation of object state and tracking of nonrigid objects are at the forefront. These seemingly separate directions (object pose and SLAM) are already converging and might coincide in the future as we are moving toward a system that is able to obtain a fully dense semantic understanding of the environment.

References

1. (2014). http://www.cs.cmu.edu/~kaess/vslam_cvpr14/media/vslam-tutorial-cvpr14-p12-densevo.pdf.
2. Google ARCore (2018). https://developers.google.com/ar
3. Microsoft Hololens (2018). https://www.microsoft.com/en-us/hololens
4. VisionLib (2018). https://visionlib.com
5. Vuforia (2018). https://www.vuforia.com
6. Adagolodjo, Y., Trivisonne, R., Haouchine, N., Cotin, S., Courtecuisse, H.: Silhouette-based pose estimation for deformable organs application to surgical augmented reality. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 539–544. IEEE, New York (2017)
7. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
8. Baker, S., Bennett, E., Kang, S.B., Szeliski, R.: Removing rolling shutter wobble. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2392–2399. IEEE, New York (2010)
9. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. Int. J. Comput. Vis. **92**(1), 1–31 (2011)
10. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. In: Proceedings of the Conference on Neural Information Processing Systems, pp. 153–160 (2007)
11. Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A., Torr, P.H.S.: Fully-convolutional Siamese networks for object tracking. In: Proceedings of the European Conference on Computer Vision, pp. 850–865. Springer, Berlin (2016)
12. Bleser, G.: Towards visual-inertial slam for mobile augmented reality. Verlag Dr. Hut, Germany (2009)
13. Bloesch, M., Czarnowski, J., Clark, R., Leutenegger, S., Davison, A.J.: CodeSLAM-learning a compact, optimisable representation for dense visual SLAM. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2560–2568. IEEE, New York (2018)
14. Bouguet, J.Y.: Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. Intel Corporation **5**(1–10), 4 (2001)
15. Brachmann, E., Michel, F., Krull, A., Ying Yang, M., Gumhold, S., et al.: Uncertainty-driven 6d pose estimation of objects and scenes from a single RGB image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3364–3372. IEEE, New York (2016)

16. Bradski, G.: The OpenCV Library. Dr Dobb's J. Softw. Tools **25**(11), 120–123 (2000)
17. Brox, T., Rosenhahn, B., Gall, J., Cremers, D.: Combined region and motion-based 3D tracking of rigid and articulated objects. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 402–415 (2009)
18. Buerli, M., Misslinger, S.: Introducing ARKit-Augmented Reality for iOS. In: Proceedings of the Apple Worldwide Developers Conference, pp. 1–187 (2017)
19. Bulat, A., Tzimiropoulos, G.: How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In: Proceedings of the IEEE International Conference on Computer Vision. IEEE, New York (2017)
20. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: Brief: Binary robust independent elementary features. In: Proceedings of the European Conference on Computer Vision, pp. 778–792. Springer, Berlin (2010)
21. Caron, G., Dame, A., Marchand, E.: Direct model based visual tracking and pose estimation using mutual information. Image Vis. Comput. **32**(1), 54–63 (2014)
22. Chen, L., Day, T.W., Tang, W., John, N.W.: Recent developments and future challenges in medical mixed reality. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 123–135. IEEE, New York (2017)
23. Chen, C., Zhu, H., Li, M., You, S.: A review of visual-inertial simultaneous localization and mapping from filtering-based and optimization-based perspectives. Robotics **7**(3), 45 (2018a)
24. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018b)
25. Clark, R., Wang, S., Wen, H., Markham, A., Trigoni, N.: Vinet: Visual-inertial odometry as a sequence-to-sequence learning problem. In: Proceedigs of the AAAI Conference on Artificial Intelligence (2017)
26. Concha, A., Civera, J.: DPPTAM: Dense piecewise planar tracking and mapping from a monocular sequence. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, ART-2015-92153. IEEE, New York (2015)
27. Davison, A., Murray, D.: Simultaneous localization and map-building using active vision. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 865–880 (2002)
28. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: MonoSLAM: Real-time single camera SLAM. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 1052–1067 (2007)
29. Dementhon, D.F., Davis, L.S.: Model-based object pose in 25 lines of code. Int. J. Comput. Vis. **15**(1–2), 123–141 (1995)
30. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 224–236. IEEE, New York (2018)
31. Drummond, T., Cipolla, R.: Real-time visual tracking of complex structures. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 932–946 (2002)
32. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. **12**(Jul), 2121–2159 (2011)
33. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-scale direct monocular SLAM. In: Proceedings of the European Conference on Computer Vision, pp. 834–849. Springer, Berlin (2014)
34. Fiala, M.: Artag revision 1, a fiducial marker system using digital techniques. Natl. Res. Counc. Publ. **47419**, 1–47 (2004)
35. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
36. Fleet, D., Weiss, Y.: Optical flow estimation. In: Handbook of mathematical models in computer vision, pp. 237–257. Springer, Berlin (2006)
37. Forster, C., Pizzoli, M., Scaramuzza, D.: SVO: Fast semi-direct monocular visual odometry. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 15–22. IEEE, New York (2014)
38. Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D.: IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. Georgia Institute of Technology, New York (2015)
39. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. IEEE Trans. Pattern Anal. Mach. Intell. **25**(8), 930–943 (2003)
40. Garon, M., Lalonde, J.F.: Deep 6-DOF tracking. IEEE Trans. Vis. Comput. Graph. **23**(11), 2410–2418 (2017)
41. Garon, M., Laurendeau, D., Lalonde, J.F.: A framework for evaluating 6-DOF object trackers. In: Proceedings of the European Conference on Computer Vision, pp. 582–597. Springer, Berlin (2018)
42. Gemeiner, P., Einramhof, P., Vincze, M.: Simultaneous motion and structure estimation by fusion of inertial and vision data. Int. J. Robot. Res. **26**(6), 591–605 (2007)
43. Getting, I.A.: Perspective/navigation-the global positioning system. IEEE Spectr. **30**(12), 36–38 (1993)
44. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT Press, Cambridge (2016)
45. Hager, G.D., Belhumeur, P.N.: Efficient region tracking with parametric models of geometry and illumination. IEEE Trans. Pattern Anal. Mach. Intell. **20**(10):1025–1039 (1998)
46. Harris, C., Stennett, C.: Rapid-a video rate object tracker. In: Proceedings of the British Machine Vision Conference, pp. 1–6 (1990)
47. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference, vol. 15, pp. 10–5244 (1988)
48. Hartley, R.I., Sturm, P.: Triangulation. Comput. Vis. Image Underst. **68**(2), 146–157 (1997)
49. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. IEEE, New York (2016)
50. Heikkila, J., Silven, O.: A four-step camera calibration procedure with implicit image correction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1106–1112. IEEE, New York (1997)
51. Hesch, J.A., Roumeliotis, S.I.: A direct least-squares (DLS) method for PnP. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 383–390. IEEE, New York (2011)
52. Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N.: Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In: Proceedings of the Asian Conference on Computer Vision, pp. 548–562. Springer, Berlin (2012)
53. Huber, P.J.: Robust statistics. In: International Encyclopedia of Statistical Science, pp. 1248–1251. Springer, Berlin (2011)
54. Jin, H., Favaro, P., Soatto, S.: Real-time feature tracking and outlier rejection with changes in illumination. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 1, pp. 684–689. IEEE, New York (2001)
55. Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings of the IEEE/ACM International Workshop on Augmented Reality, pp. 85–94. IEEE, New York (1999)
56. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874. IEEE, New York (2014)
57. Kehl, W., Manhardt, F., Tombari, F., Ilic, S., Navab, N.: SSD-6D: Making RGB-based 3D detection and 6D pose estimation great

again. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 22–29. IEEE, New York (2017a)
58. Kehl, W., Tombari, F., Ilic, S., Navab, N.: Real-time 3D model tracking in color and depth on a single CPU core. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 745–753. IEEE, New York (2017b)
59. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980 (2014)
60. Klein, G., Murray, D.: Parallel tracking and mapp.ing for small AR workspaces. In: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 225–234. IEEE, New York (2007)
61. Klein, G., Murray, D.: Parallel tracking and mapping on a camera phone. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 83–86. IEEE, New York (2009)
62. Köhler, J., Pagani, A., Stricker, D.: Detection and identification techniques for markers used in computer vision. In: Proceedings of the Visualization of Large and Unstructured Data Sets-Applications in Geospatial Planning, Modeling and Engineering, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2011)
63. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the Conference on Neural Information Processing Systems, pp. 1097–1105 (2012)
64. Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: Proceedings of the International Conference on 3D Vision, pp. 239–248. IEEE, New York (2016)
65. LeCun, Y., et al. Generalization and network design strategies. Connectionism Perspect. **19**, 143–155 (1989)
66. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: Proceedings of the IEEE International Symposium on Circuits and Systems, pp. 253–256. IEEE, New York (2010)
67. Lepetit, V., Moreno-Noguer, F., Fua, P.: EPnP: An accurate o(n) solution to the PnP problem. Int. J. Comput. Vis. **81**(2), 155 (2009)
68. Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P.: Keyframe-based visual–inertial odometry using nonlinear optimization. Int. J. Rob. Res. **34**(3), 314–334 (2015)
69. Li, M., Mourikis, A.I.: High-precision, consistent EKF-based visual-inertial odometry. Int. J. Rob. Res. **32**(6), 690–711 (2013)
70. Li, Y., Wang, G., Ji, X., Xiang, Y., Fox, D.: DeepIM: Deep iterative matching for 6D pose estimation. In: Proceedings of the European Conference on Computer Vision, pp. 683–698. Springer, Berlin (2018)
71. Liang, C.K., Chang, L.W., Chen, H.H.: Analysis and compensation of rolling shutter effect. IEEE Trans. Image Process. **17**(8), 1323–1330 (2008)
72. Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. arXiv preprint arXiv:150202791 (2015)
73. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature **293**(5828), 133 (1981)
74. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157. IEEE, New York (1999)
75. Lucas, B.D., Kanade, T., et al. An iterative image registration technique with an application to stereo vision. Technical Report (1981)
76. Luo, W., Schwing, A.G., Urtasun, R.: Efficient deep learning for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5695–5703. IEEE, New York (2016)
77. McCormac, J., Handa, A., Davison, A., Leutenegger, S.: SemanticFusion: Dense 3D semantic mapping with convolutional neural networks. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 4628–4635. IEEE, New York (2017)
78. McCormac, J., Clark, R., Bloesch, M., Davison, A., Leutenegger, S.: Fusion++: Volumetric object-level slam. In: Proceedings of the International Conference on 3D Vision, pp. 32–41. IEEE, New York (2018)
79. Mukherjee, D., Wu, Q.M.J., Wang, G.: A comparative experimental study of image feature detectors and descriptors. Mach. Vision Appl. **26**(4), 443–466 (2015)
80. Mur-Artal, R., Montiel, J.M.M., Tardos, J.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Rob. **31**(5), 1147–1163 (2015)
81. Naimark, L., Foxlin, E.: Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, p. 27. IEEE, New York (2002)
82. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the International Conference on Machine Learning, pp. 807–814 (2010)
83. Newcombe, R., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136. IEEE, New York (2011a)
84. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: Dense tracking and mapping in real-time. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2320–2327. IEEE, New York (2011b)
85. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 0756–777 (2004)
86. Nistér, D.: Preemptive RANSAC for live structure and motion estimation. Mach. Vision Appl. **16**(5), 321–329 (2005)
87. Oth, L., Furgale, P., Kneip, L., Siegwart, R.: Rolling shutter camera calibration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1360–1367. IEEE, New York (2013)
88. Pagani, A.: Reality Models for efficient registration in Augmented Reality. Verlag Dr. Hut, Germany (2014)
89. Park, Y., Lepetit, V., Woo, W.: Multiple 3D object tracking for augmented reality. In: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 117–120. IEEE, New York (2008)
90. Paulus, C.J., Haouchine, N., Cazier, D., Cotin, S.: Augmented reality during cutting and tearing of deformable objects. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 54–59. IEEE, New York (2015)
91. Poultney, C., Chopra, S., Cun, Y.L., et al.: Efficient learning of sparse representations with an energy-based model. In: Proceedings of the Conference on Neural Information Processing Systems, pp. 1137–1144 (2007)
92. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes 3rd edition: The art of scientific computing. Cambridge University, Cambridge (2007)
93. Prisacariu, V.A., Reid, I.D.: PWP3D: Real-time segmentation and tracking of 3D objects. Int. J. Comput. Vis. **98**(3), 335–354 (2012)
94. Puerto-Souza, G.A., Mariottini, G.L.: A fast and accurate feature-matching algorithm for minimally-invasive endoscopic images. IEEE Trans. Med. Imaging **32**(7), 1201–1214 (2013)
95. Qin, T., Li, P., Shen, S.: Vins-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Trans. Rob. **34**(4), 1004–1020 (2018)
96. Rad, M., Lepetit, V.: BB8: A scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3848–3856. IEEE, New York (2017)

97. Radkowski, R., Herrema, J., Oliver, J.: Augmented reality-based manual assembly support with visual features for different degrees of difficulty. Int. J. Hum.-Comput. Interact. **31**(5), 337–349 (2015)
98. Rambach, J., Tewari, A., Pagani, A., Stricker, D.: Learning to fuse: A deep learning approach to visual-inertial camera pose estimation. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 71–76. IEEE, New York (2016)
99. Rambach, J., Pagani, A., Stricker, D.: Augmented things: enhancing AR applications leveraging the internet of things and universal 3D object tracking. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 103–108. IEEE, New York (2017)
100. Rambach, J., Deng, C., Pagani, A., Stricker, D.: Learning 6DoF object poses from synthetic single channel images. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. IEEE, New York (2018)
101. Rambach, J., Lesur, P., Pagani, A., Stricker, D.: SlamCraft: Dense planar RGB monocular SLAM. In: Proceedings of the International Conference on Machine Vision Applications. Springer, Berlin (2019)
102. Reina, S.C., Solin, A., Kannala, J.: Robust gyroscope-aided camera self-calibration. In: Proceedings of the International Conference on Information Fusion, pp. 772–779. IEEE, New York (2018)
103. Renaudin, V., Afzal, M.H., Lachapelle, G.: Complete triaxis magnetometer calibration in the magnetic domain. J. Sens. (2010)
104. Ricolfe-Viala, C., Sanchez-Salmeron, A.J.: Lens distortion models evaluation. Appl. Opt. **49**(30), 5914–5928 (2010)
105. Rosten, E., Drummond, T.: Fusing points and lines for high performance tracking. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 2, pp. 1508–1515. IEEE, New York (2005)
106. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2564–2571. IEEE, New York (2011)
107. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533 (1986)
108. Runz, M., Buffier, M., Agapito, L.: Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 10–20. IEEE, New York (2018)
109. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Proceedings of the International Conference on 3-D Digital Imaging and Modeling, vol. 1, pp. 145–152 (2001)
110. Salas-Moreno, R., Newcombe, R., Strasdat, H., Kelly, P., Davison, A.: Slam++: Simultaneous localisation and mapping at the level of objects. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1352–1359. IEEE, New York (2013)
111. Salas-Moreno, R., Glocken, B., Kelly, P., Davison, A.: Dense planar SLAM. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 157–164. IEEE, New York (2014)
112. Sansoni, G., Trebeschi, M., Docchio, F.: State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. Sensors **9**(1), 568–601 (2009)
113. Seo, B.K., Wuest, H.: A direct method for robust model-based 3D object tracking from a monocular RGB image. In: Proceedings of the European Conference on Computer Vision, pp. 551–562. Springer, Berlin (2016)
114. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556 (2014)
115. Su, Y., Rambach, J., Minaskan, N., Lesur, P., Pagani, A., Stricker, D.: Deep multi-state object pose estimation for augmented reality assembly. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. IEEE, New York (2019)
116. Subbarao, R., Meer, P.: Beyond RANSAC: user independent robust regression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 101–101. IEEE, New York (2006)
117. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: Proceedings of the European Conference on Computer Vision, pp. 443–450. Springer, Berlin (2016)
118. Sundermeyer, M., Marton, Z.C., Durner, M., Brucker, M., Triebel, R.: Implicit 3D orientation learning for 6D object detection from RGB images. In: Proceedings of the European Conference on Computer Vision, pp. 699–715. Springer, Berlin (2018)
119. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9. IEEE, New York (2015)
120. Tan, D.J., Navab, N., Tombari, F.: Looking beyond the simple scenarios: Combining learners and optimizers in 3D temporal tracking. IEEE Trans. Visual Comput. Graphics **23**(11), 2399–2409 (2017)
121. Tateno, K., Tombari, F., Laina, I., Navab, N.: CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2. IEEE, New York (2017)
122. Tekin, B., Sinha, S.N., Fua, P.: Real-time seamless single shot 6D object pose prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 292–301. IEEE, New York (2018)
123. Titterton, D., Weston, J.L., Weston, J.: Strapdown inertial navigation technology, vol. 17. IET, United Kingdom (2004)
124. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: Proceedings of the International Workshop on Vision Algorithms, pp. 298–372. Springer, Berlin (1999)
125. Vacchetti, L., Lepetit, V., Fua, P.: Combining edge and texture information for real-time accurate 3D camera tracking. In: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 48–57. IEEE, New York (2004a)
126. Vacchetti, L., Lepetit, V., Fua, P.: Stable real-time 3D tracking using online and offline information. IEEE Trans. Pattern Anal. Mach. Intell. **26**(10), 1385–1391 (2004b)
127. Wagner, D., Schmalstieg, D.: Artoolkitplus for pose tracking on mobile devices. In: Proceedings of 12th Computer Vision Winter Workshop, 139–146 (2007)
128. Wang, S., Clark, R., Wen, H., Trigoni, N.: Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2043–2050. IEEE, New York (2017)
129. Wasenmüller, O., Stricker, D.: Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. In: Proceedings of the Asian Conference on Computer Vision Workshops, pp. 34–45. Springer, Berlin (2016)
130. Whelan, T., Kaess, M., Fallon, M.F.: Kintinuous: Spatially extended {K}inect{F}usion. In: Proceedings of the Workshop on RGB-D: Advanced Reasoning with Depth Cameras (2012)
131. Whelan, T., Salas-Moreno, R., Glocker, B., Davison, A., Leutenegger, S.: ElasticFusion: Real-time dense SLAM and light source estimation. Int. J. Rob. Res. **35**(14), 1697–1716 (2016)
132. Wuest, H., Vial, F., Stricker, D.: Adaptive line tracking with multiple hypotheses for augmented reality. In: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 62–69. IEEE, New York (2005)

133. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Proceedings of the European Conference on Computer Vision, pp. 818–833. Springer, Berlin (2014)
134. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vis. **13**(2), 119–152 (1994)
135. Zhang, Z., et al. Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 99, pp. 666–673. IEEE, New York (1999)
136. Zhang, Z., Li, M., Huang, K., Tan, T.: Practical camera autocalibration based on object appearance and motion for traffic scene visual surveillance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE, New York (2008)
137. Zhi, S., Bloesch, M., Leutenegger, S., Davison, A.J.: SceneCode: Monocular dense semantic reconstruction using learned encoded scene representations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11776–11785. IEEE, New York (2019)
138. Zhou, H., Ummenhofer, B., Brox, T.: Deeptam: Deep tracking and mapping. In: Proceedings of the European Conference on Computer Vision, pp. 822–838. Springer, Berlin (2018)

Jason Rambach received his PhD from the Computer Science department of the University of Kaiserslautern in 2020. He leads the research team Spatial Sensing and Machine Perception at DFKI Augmented Vision. He has several publications in computer vision and augmented reality conferences and a best paper award from ISMAR 2017. He acts as a reviewer for ISMAR, CVPR, and ECCV.

Alain Pagani is deputy director of the Augmented Vision department at the German Research Center for Artificial Intelligence (DFKI). His research interests include computer vision, artificial intelligence, image understanding, augmented reality, and machine learning, and he published over 50 articles in conferences and journals. Parallel to his work at DFKI, he gives lectures at the Technical University of Kaiserslautern.

Didier Stricker is professor in computer science at the University of Kaiserslautern and scientific director at the German Research Center for Artificial Intelligence (DFKI GmbH) in Kaiserslautern, where he leads the research department "Augmented Vision" focusing on computer vision, on-body sensor networks, and augmented reality. He was general chair of the first ISMAR conference and has over 300 publications.

3D Object and Hand Pose Estimation

4

Vincent Lepetit

Contents

4.1 3D Object and Hand Pose Estimation for Augmented Reality . . . 85
4.2 Formalization . . . 86
4.3 Challenges of 3D Pose Estimation Using Computer Vision . . . 87
4.4 Early Approaches to 3D Pose Estimation and Their Limits . . . 89
4.5 Machine Learning and Deep Learning . . . 90
4.6 Datasets . . . 91
4.6.1 Datasets for Object Pose Estimation . . . 91
4.6.2 Datasets for Hand Pose Estimation . . . 92
4.6.3 Datasets for Object and Hand Pose Estimation . . . 93
4.6.4 Metrics . . . 93
4.7 Modern Approaches to 3D Object Pose Estimation . . . 93
4.7.1 BB8 . . . 93
4.7.2 SSD-6D . . . 95
4.7.3 YOLO-6D . . . 95
4.7.4 PoseCNN . . . 95
4.7.5 DeepIM . . . 96
4.7.6 Augmented Autoencoders . . . 96
4.7.7 Robustness to Partial Occlusions: Oberweger's Method, Segmentation-Driven MeThod, PVNet . . . 96
4.7.8 DPOD and Pix2Pose . . . 97
4.7.9 Discussion . . . 97
4.8 3D Pose Estimation for Object Categories . . . 97
4.9 3D Hand Pose Estimation from Depth Maps . . . 98
4.9.1 DeepPrior++ . . . 99
4.9.2 V2V-PoseNet . . . 99
4.9.3 A2J . . . 99
4.9.4 Discussion . . . 100
4.10 3D Hand Pose Estimation from an RGB Image . . . 100
4.10.1 Zimmerman's Method . . . 100
4.10.2 Iqbal's Method . . . 100
4.10.3 GANerated Hands . . . 101
4.10.4 3D Hand Shape and Pose Estimation: Ge's and Boukhayma's Methods . . . 101
4.10.5 Implementation in MediaPipe . . . 101
4.10.6 Manipulating Virtual Objects . . . 102
4.11 3D Object+Hand Pose Estimation . . . 102
4.11.1 ObMan and HOPS-Net . . . 102
4.11.2 H+O . . . 103
4.11.3 HOnnotate . . . 103
4.12 The Future of 3D Object and Hand Pose Estimation . . . 103
References . . . 104

Abstract

3D object and hand pose estimation have huge potentials for Augmented Reality, to enable tangible interfaces, natural interfaces, and blurring the boundaries between the real and virtual worlds. In this chapter, we first motivate the topic and explain the challenges. After a brief review of early work and Deep Learning techniques on which recent works are based, we present the recent developments for 3D object and hand pose estimation using cameras, when considered separately and together. We examine the abilities and limitations of each technique. We conclude by discussing the possible future developments of the field.

Keywords

Hand-object interaction · 3D object pose estimation · 3D hand pose estimation · Tangible interfaces · Deep Learning

4.1 3D Object and Hand Pose Estimation for Augmented Reality

An Augmented Reality system should not be limited to the visualization of virtual elements integrated to the real world,

V. Lepetit (✉)
LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France
e-mail: vincent.lepetit@enpc.fr

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_4

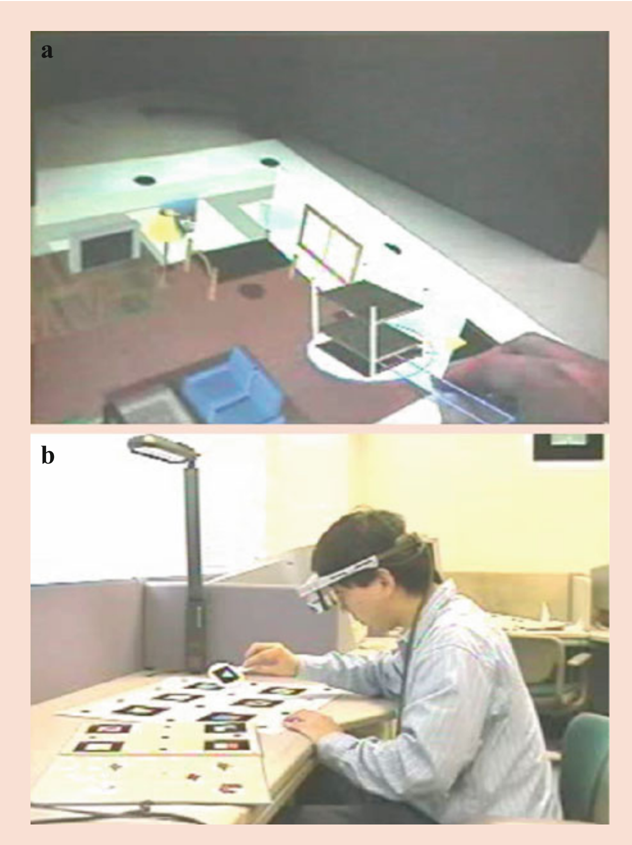


Fig. 4.1 (**a**) Augmented Reality and (**b**) external views of the early magic book developed by Kato and Billinghurst. This visionary system shows the importance of both virtual and real object manipulation in AR applications. (Images courtesy of the authors)

it should also *perceive* the user and the real world surrounding the user. As even early Augmented Reality applications demonstrated, this is required to provide rich user experience and unlock new possibilities. For example, the magic book application developed by Billinghurst and Kato [8] and illustrated by Fig. 4.1 shows the importance of being able to manipulate real objects in Augmented Reality. It featured a real book, from which virtual objects would pop up, and these virtual objects could be manipulated by the user using a small real "shovel."

This early system relied on visual markers on the real book and the shovel. These markers would become the core of the popular software framework ARToolkit [3], and were essential to robustly estimate the locations and orientations of objects in the 3D space, a geometric information required for the proper integration of the virtual objects with the real book, and their manipulation by the shovel.

Visual markers, however, require a modification of the real world, which is not always possible nor desirable in practice. Similarly, magnetic sensors have also been used to perceive the spatial positions of real objects, in order to integrate them to the augmented world. While they can be made invisible by contrast with visual markers, they also require a preparation of the scene, and metallic objects can affect the magnetic field.

Being able to perceive the user's hands is also very important, as the hands can act as an interface between the user's intention and the virtual world. Haptic gloves or various types of joysticks have been used to capture the hands' locations and motions. In particular, input pads are still popular with current augmented and virtual reality platforms, as they ensure robustness and accuracy for the perception of the user's actions. Getting rid of such hardware is extremely desirable, as it makes Augmented Reality user interfaces intuitive and natural.

Researchers have thus been developing computer vision approaches to perceiving the real world using simple cameras and without having to engineer the real objects with markers or sensors or the user's hands with gloves or handles. Such perception problems also appear in fields such as robotics, and the scientific literature on this topic is extremely rich. The goal of this chapter is to introduce the reader to 3D object and hand perception based on cameras for Augmented Reality applications, and to the scientific literature on the topic, with a focus on the most recent techniques.

In the following, we will first detail the problem of 3D pose estimation, explain why it is difficult, and review early approaches to motivate the use of Machine Learning techniques. After a brief introduction to Deep Learning, we will discuss the literature on 3D pose estimation for objects, hands, and hands manipulating objects through representative works.

4.2 Formalization

In the rest of this chapter, we will define the 3D pose of a rigid object as its 3D location and orientation in some coordinate system, as shown in Fig. 4.2. In the case of Augmented Reality, this coordinate system should be directly related to the headset, or to the tablet or smartphone, depending on the visualization system, *and* to the object. This is thus different from simultaneous localization and mapping (SLAM), where the 3D pose of the camera can be estimated in an arbitrary coordinate system as long as it is consistent over time.

In general, this 3D pose has thus 6 degrees of freedom: 3 for the 3D translation and 3 for the 3D orientation, which can be represented, for example, with a 3D rotation matrix or a quaternion. Because of these 6 degrees of freedom, the 3D pose is also called "6D pose" in the literature, the terms "3D pose" and "6D pose" thus refer to the same notion.

Articulated objects, such as scissors, or even more deformable objects such as clothes or sheets of papers have also been considered in the literature [75,84]. Their positions

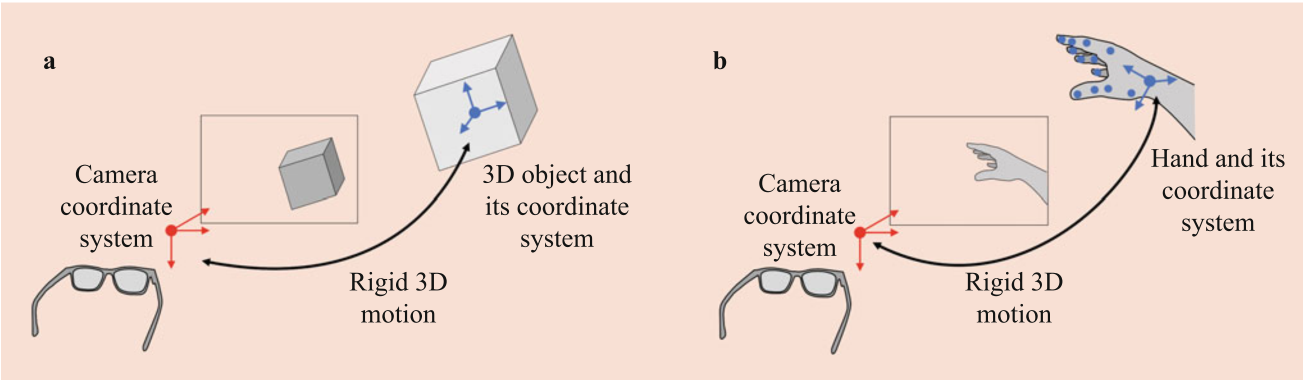


Fig. 4.2 (**a**) The 3D pose of a rigid object can be defined as a 3D rigid motion between a coordinate system attached to the object and another coordinate system, for example, one attached to an Augmented Reality headset. (**b**) One way to define the 3D pose of a hand can be defined as a rigid motion between a coordinate system attached to one of the joints (e.g., the wrist) and another coordinate system, plus the 3D locations of the joints in the first coordinate system. 3D rigid motions can be defined as the composition of a 3D rotation and a 3D translation

and shapes in space have many more degrees of freedom, and specific representations of these poses have been developed. Estimating these values from images robustly remains very challenging.

The 3D pose of the hand is also very complex, since we would like to also consider the positions of the individual fingers. This is by contrast with gesture recognition, for example, which aims at assigning a distinctive label such as "open" or "close" to a certain hand pose. Instead, we would like to estimate continuous values in the 3D space. One option, among others, to represent the 3D pose of a hand is to consider the 3D locations of each joint in some coordinate system. Given a hand model with bone length and rotation angles for all degrees of freedom (DoF) of the joints, forward kinematics [28] can be used to calculate the 3D joint locations. Reciprocally, inverse kinematics [101] can be applied to obtain joint angles and bone length from the 3D joint locations. In addition, one may also want to estimate the shapes of the user's hands, such as their sizes or the thickness of the fingers.

To be useful for Augmented Reality, 3D pose estimation has to run continuously, in real time. When using computer vision, it means that it should be done from a single image, or stereo images captured at the same time, or RGB-D images that provide both color and depth data, maybe also relying on the previous images to guide or stabilize the 3D pose estimations. Cameras with large fields of view help, as they limit the risk of the target objects to leave the field of view compared to more standard cameras.

Stereo camera rigs, made of two or more cameras, provide additional spatial information that can be exploited for estimating the 3D pose but make the device more complex and more expensive. RGB-D cameras are currently a good trade-off, as they also provide, in addition to a color image, 3D information in the form of a depth map, i.e., a depth value for each pixel, or at least most of the pixels of the color image. "Structured light" RGB-D cameras measure depth by projecting a known pattern in infrared light and capturing the projection with an infrared camera. Depth can then be estimated from the deformation of the pattern. "Time-of-flight" cameras are based on pulsed light sources and measure the time a light pulse takes to travel from the emitter to the scene and come back after reflection.

With structured-light technology, depth data can be missing at some image locations, especially along the silhouettes of objects or on dark or specular regions. This technology also struggles to work outdoor, because of the ambient infrared sunlight. Time-of-flight cameras are also disturbed outdoor, as the high intensity of sunlight causes a quick saturation of the sensor pixels. Multiple reflections produced by concave shapes can also affect the time of flight.

But despite these drawbacks, RGB-D cameras remain however very useful in practice when they can be used, and algorithms using depth maps perform much better than algorithms relying only on color information – even though the gap is decreasing in recent research.

4.3 Challenges of 3D Pose Estimation Using Computer Vision

There are many challenges that need to be tackled in practice when estimating the 3D poses of objects and hands from images. We list below some of these challenges, illustrated in Fig. 4.3.

High Degrees of Freedom As mentioned above, the pose of a rigid object can be represented with six scalar values, which is already a large number of degrees of freedom to estimate. The 3D pose of a human hand lives in an even much higher dimensional space, as it is often represented with about 32 degrees of freedom. The risk of making an

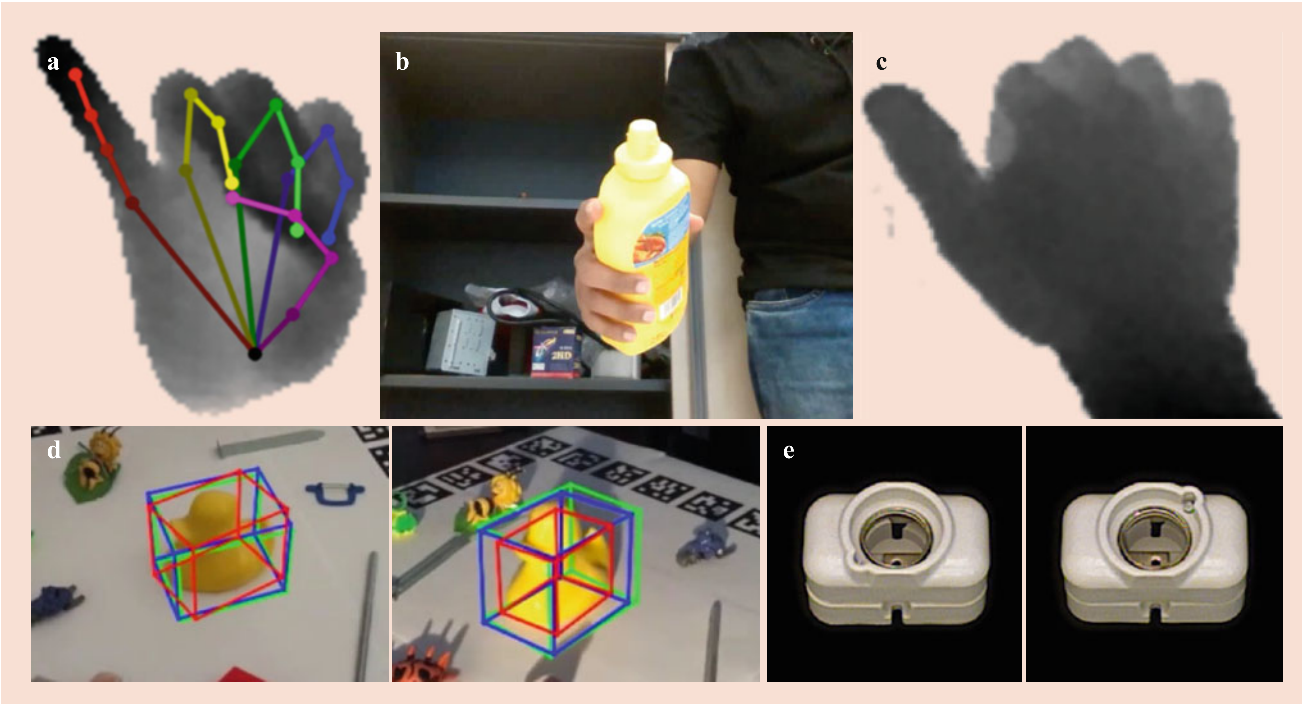


Fig. 4.3 Some challenges when estimating the 3D poses of objects and hands. (**a**) Degrees of freedom. The human hand is highly articulated, and many parameters have to be estimated to correctly represent its pose. (**b**) Occlusions. In this case, the hand partially occludes the object, and the object partially occludes the hand. (**c**) Self-occlusions can also occur in the case of hands, especially in egocentric views. (**d**) Illuminations. Like the rubber duck in these examples, the appearance of objects can vary dramatically with illumination. (**e**) Ambiguities. Many manufactured objects have symmetric or almost symmetric shapes, which may make pose estimation ambiguous. (Images from the T-LESS dataset [38], courtesy of the authors)

error when estimating the pose increases with this number of degrees of freedom.

Occlusions Occlusions of the target objects or hands, even partial, often disturb pose estimation algorithms. It can be difficult to identify which parts are visible and which parts are occluded, and the presence of occlusions may make pose estimation completely fail or be very inaccurate. Occlusions often happen in practice. For example, when a hand manipulates an object, both the hand and the object are usually partially occluded.

Cluttered Background Objects in the background can act as distractors, especially if they look like target objects or have similar parts.

Changing Illumination Conditions In practice, illumination cannot be controlled and will change the appearance of the objects, not only because the lighting is different but also because shadows can be cast on them. This requires robust algorithms. Moreover, cameras may struggle to keep a good balance between bright glares and dark areas, and using high dynamic range cameras becomes appealing under such conditions. Moreover, sunlight may make depth cameras fail as discussed above.

Material and Textures Early approaches to pose estimation relied on the presence of texture or pattern on the objects' surfaces, because stable features can be detected and matched relatively easily and efficiently on patterns. However, in practice, many objects lack textures, making such approach fail. Non-Lambertian surfaces such as metal and glass make the appearance of objects change with the objects' poses because of specularities and reflections appearing on the objects' surfaces. Transparent objects are also of course problematic, as they do not appear clearly in images.

Ambiguities Many manufactured objects are symmetrical or almost symmetrical or have repetitive patterns. This generates possible ambiguities when estimating the poses of these objects that need to be handled explicitly.

3D Model Requirement Most of the existing algorithms for pose estimation assume the knowledge of some 3D models of the target objects. Such 3D models provide very useful geometric constraints, which can be exploited to estimate the objects' 3D poses. However, building such 3D models takes time and expertise, and in practice, a 3D model is not readily available. Hand pose estimation suffers much less from this problem, since the 3D geometry of hands remains very similar from one person to another. Some recent works have also considered the estimation of the objects' geometry together with their poses [25]. These works are still limited to a few object categories, such as chairs or cars, but are very interesting for future developments.

4.4 Early Approaches to 3D Pose Estimation and Their Limits

3D pose estimation from images has a long history in computer vision. Early methods were based on simple image features, such as edges or corners. Most importantly, they strongly relied on some prior knowledge about the 3D pose, to guide the pose estimation in the high-dimensional space of possible poses. This prior may come from the poses estimated for the previous images or could be provided by a human operator.

3D Tracking Methods For example, pioneer works such as [30] and [55] describe the object of interest as a set of 3D geometric primitives such as lines and conics. These features were matched with contours in the image to find a 3D pose estimate by solving an optimization problem to find the 3D pose that reprojects the 3D geometric primitives to the matched image contours. The entire process is computationally light, and careful implementations were able to achieve high frame rates with computers that would appear primitive to us.

Unfortunately, edge-based approaches are quite unreliable in practice. Matching the reprojection of the 3D primitives with image contours is difficult to achieve: 3D primitives do not necessarily appear as strong image contours, except for carefully chosen objects. As a result, these primitives are likely to be matched with the wrong image contours, especially in case of cluttered background. Incorrect matches will also occur in case of partial occlusions. Introducing robust estimators into the optimization problem [17] helps, but it appears that it is impossible in general to be robust to such mismatches in edge-based pose estimation: Most of the time, a match between two contours provides only limited constraints on the pose parameters, as the two contours can "slide" along each other and still be a good match. Contours thus do not provide reliable constraints for pose estimation.

Relying on feature points [31] instead of contours provides more constraints as point-to-point matching does not have the "sliding ambiguity" of contour matching. Feature point matching has been used for 3D object tracking with some success, for example, in [107]. However, such approach assumes the presence of feature points that can be detected on the object's surface, which is not true for all the objects.

The Importance of Detection Methods The two approaches described above assume that prior knowledge on the object pose is available to guide the matching process between contours or points. Typically, the object pose estimated at time t is exploited to estimate the pose at time $t+1$. In practice, this makes such approaches fragile, because the poses at time t and $t + 1$ can be very different if the object moves fast or because the pose estimated at time t can be wrong.

Being able to estimate the 3D pose of an object without relying too much on prior knowledge is therefore very important in practice. As we will see, this does not mean that methods based on strong prior knowledge on the pose are not useful. In fact, they tend to be much faster and/or more accurate than "detection methods," which are more robust. A natural solution is thus to combine both, and this is still true in the modern era of object and pose estimation.

One early popular method for object pose estimation without pose prior was also based on feature points, often referred to as keypoints in this context. This requires the ability to match keypoints between an input image and a reference image of the target object, which is captured offline and in which the 3D pose and the 3D shape of the object are known. By using geometry constraints, it is then possible to estimate the object's 3D pose. However, this so-called wide baseline point matching is much more difficult than short baseline matching used by tracking methods. SIFT keypoints and descriptors [56] were a breakthrough that made many computer vision applications possible, including 3D pose estimation. They were followed by faster methods, including SURF [6] and ORB [82].

As for tracking methods based on feature points, the limitation of keypoint-based detection and pose estimation is that it is limited to objects exhibiting enough keypoints, which is not the case in general. This approach was still very successful for Augmented Reality in magazines, for example, where the "object" is an image printed on paper – so it has a simple planar Geometry – and selected to guarantee that the approach will work well by exhibiting enough keypoints [45].

To be able to detect and estimate the 3D pose of objects with almost no keypoints, sometimes referred to as "textureless" objects, some works attempted to use "templates." The templates of [36] aim at capturing the possible appearances of the object by discretizing the 3D pose space and representing the object's appearance for each discretized pose by a template covering the full object, in a way that is robust to lighting variations. Then, by scanning the input images looking for templates, the target object can be detected in 2D and its pose estimated based on the template that matches best the object appearance. However, such approach requires the creation of many templates and is poorly robust to occlusions.

Conclusion We focused in this section on object pose estimation rather than hand pose Estimation; however, the conclusion would be the same. Early approaches were based on handcrafted methods to extract features from images, with the goal of estimating the 3D pose from these features. This is however very challenging to do and almost doomed to fail in the general case. Since then, Machine Learning-based methods have been shown to be more adapted, even if they also come with their drawbacks, and will be discussed in the rest of this chapter.

4.5 Machine Learning and Deep Learning

Fundamentally, 3D pose estimation of objects or hands can be seen as a mapping from an image to a representation of the pose. The input space of this mapping is thus the space of possible images, which is an incredibly large space: For example, a RGB VGA image is made of almost 1 million pixel values. Not many fields deal with 1 million dimension data! The output space is much smaller, since it is made of six values for the pose of a rigid object or a few tens for the pose of a hand. The natures of the input and output spaces of the mapping sought in pose estimation are therefore very different, which makes this mapping very complex. From this point of view, we can understand that it is very difficult to hope for a pure "algorithmic" approach to encode this mapping.

This is why Machine Learning techniques, which use data to improve algorithms, have become successful for pose estimation problems and computer vision problems in general. Because they are data driven, they can find automatically an appropriate mapping, by contrast with previous approaches that required hardcoding mostly based on intuition, which can be correct or wrong.

Many Machine Learning methods exist, and random forests, also called random decision forests or randomized trees, were an early popular method in the context of 3D pose estimation [10, 50, 87]. Random forests can be efficiently applied to image patches and discrete (multi-class) and continuous (regression) problems, which makes them flexible and suitable to fast, possibly real-time, applications.

Deep Learning For multiple reasons, Deep Learning, a Machine Learning technique, took over the other methods almost entirely during the last years in many scientific fields, including 3D pose estimation. It is very flexible, and, in fact, it has been known for a long time that any continuous mapping can be approximated by two-layer networks, as finely as wanted [40, 72]. In practice, networks with more than two layers tend to generalize better, and to need dramatically less parameters than two-layer networks [18], which makes them a tool of choice for computer vision problems.

Many resources can now be easily found to learn the fundamentals of Deep Learning, for example, [23]. To stay brief, we can say here that a deep network can be defined as a composition of functions ("layers"). Figure 4.4 shows a basic deep network architecture. These functions may depend on parameters, which need to be estimated for the network to perform well. Almost any function can be used as layer, as long as it is useful to solve the problem at hand and if it is *differentiable*. This differentiable property is indeed required to find good values for the parameters by solving an optimization problem.

Deep Network Training For example, if we want to make a network F predict the pose of a hand visible in an image, one way to find good values $\hat{\Theta}$ for the network parameters is to solve the following optimization problem:

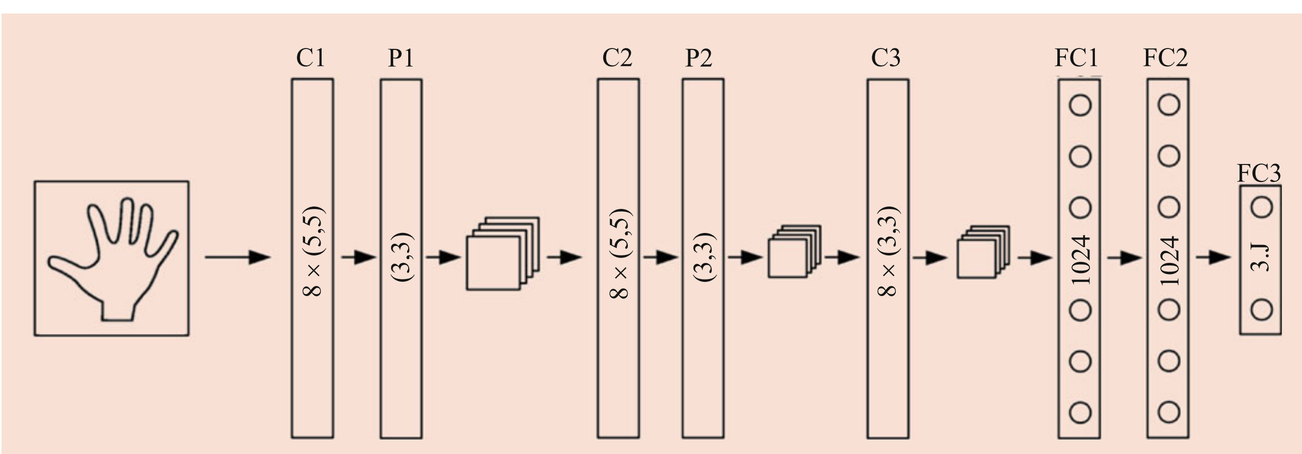


Fig. 4.4 A basic deep network architecture applied to hand pose prediction. C1, C2, and C3 are convolutional layers, P1 and P2 are pooling layers, and FC1, FC2, and FC3 are fully connected layers. The numbers in each bar indicates either. The number and size of the linear filters in case of the convolutional layers, the size of the pooling region for the pooling layers, and the size of the output vector in case of the fully connected layers

$$\hat{\Theta} = \arg\min_{\Theta} \mathcal{L}(\Theta) \quad \text{with}$$
$$\mathcal{L}(\Theta) = \frac{1}{N}\sum_{i=1}^{N} \|F(I_i;\Theta) - \mathbf{e}_i\|^2 \,, \tag{4.1}$$

where $\{(I_1, \mathbf{e}_1), .., (I_N, \mathbf{e}_N)\}$ is a *training set* containing pairs made of images I_i and the corresponding poses $\mathbf{e}_i$, where the $\mathbf{e}_i$ are vectors that contain, for example, the 3D locations of the joints of the hand as was done in [63]. Function $\mathcal{L}$ is called a loss function. Optimizing the network parameters Θ is called training. Many optimization algorithms and tricks have been proposed to solve problems like Eq. (4.1), and many software libraries exist to make the implementation of the network creation and its training an easy task.

Because optimization algorithms for network training are based on gradient descent, any differentiable loss function can be used in principle, which makes Deep Learning a very flexible approach, as it can be adapted easily to the problem at hand.

Supervised Training and the Requirement for Annotated Training Sets Training a network by using a loss function like the one in Eq. (4.1) is called supervised training, because we assume the availability of an annotated dataset of images. Supervised training tends to perform very well, and it is used very often in practical applications.

This is however probably the main drawback of Deep Learning-based 3D pose estimation. While early methods relied only on a 3D model, and for some of them only on a small number of images of the object, modern approaches based on Deep Learning require a large dataset of images of the target objects, annotated with the ground truth 3D poses.

Such datasets are already available (see Sect. 4.6), but they are useful mostly for evaluation and comparison purposes. Applying current methods to new objects requires creating a dataset for these new objects, and this is a cumbersome task.

It is also possible to use synthetic data for training, by generating images using computer graphics techniques. This is used in many works, often with special care to take into account the differences between real and synthetic images, as we will see below.

4.6 Datasets

Datasets have become an important aspect of 3D pose estimation, for training, evaluating, and comparing methods. We describe some for object, hand, and hand+object pose estimation below.

4.6.1 Datasets for Object Pose Estimation

Figure 4.5 shows images from the most popular datasets for object pose estimation.

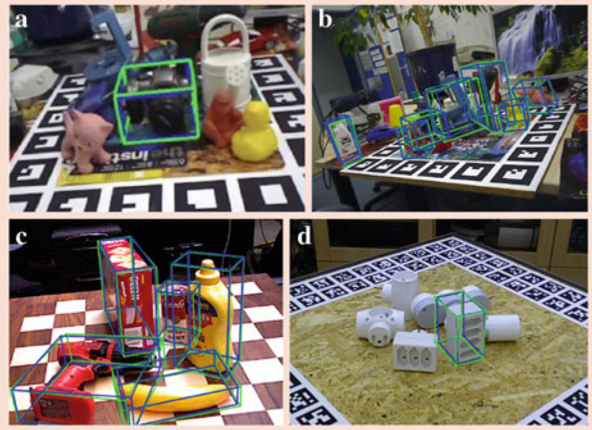


Fig. 4.5 Images from popular datasets for 3D object pose estimation. (**a**) LineMOD; (**b**) Occluded LineMOD; (**c**) YCB-Video; (**d**) T-LESS. (Images courtesy of the datasets' authors)

LineMOD Dataset The LineMOD dataset [36] predates most Machine Learning approaches, and as such, it is not divided into a training and a test set. It is made of 15 small objects, such as a camera, a lamp, and a cup. For each object, it offers a set of 1200 RGB-D images of the object surrounded by clutter. The other objects are often visible in the clutter, but only the 3D pose of the target object is provided for each set. The 3D models of the objects are also provided.

Occluded LineMOD Dataset The Occluded LineMOD dataset was created by the authors of [10] from LineMOD by annotating the 3D poses of the objects belonging to the dataset but originally not annotated because they were considered as part of the clutter. This results into a sequence of 1215 frames, each frame labeled with the 3D poses of eight objects in total, as well as the objects' masks. The objects show severe occlusions, which makes pose estimation challenging.

YCB-Video Dataset This dataset [112] consists of 92 video sequences, where 12 sequences are used for testing and the remaining 80 sequences for training. In addition, the dataset contains 80k synthetically rendered images, which can be used for training as well. There are 21 "daily life" objects in the dataset, from cereal boxes to scissors or plates. These objects were selected from the YCB dataset [14] and are available for purchase. The dataset is captured with two different RGB-D sensors. The test images are challenging due to the presence of significant image noise, different illumination levels, and large occlusions. Each image is annotated with the 3D object poses, as well as the objects' masks.

T-LESS Dataset The T-LESS dataset [38] is made from 30 "industry-relevant" objects. These objects have no discriminative color nor texture. They present different types of symmetries and similarities between them, making pose estimation often almost ambiguous. The images were captured using three synchronized sensors: two RGB-D cameras, one structured-light based and one time-of-flight based, and one high-resolution RGB camera. The test images (10K from each sensor) are from 20 scenes with increasing difficulties, with partial occlusions and contacts between objects. This dataset remains extremely challenging.

4.6.2 Datasets for Hand Pose Estimation

Early Datasets The NYU dataset [102] contains over 72k training and 8k test RGB-D images data, captured from three different viewpoints using a structured-light camera. The images were annotated with 3D joint locations with a semiautomated approach, by using a standard 3D hand tracking algorithm reinitialized manually in case of failure. The ICVL dataset [96] contains over 180k training depth frames showing various hand poses, and two test sequences with each approximately 700 frames, all captured with a time-of-flight camera. The depth images have a high quality with hardly any missing depth values and sharp outlines with little noise. Unfortunately, the hand pose variability of this dataset is limited compared to other datasets, and annotations are rather inaccurate [94]. The MSRA dataset [90] contains about 76k depth frames, captured using a time-of-flight camera from nine different subjects.

BigHands Dataset The BigHands dataset [116] contains an impressive 2.2 million RGB-D images captured with a structured-light camera. The dataset was automatically annotated by using six 6D electromagnetic sensors and inverse kinematics to provide 21 3D joint locations per frame. A significant part is made of egocentric views. The depth images have a high quality with hardly any missing depth values and sharp outlines with little noise. The labels are sometimes inaccurate because of the annotation process, but the dataset has a large hand pose variability, from ten users (Fig. 4.6).

CMU Panoptic Hand Dataset The CMU Panoptic hand dataset [88] is made of 15k real and synthetic RGB images, from a third-person point of view. The real images were recorded in the CMU's Panoptic studio and annotated by the method proposed in the paper, based on multiple views. The annotations are only in 2D but can still be useful for multi-view pose estimation.

Fig. 4.6 Images from popular datasets for 3D hand (**a**) and hand+object (**b**)–(**f**) pose estimation. (**a**) BigHand; (**b**) GANerated hand dataset; (**c**) First-Person Hand Action dataset; (**d**) Obman dataset; (**e**) FreiHAND dataset; (**f**) HO-3D dataset. (Images courtesy of the datasets' authors)

4.6.3 Datasets for Object and Hand Pose Estimation

GANerated Hand Dataset The GANerated hand dataset [60] is a large dataset made of 330K synthetic images of hands, sometimes holding an object, in front of a random background. The images are annotated with the 3D poses of the hand. The images were made more realistic by extending CycleGAN [118].

First-Person Hand Action Dataset The First-Person Hand Action dataset [20] provides a dataset of hand and object interactions with 3D annotations for both hand joints and object pose. They used a motion capture system made of magnetic sensors attached to the user's hand and to the object in order to obtain hand 3D pose annotations in RGB-D video sequences. Unfortunately, this changes the appearance of the hand in the color images as the sensors and the tape attaching them are visible, but the dataset proposes a large number of frames under various conditions (more than 100K egocentric views of 6 subjects doing 45 different types of daily-life activities).

ObMan Dataset Very recently, [32] introduced ObMan, a large dataset of images of hands grasping objects. The images are synthetic, but the grasps are generated using an algorithm from robotics, and the grasps still look realistic. The dataset provides the 3D poses and shapes of the hand as well as the object shapes.

FreiHand Dataset Zimmermann et al. [120] proposed a multi-view RGB dataset, FreiHAND, which includes hand-object interactions and provides the 3D poses and shapes of the hand. It relies on a green screen background environment so that it is easy to change the background for training purposes.

HO-3D Dataset Hampali et al. [29] proposed a method to automatically annotate video sequences captured with one or more RGB-D cameras with the object and hand poses and shapes. This results in a dataset made of 75,000 real RGB-D images, from 10 different objects and 10 different users. The objects come from the YCB dataset (see Sect. 4.6.1). The backgrounds are complex, and the mutual occlusions are often large, which makes the pose estimation realistic but very challenging.

4.6.4 Metrics

Metrics are important to evaluate and compare methods. Many metrics exist, and we describe here only the main ones for object pose estimation. Discussions on metrics for 3D object pose estimation can be found in [38] and [12].

ADD, ADI, ADD-S, and the 6D Pose Metrics The ADD metric [35] calculates the average distance in 3D between the model points, after applying the ground truth pose and the predicted pose. This can be formalized as

$$\text{ADD} = \frac{1}{|\mathcal{V}|} \sum_{\mathbf{M} \in \mathcal{V}} \|\text{Tr}(\mathbf{M}; \hat{\mathbf{p}}) - \text{Tr}(\mathbf{M}; \bar{\mathbf{p}})\|_2 \,, \tag{4.2}$$

where $\mathcal{V}$ is the set of the object's vertices, $\hat{\mathbf{p}}$ is the estimated pose, $\bar{\mathbf{p}}$ is the ground truth pose, and $\text{Tr}(\mathbf{M}; \mathbf{p})$ is the rigid transformation in $\mathbf{p}$ applied to 3D point $\mathbf{M}$. In the 6D Pose metric, a pose is considered when the ADD metric is less than 10% of the object's diameter.

For the objects with ambiguous poses due to symmetries, [35] replaces the ADD metric by the ADI metric, also referred to as the ADD-S metric in [113], computed as follows:

$$\text{ADD-S} = \frac{1}{|\mathcal{V}|} \sum_{\mathbf{M}_1 \in \mathcal{V}} \min_{\mathbf{M}_2 \in \mathcal{V}} \|\text{Tr}(\mathbf{M}_1; \hat{\mathbf{p}}) - \text{Tr}(\mathbf{M}_2; \bar{\mathbf{p}})\|_2 \,, \tag{4.3}$$

which averages the distances from points after applying the predicted pose to the *closest* points under the ground truth pose. The advantage of this metric is that it is indeed equal to zero when the pose is retrieved up to a symmetry, even if it does not exploit the symmetries of the object.

4.7 Modern Approaches to 3D Object Pose Estimation

Over the past years, many authors realized that Deep Learning is a powerful tool for 3D object pose estimation from images. We discuss here the development of Deep Learning applied to 3D object pose estimation over time. This development was and is still extremely fast, with improving accuracy, robustness, and computation times. We present this development through several milestone methods, but much more methods could also be included here.

4.7.1 BB8

One of the first Deep Learning methods for 3D object pose estimation is probably BB8 [76]. As it is the first method we describe, we will present it in some details.

This method proceeds in three steps. It first detect the target objects in 2D using coarse object segmentation. It then applies a deep network on each image window centered

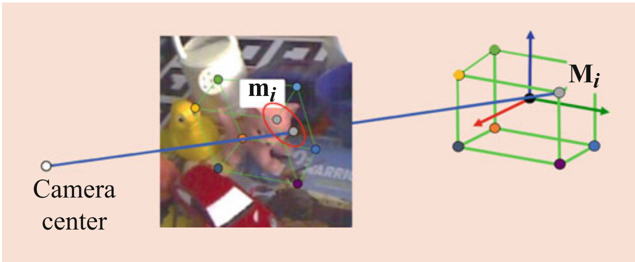


Fig. 4.7 Some 3D object pose estimation methods predict the pose by first predicting the 2D reprojections $\mathbf{m}_i$ of some 3D points $\mathbf{M}_i$ and then computing the 3D rotation and translation from the 2D-3D correspondences between the $\mathbf{m}_i$ and $\mathbf{M}_i$ points using a PnP algorithm

on detected objects. Instead of predicting the 3D pose of the detected objects in the form of a 3D translation and a 3D rotation, it predicts the 2D projections of the corners of the object's bounding box and computes the 3D pose from these 2D-3D correspondences with a PnP algorithm [19] – hence the name for the method, from the eight corners of the bounding box, as illustrating in Fig. 4.7. Compared to the direct prediction of the pose, this avoids the need for a meta-parameter to balance the translation and rotation terms. It also tends to make network optimization easier. This representation was used in some later works.

Since it is the first Deep Learning method for pose estimation we describe, we will detail the loss function (see Sect. 4.5) used to train the network. This loss function $\mathcal{L}$ is a least-squares error between the predicted 2D points and the expected ones and can be written as

$$\mathcal{L}(\Theta) = \frac{1}{8} \sum_{(W,\mathbf{p})\in\mathcal{T}} \sum_i \|\mathrm{Proj}_{\mathbf{p}}(\mathbf{M}_i) - F(W;\Theta)_i\|^2 , \quad (4.4)$$

where F denotes the trained network and Θ its parameters. $\mathcal{T}$ is a training set made of image windows W containing a target object under a pose $\mathbf{p}$. The $\mathbf{M}_i$ are the 3D coordinates of the corners of the bounding box of for this object, in the object coordinate system. $\mathrm{Proj}_{\mathbf{e},\mathbf{t}}(\mathbf{M})$ projects the 3D point $\mathbf{M}$ on the image from the pose defined by $\mathbf{e}$ and $\mathbf{t}$. $F(W;\Theta)_i$ returns the two components of the output of F corresponding to the predicted 2D coordinates of the i-th corner.

The Problem of Symmetrical and "Almost Symmetrical" Objects Predicting the 3D pose of objects as a standard least-squares problem, using a standard representation of the pose or point reprojections as in BB8, yields to large errors on symmetrical objects, such as many objects in the T-Less dataset (Fig. 4.5d). This dataset is made of manufactured objects that are not only similar to each other but also have one axis of rotational symmetry. Some objects are not perfectly symmetrical but only because of small details, like a screw.

The approach described above fails on these objects because it tries to learn a mapping from the image space to the pose space. Since two images of a symmetrical object under two different poses look identical, the image-pose correspondence is in fact a one-to-many relationship.

For objects that are perfectly symmetrical, [76] proposed a solution that we will not describe here, as simpler solutions have been proposed since. Objects that are "almost symmetrical" can also disturb pose prediction, as the mapping from the image to the 3D pose is difficult to learn even though this is a one-to-one mapping. Most recent methods ignore the problem and consider that these objects are actually perfectly symmetrical and consider a pose recovered up to the symmetries as correct. For such object, BB8 proposes to first consider these objects as symmetrical and then to train a classifier (also a deep network) to predict which pose is actually the correct one. For example, for an object with a rotational symmetry of 180°, there are two possible poses in general, and the classifier has to decide between two classes. This is a much simpler problem than predicting 6 degrees of freedom for the pose, and the classifier can focus on the small details that break the symmetry.

Refinement Step The method described above provides an estimate for the 3D pose of an object using only feedforward computation by a network from the input image to the 3D pose. It is relatively natural to aim at refining this estimate, which is a step also present in many other methods. In BB8, this is performed using a method similar to the one proposed in [62] for hand detection in depth images. A network is trained to improve the prediction of the 2D projections by comparing the input image and a rendering of the object for the initial pose estimate, as illustrated in Fig. 4.8. Such refinement can be iterated multiple times to improve the pose estimate.

One may wonder why the refinement step, in BB8 but also in more recent works such as DeepIM [52] or DPOD [117], can improve pose accuracy while it is (apparently) trained with the same data as the initial pose prediction. This can be understood by looking more closely at how the network predicting the update is trained. The input part of one training sample is made of a regular input image, plus a rendered image for a pose close to the pose for input image, and the output part is the difference between the two poses. In practice, the pose for the rendered image is taken as the ground truth pose for the real image plus some random noise. In other words, from one sample of the original dataset trained to predict the network providing the first estimate, it is possible to generate a virtually infinite number of samples for training the refinement network, by simply adding noise to the ground truth pose. The refinement network is thus trained with much more samples than the first network.

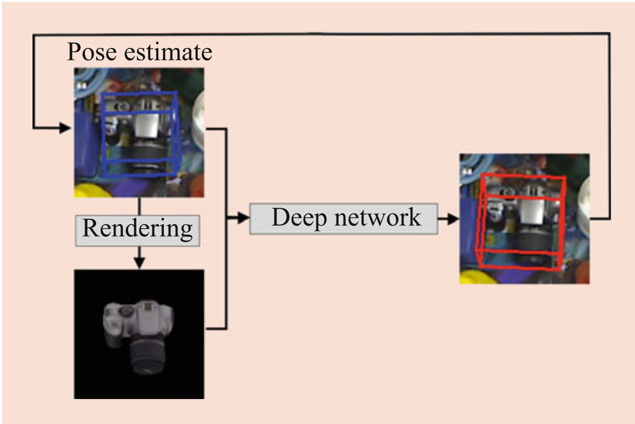


Fig. 4.8 Pose refinement in BB8. Given a first pose estimate, shown by the blue bounding box, BB8 generates a color rendering of the object. A network is trained to predict an update for the object pose given the input image and this rendering, to get a better estimate shown by the red bounding box. This process can be iterated

Data Augmentation Data augmentation, i.e., generating additional training images from the available ones, is often critical in Deep Learning. In the case of BB8, the objects' silhouettes are extracted from the original training images, which can be done using the ground truth poses and the objects' 3D models. The background is replaced by a patch extracted from a randomly picked image from the ImageNet dataset [83]. Note that this procedure removes context by randomly replacing the surroundings of the objects. Some information is thus lost, as context could be useful for pose estimation.

4.7.2 SSD-6D

BB8 relies on two separate networks to first detect the target objects in 2D and then predict their 3D poses, plus a third one if refinement is performed. Instead, SSD-6D [44] extends a deep architecture (the SSD architecture [54]) developed for 2D object detection to 3D pose estimation (referred in the paper as 6D estimation). SSD had already been extended to pose estimation in [74], but SS6-6D performs full 3D pose prediction.

This prediction is done by first discretizing the pose space. Each discretized pose is considered as a class, to turn the pose prediction into a classification problem rather than a regression one, as this was performing better in the authors' experience. This discretization was done on the 3D pose decomposition into direction of view over a half-sphere and in-plane rotation. The 3D translation can be computed from the 2D bounding box. To deal with symmetrical objects, views on the half-sphere corresponding to identical appearance for the object are merged into the same class.

A single network is therefore trained to perform both 2D object detection and 3D pose prediction, using a single loss function made of a weighted sum of different terms. The weights are hyperparameters and have to be tuned, which can be difficult. However, having a single network is an elegant solution, and more importantly for practical applications, this allows to save computation times: Image feature extraction, the slowest part of the network, is performed only once even if it is used to predict the 2D bounding boxes and the 3D pose estimation.

A refinement procedure is then run on each object detection to improve the pose estimate predicted by the classifier. SSD-6D relies on a method inspired by an early approach to 3D object tracking based on edges [17]. Data augmentation was done by rendering synthetic views of the objects using their 3D models over images from COCO [53], which helps but only to some extent as the differences between the real and synthetic images (the "domain gap") remain high.

4.7.3 YOLO-6D

The method proposed in [98], sometimes referred as YOLO-6D, is relatively similar to SSD-6D but makes different choices that makes it faster and more accurate. It relies on the YOLO architecture [78, 79] for 2D object detection and predicts the 3D object poses in a form similar to the one of BB8. The authors report that training the network using this pose representation was much simpler than when using quaternions to represent the 3D rotation. As YOLO is much faster and as they do not discretize the pose space, [98] also reports much better performance than SSD-6D in terms of both computation times and accuracy.

4.7.4 PoseCNN

PoseCNN is a method proposed in [113]. It relies on a relatively complex architecture, based on the idea of decoupling the 3D pose estimation task into different subtasks. This architecture predicts for each pixel in the input image (1) the object label, (2) a unit vector toward the 2D object center, and (3) the 3D distance between the object center and the camera center, plus (4) 2D bounding boxes for the objects and (5) a 3D rotation for each bounding box in the form of a quaternion. From (1), (2), and (3), it is possible to compute the 3D translation vector for each visible object, which is combined with the 3D rotation to obtain the full 3D pose for each visible object.

Maybe the most interesting contribution of PoseCNN is the loss function. It uses the ADD metric as the loss (see Sect. 4.6.4), and even more interestingly, the paper shows that the ADD-S metric, used to evaluate the pose accuracy for symmetrical objects (see Sect. 4.6.1), can be used as a loss function to deal with symmetrical objects. The ADI metric makes use of the min operator, which makes it maybe an unconventional loss function, but it is still differentiable and can be used to train a network. This results in an elegant and efficient way of handling object symmetries.

Another contribution of [113] is the introduction of the YCB-Video dataset (see Sect. 4.6.1).

4.7.5 DeepIM

DeepIM [52] proposes a refinement step that resembles the one in BB8 (see the discussion on refinement steps in Sect. 4.7.1) but with several main differences. As BB8, it trains a network to compare the input image with a rendering of an object that has already been detected and for which a pose estimate is already available. The network outputs an update for the 3D object pose that will improve the pose estimate, and this process can be iterated until convergence. The first difference with BB8 is that the rendered image has a high resolution and the input image is centered on the object and upscaled to the same resolution. This allows a higher precision when predicting the pose update. The loss function also includes terms to make the network output the flow and the object mask, to introduce regularization.

The second difference is more fundamental, as it introduces a coordinate system well suited to define the rotation update that must be predicted by the network. The authors first remark that predicting the rotation in the camera coordinate system is wrong because this rotation would also *translate* the object. They thus set the center of rotation to the object center. For the axes of the coordinate system, the authors remark that using those of the object's 3D model is not a good option, as they are arbitrary, and this would force the network to learn them for each object. They therefore propose to use the axes of the camera coordinate system, which makes the network generalize much better. The translation update is predicted as a 2D translation on the image plane, plus a delta along the z axis of the camera in a log-scale. To learn this update, DeepIM uses the same loss function as [113] (ADD, or ADD-S for symmetrical objects). This approach is even shown to generalize to unseen objects, but this is admittedly demonstrated in the paper only on very simple object renderings.

4.7.6 Augmented Autoencoders

The most interesting aspect of the Augmented Autoencoders method proposed in [92] is the way it deals with symmetrical objects. It proposes to first learn an embedding that can be computed from an image of the object. This embedding should be robust to imaging artefacts (illumination, background, etc.) and depends only on the object's appearance: The embeddings for two images of the object under ambiguous rotations should thus be the same. At run time, the embedding for an input image can then be mapped to all the possible rotations. This is done efficiently by creating a codebook offline, by sampling views around the target objects, and associating the embeddings of these views to the corresponding rotations. The translation can be recovered from the object bounding box's 2D location and scale.

To learn to compute the embeddings, [92] relies on an autoencoder architecture. An encoder network predicts an embedding from an image of an object with different image nuisances; a decoder network takes the predicted embedding as input and should generate the image of the object under the same rotation but without the nuisances. The encoder and decoder are trained together on synthetic images of the objects, to which nuisances are added. The loss function encourages the composition of the two networks to output the original synthetic images, despite using the images with nuisances as input. Another motivation for this autoencoder architecture is to learn to be robust to nuisances, even though a more standard approach outputting the 3D pose would probably do the same when trained on the same images.

4.7.7 Robustness to Partial Occlusions: Oberweger's Method, Segmentation-Driven MeThod, PVNet

Oberweger et al. [64] (Oberweger's method), [41] (segmentation-driven method), and PVNet [70] developed almost in parallel similar methods that provide accurate 3D poses even under large partial occlusions. Indeed, [64] shows that deep networks can be robust to partial occlusions when predicting a pose from an image containing the target object when trained on examples with occlusions but only to some extent. To become more robust and more accurate under large occlusions, [41, 64, 70] proposed to predict the 3D pose in the form of the 2D reprojections of some 3D points as in BB8 but by combining multiple predictions, where each prediction is performed from different local image information. The key idea is that local image information that is not disturbed by occlusions will result into good predictions; local image information disturbed by occlusions will predict erroneous reprojections, but these can be filtered out with a robust estimation.

Oberweger et al. [64] predicts the 2D reprojections in the form of heatmaps, to handle the ambiguities of mapping between image locations and the reprojections, as many image locations can look the same. However, this makes predictions relatively slow. Instead, [41] predicts a single 2D displacement between the image location and each 2D reprojection. These result in many possible reprojections, some noisy, but the correct 3D pose can be retrieved using RANSAC. PVNet [70] chose to predict the directions toward the reprojections, rather than a full 2D displacement, but relies on a similar RANSAC procedure to estimate the final 3D pose. Hu et al. [41] and Peng et al. [70] also predict the masks of the target objects, in order to consider the predictions from the image locations that lie on the objects, as they are the only informative ones.

4.7.8 DPOD and Pix2Pose

DPOD [117] and Pix2Pose [68] are also two methods that have been presented at the same conference and present similarities. They learn to predict for each pixel of an input image centered on a target object its 3D coordinates in the object coordinate system. More exactly, DPOD predicts the pixels' texture coordinates, but this is fundamentally the same thing as they both provide 2D-3D correspondences between image locations and their 3D object coordinates. Such representation is often called "object coordinates" and more recently "location field" and has already been used for 3D pose estimation in [10, 11] and before that in [97] for human pose estimation. From these 2D-3D correspondences, it is then possible to estimate the object pose using RANSAC and P*n*P. Park et al. [68] relies on a GAN [22] to learn to perform this prediction robustly even under occlusion. Note that DPOD has been demonstrated to run in real time on a tablet.

Pix2Pose [68] also uses a loss function that deals with symmetrical objects and can be written as

$$\mathcal{L} = \min_{\mathbf{R} \in \mathrm{Sym}} \frac{1}{|\mathcal{V}|} \sum_{\mathbf{M} \in \mathcal{V}} \|\mathrm{Tr}(\mathbf{M};\hat{\mathbf{p}}) - \mathrm{Tr}(\mathbf{M};\mathbf{R}.\bar{\mathbf{p}})\|_2 , \quad (4.5)$$

where $\hat{\mathbf{p}}$ and $\bar{\mathbf{p}}$ denote the predicted and ground truth poses, respectively, Sym is a set of rotations corresponding to the object symmetries, and $\mathbf{R}.\mathbf{p}$ denotes the composition of such a rotation and a 3D pose. This deals with symmetries in a much more satisfying way than the ADD-S loss (Eq. (4.3)).

4.7.9 Discussion

As can be seen, the last recent years have seen rich developments in 3D object pose estimation, and methods have become even more robust, more accurate, and faster. Many methods rely on 2D segmentation to detect the objects, which seems pretty robust but assumes that only several instances of the same object do not overlap in the image. Most methods also rely on a refinement step, which may slow things down but relax the need for having a single strong detection stage.

There are still several caveats though. First, it remains difficult to compare methods based on Deep Learning, even though the use of public benchmarks helped to promote fair comparisons. Quantitative results depend not only on the method itself but also on how much effort the authors put into training their methods and augmenting training data. Also, the focus on the benchmarks may make the method overfit to their data, and it is not clear how well current methods generalize to the real world. Also, these methods rely on large numbers of registered training images and/or on textured models of the objects, which can be cumbersome to acquire.

4.8 3D Pose Estimation for Object Categories

So far, we discussed methods that estimate the 3D pose of specific objects, which are known in advance. As we already mentioned, this requires the cumbersome step of capturing a training set for each object. One possible direction to avoid this step is to consider a "category"-level approach, for objects that belong to a clear category, such as "car" or "chair." The annotation burden moves then to images of objects from the target categories, but we can then estimate the 3D pose of new, unseen objects from these categories.

Some early category-level methods only estimate 3 degrees of freedom (2 for the image location and 1 for the rotation over the ground plane) of the object pose using regression, classification or hybrid variants of the two. For example, [111] directly regresses azimuth, elevation, and in-plane rotation using a convolutional neural network (CNN), while [104, 105] perform viewpoint classification by discretizing the range of each angle into a number of disjoint bins and predicting the most likely bin using a CNN.

To estimate a full 3D pose, many methods rely on "semantic keypoints," which can be detected in the images and correspond to 3D points on the object. For example, to estimate the 3D pose of a car, one may consider the corners of the roof and the lights as semantic keypoints. Pepik et al. [71] recovers the pose from keypoint predictions and CAD models using a P*n*P algorithm, and [69] predicts semantic keypoints and trains a deformable shape model which takes keypoint uncertainties into account. Relying on such keypoints, however, is even more depending in terms of annotations as keypoints have to be carefully chosen and manually located in many images and as they do not generalize across categories.

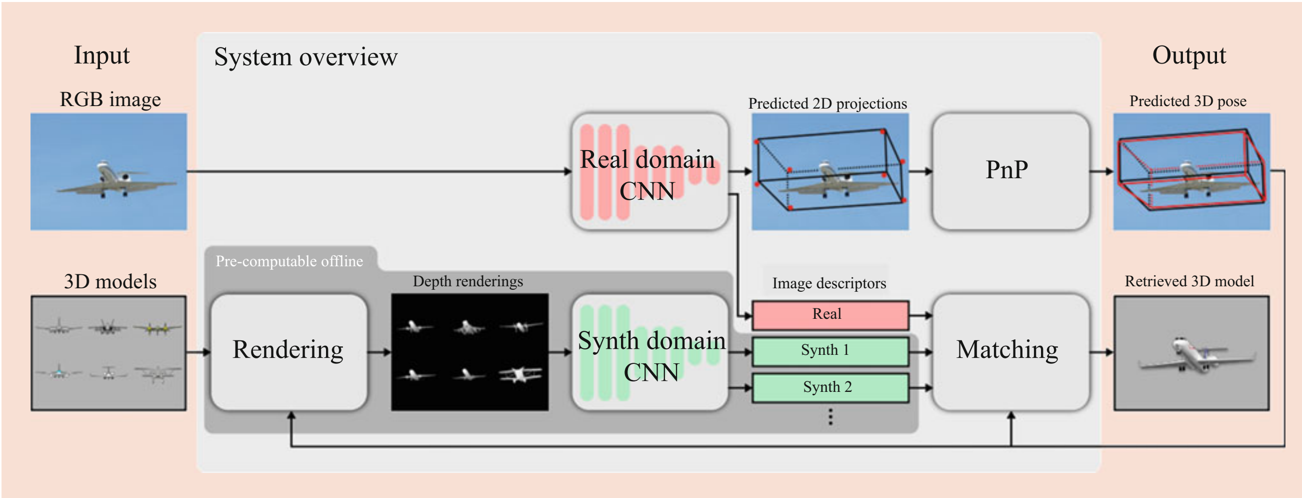


Fig. 4.9 Estimating the 3D pose of an unknown object from a known category. Grabner et al. [25] predicts the 2D reprojections of the corners of the 3D bounding box *and* the size of the bounding box. For this information, it is possible to compute the pose of the object using a PnP algorithm. (Image from [25])

In fact, if one is not careful, the concept of 3D pose for object categories is ill defined, as different objects from the same category are likely to have different sizes. To properly define this pose, as shown in Fig. 4.9, [25] considers that the 3D pose of an object from a category is defined as the 3D pose of its 3D bounding box. It then proposes a method that extends BB8 [76]. BB8 estimates a 3D pose by predicting the 2D reprojections of the corners of its 3D bounding box. Grabner et al. [25] predicts similar 2D reprojections, plus the size of the 3D bounding box in the form of three values (length, height, width). Using these three values, it is possible to compute the 3D coordinates of the corners in a coordinate system related to the object. From these 3D coordinates, and from the predicted 2D reprojections, it is possible to compute the 3D pose of the 3D bounding box.

3D Model Retrieval Since objects from the same category have various 3D models, it may also be interesting to recover a 3D model for these objects, in addition to their 3D poses. Different approaches are possible. One is to recover an existing 3D model from a database that fits well the object. Large datasets of light 3D models exist for many categories, which makes this approach attractive [15]. To make this approach efficient, it is best to rely on metric learning, so that an embedding computed for the object from the input image can be matched against the embeddings for the 3D models [4, 25, 43]. If the similarity between these embeddings can be estimated based on their Euclidean distance or their dot product, then it is possible to rely on efficient techniques to perform this match.

The challenge is that the object images and the 3D models have very different natures, while this approach needs to compute embeddings that are comparable from these two sources. The solution is then to replace the 3D models by image renderings; the embeddings for the image renderings can then be compared more easily with the embeddings for the input images. The domain gap remains between the real input images and the synthetic image renderings. In [26], illustrated in Fig. 4.10, this is solved by predicting the object coordinates for the input image using a deep network, and by rendering the object coordinates for the synthetic images, rather than a regular rendering. This brings the representations for the input images and the 3D models even closer. From these representations, it is then easier to compute suitable embeddings.

Another approach is to predict the object 3D geometry directly from images. This approach learns a mapping from the appearance of an object to its 3D geometry. This is more appealing than recovering a 3D model from a database that has no guarantee to fit perfectly the object, since this latter method can potentially adapt better to the object's geometry [27, 67, 86]. This is however much more challenging, and current methods are still often limited to clean images with blank background and sometimes even synthetic renderings; however, this is a very active area, and more progress can be expected in the near future.

4.9 3D Hand Pose Estimation from Depth Maps

While related to 3D object pose estimation, hand pose estimation has its own specificities. On one hand (no pun intended), it is more challenging, if only because more degrees of freedom must be estimated. On the other hand, while considering new 3D objects still requires acquiring new data and/or specific 3D models, a single model can generalize to

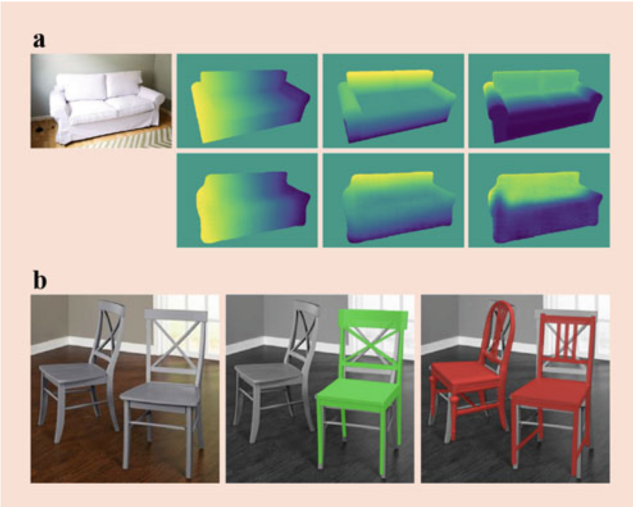


Fig. 4.10 Recovering a 3D model for an unknown object from a known category using an image. (**a**) [26] generates renderings of the object coordinates (called location fields in the paper) from 3D models (top) and predicts similar object coordinates from the input images (bottom). Images of objects and 3D models are matched based on embeddings computed from the object coordinates. (**b**) An example of recovered 3D poses and 3D models for two chairs. The 3D models capture the general shapes of the real chairs but do not fit perfectly

unseen hands, despite differences in size, shape, and skin color, as these differences remain small and considerably smaller than differences between two arbitrary objects. In practice, this means that a model does not need new data nor retraining to adapt to new users, which is very convenient.

However, the motivation for 3D hand pose in AR applications is mostly for making possible natural interaction with virtual objects. As we will briefly discuss in Sect. 4.10.6, convincing interaction requires very high accuracy, which current developments are aiming to achieve.

We will start with 3D hand pose estimation from a depth map. As for 3D object pose estimation, the literature is extremely large, and we will focus here as well on a few representative methods. This is the direction currently taken in the industry, because it can provide a robust and accurate estimate. It is currently being deployed on commercial solutions, such as HoloLens 2 and Oculus. We will then turn to hand pose estimation from color images, which is more challenging but has also the potential to avoid the drawbacks of using depth cameras.

4.9.1 DeepPrior++

Estimating the 3D pose of a hand from a depth map is in fact now relatively easy and robust, when using a well-engineered solution based on Deep Learning. For example, [64] discusses the implementation of a method called DeepPrior++, which is simple in principle and performs well.

DeepPrior++ first selects a bounding box on the hand and then predicts the 3D joint locations from the depth data within this bounding box. An accurate localization of the bounding appears to improve the final accuracy, and the method from the center of mass of the depth data within a threshold, and then uses a refinement step that can be iterated. This refinement step is performed by a regression network predicting, from a 3D bounding box centered on the center of mass, the 3D location of one of the joints used as a referential. Then, the 3D joint locations can be predicted using a second network that is not very different from the one presented in Fig. 4.4, except that it relies on ResNet block [33] for more accurate results. The input to this network is therefore the input depth map cropped to contain only the hand, and its output is made of the 3D coordinates of all the hand joints.

Simple data augmentation appears to improve accuracy, even when starting from an already large dataset. Oberweger et al. [64] applies random in-plane rotations, scales, and 3D offsets to the input data.

4.9.2 V2V-PoseNet

V2V-PoseNet [59] is a method that has been identified by Xiong et al. [115] as one of the best performing methods at the time, based on the results on a public benchmark. Moon et al. [59] argues that using a cropped depth map as input to the network makes training difficult, as the same hand pose appears (slightly) differently in a depth map depending on where it reprojects.

To avoid distortion, V2V-PoseNet converts the depth map information in the hand bounding box into a voxel-based representation (it uses a refinement step similar to the one in [64], so accuracy of the hand detection seems important). The conversion is done by voxelizing the 3D bounding box for the hand, and each voxel that contains at least one pixel from the depth map is set to occupied. The voxel-based representation therefore corresponds to a discretization of the hand surface.

Then, a network is trained to take this voxel-based representation as input and predicts the 3D joint locations in the form of 3D heatmaps. A 3D heatmap is a 3D array containing confidence values. There is one 3D heatmap for each joint, and the joint location is taken as the 3D location with the higher confidence.

4.9.3 A2J

The A2J method [115] performs similarly or better than V2V-PoseNet but faster. The prediction of the 3D joint locations is performed via what are called in the paper “anchor points.” Anchor points are 2D locations regularly sampled over the bounding box of the hand, obtained as in [59] and [64]. A network is trained to predict, for each anchor point, and for each joint, (1) a 2D displacement from the anchor point to the joint, (2) the depth of the joint, and (3) a weight reflecting the confidence for the prediction. The prediction of 2D displacements makes this method close to [41] and PVNet [70], which were developed for 3D object pose estimation under occlusions.

From the output of the network, it is possible to estimate the 3D joint locations as weighted sums of the 2D locations and depths, where the weights of the sums are the ones predicted by the network. Special care is taken to make sure the weights sum to 1. The paper argues that this is similar to predictions by an "ensemble," which are known to generalize well [13]. The weights play a critical role here. Without them, the predicted locations would be a linear combination of the network output, which could be done by a standard layer without having to introduce anchor points. The product between weights and 2D displacements and depths make the predicted joint locations a more complex, nonlinear transformation of the network output.

4.9.4 Discussion

As discussed in detail in [2], the coverage of the space of possible hand poses is critical to achieve good performance. This should not be surprising: Deep networks essentially learn a mapping between the input (here the depth map) and the output (here the hand pose) from the training set. They can interpolate very well between samples but poorly extrapolate: if the input contains a pose too different from any sample in the training set, the output is likely to be wrong. When using as input depth maps, which do not suffer from light changes or cluttered background like color images, having a good training set is the main important issue.

That was in fact already discussed in [64] and even before that in [94]. Combining the DeepPrior++ that relies on a simple mapping with a simple domain adaptation method [77] in order to use more synthetic data appears to perform very well and still outperforms more complex methods on the NYU dataset, for example. While it is not entirely clear, the direction taken by HoloLens 2 and Oculus is to train a simple model on a large set of synthetic training data and refine the pose estimate using an ICP algorithm [85] aligning a 3D model of the hand to the depth data.

4.10 3D Hand Pose Estimation from an RGB Image

We now turn to modern approaches to 3D hand pose estimation from an RGB image, which is significantly more challenging than from a depth map but which is a field that has been impressively fast progress over the past few years.

4.10.1 Zimmerman's Method

Zimmermann and Brox [119] was one of the first methods predicting the 3D hand pose from a single color image, without using depth information. The hand is first segmented in 2D, as in recent approaches for 3D object pose estimation, to obtain a 2D bounding box centered on it. From this window, a 2D heatmap is predicted for each joint, using convolutional pose machines [110]. This provides information on the 2D joint locations. To obtain the final 3D pose, the heatmaps are lifted in 3D by using two networks: one network predicts the 3D joint locations in a canonical system attached to the hand; the other network predicts the 3D pose (3D rotation and translation) of the hand in the camera coordinate system. From these two sets of information, it is possible to compute the 3D joint coordinates in the camera coordinate system. This is shown to work slightly better than a single network directly predicting the 3D joint locations in the camera coordinate system.

To train the segmentation, heatmap prediction, and 3D pose predictions, [119] created a dataset of synthetic images of hands, as the 3D pose can be known exactly. To do so, they used freely available 3D models of humans with corresponding animations from Mixamo, a company specialized in character animation, and Blender to render the images over random background. Lighting, viewpoint, and skin properties were also randomized. The dataset is publicly available online.

4.10.2 Iqbal's Method

Like [42, 119] relies on 2D heatmaps from a color image to lift them to 3D, but with several differences that improve the pose accuracy. They oppose two approaches to 3D pose prediction: The first approach relies on heatmaps; it is generally accurate, but keeping this accuracy with 3D heatmaps is still intractable, as a finely sampled 3D volume can be very large. The second approach is to predict the 3D joints in a "holistic" way, meaning that the values of the 3D joint locations are directly predicted from the input image, as was done in [91, 103] for human body pose prediction and [63] for hand pose prediction from depth data. Unfortunately, "holistic" approaches tend to be less accurate.

To keep the accuracy of heatmaps, and to predict the joint depths with a tractable method, [42] also predicts, in addition to the 2D heatmaps, a depth map for each joint, where the depth values are predicted depths for the corresponding joint relative to a reference joint. The final 3D pose is computed using geometric constraints from this information. Instead of learning to predict the heatmaps in a supervised way, [42] learns to predict "latent" heatmaps. This means that the deep model is pretrained to predict heatmaps and depth maps, but the constraints on the predicted heatmaps and depth maps are removed from the full loss, and the model is trained “end to end,” with a loss involving only the final predicted 3D pose. This lets the optimization find heatmaps and depth maps that are easier to predict while being more useful to predict the correct pose.

4.10.3 GANerated Hands

As already mentioned above, training data and its diversity is of crucial importance in Deep Learning. For 3D hand pose estimation from color images, this means that training data need to span the hand pose space, the hand shape space, possible lighting, possible background, etc. Moreover, this data has to be annotated in 3D, which is very challenging. Generating synthetic images is thus attractive, but because of the “domain gap” between real and synthetic images, this may impact the accuracy of a model trained on synthetic images and applied to real images.

To bridge the domain gap between synthetic and real images, [61] extends CycleGAN [118], which is itself based on generative adversarial networks (GANs). In original GANs [22], a network is trained to generate images from random vectors. In order to ensure the generated images are “realistic,” a second network is trained jointly with the first one to distinguish the generated images from real ones. When this network fails and cannot classify the images generated by the first network as generated, it means these generated images are similar to real ones. CycleGAN [118] builds on this idea to change an image from a domain into an image with the *same content* but similar to the images from another domain. An image generator is thus trained to take an image from some domain as input and to generate as output another image. A second network ensures that the output looks similar to the target domain. To make sure the content of the image is preserved, a second pair of networks is jointly trained with the first one: This pair is trained jointly with the first one, to regenerate the original input image from the image generated by the first generator.

Mueller et al. [61] extends CycleGAN by adding a constraint between the original image (here a synthetic image) and the image returned by the first generator. The synthetic image is a hand rendered over a uniform background, and a segmentation network is trained jointly to segment the generated image, so that the predicted segmentation matches the segmentation of the hand in the synthetic image. This helps preserving, at least to some extent, the 3D pose of the hand in the transformed image. Another advantage is that the plain background can be changed by a random background to augment the dataset of transformed images.

Besides this data generation, the 3D hand pose prediction in [61] is also interesting. A network is trained to predict *both* 2D heatmaps for the 2D joint locations and the 3D joint locations. Given the output of this network over a video stream, an optimization is performed to fit a kinematic model to these 2D and 3D predictions under joint angle constraints and temporal smoothness constraints. The output of the network is therefore not used directly as the 3D hand pose but used as an observation in this final optimization.

4.10.4 3D Hand Shape and Pose Estimation: Ge’s and Boukhayma’s Methods

Ge et al. [21] and Boukhayma et al. [9] were published at the same conference (CVPR’19), and both propose a method for predicting the hand shape in addition to the hand pose.

Ge et al. [21] first generates a training set of hands under various poses, where each image is annotated with the pose and shape parameters, with special care to make the rendered images as realistic as possible. Using this dataset, they train a network to predict a “latent feature vector,” which is fed to a graph CNN [16]. The motivation for using a graph CNN is to predict a 3D mesh for the hand, which can be represented as a graph. The loss function thus compares, for synthetic training images, the mesh predicted by the graph CNN and the ground truth mesh and the predicted 3D pose and the ground truth pose. For real training images, which do not have ground truth mesh available, they use the pose loss term but also a term that compares the predicted 3D mesh with the depth map for the input RGB image when available and another term that compares the same predicted 3D mesh with a mesh predicted from the ground truth 2D heatmaps.

The method of [9] is simpler. It relies on the MANO model [81], which is a popular deformable and parameterized model of the human hand, created from many captures of real hands. The MANO model provides a differentiable function that generates a 3D model of a hand, given pose parameters and shape parameters. It does not require a dataset annotated with the shape parameters, but instead it can reuse existing datasets, annotated only with the 3D hand pose. To do so, they train a network to predict the hand shape and pose parameters, and the loss function is the 2D distances between the reprojections of the 3D joints, as computed from the predicted shape and pose parameters and the annotated 2D joints.

4.10.5 Implementation in MediaPipe

We can also mention a Google Project [7] for real-time 3D hand pose implementation on mobile devices. They first detect the palm, rather than the full hand: This allows them to only consider squared bounding boxes and to avoid confusion between multiple hand detections. The 3D hand pose is directly predicted as the 3D joint locations, using a model trained on real and synthetic images.

Unfortunately, the work is not presented in a formal research publication, and it is difficult to compare their results with the ones obtained by other methods. However, the implementation is publicly available and appears to work well. This may show again that good engineering can make simple approaches perform well.

4.10.6 Manipulating Virtual Objects

One of the main motivations to track the 3D pose of a hand for an AR system is to make possible natural interaction with virtual objects. It is therefore important to be able to exploit the estimated 3D hand pose to compute how the virtual objects should move when they are in interaction with the user's hand.

This is not a trivial problem. In the real world, our ability to interact with objects is due to the presence of friction between the surfaces of the objects and our hands, as friction is a force that resists motion. However, friction is very challenging to simulate correctly. Also, nothing prevents the real hand to penetrate the virtual objects, which makes realism fail.

Some approaches rely on heuristics to perform a set of object interactions. It may be reasonable to focus on object grasping for AR systems, as it is the most interesting interaction. Some methods are data driven and synthesize prerecorded, real hand data to identify the most similar one during runtime from a predefined database [51, 58, 80].

A friction model from physics, the *Coulomb-Contensou* model was used in [95] to reach more accurate results, but the computational complexity becomes very high and not necessarily compatible with real-time constraints. Hoell et al. [39] relies on the simpler Coulomb model that appears to be a good trade-off between realism and tractability. The force applied by the user is taken proportional to how much their real hand penetrates the virtual objects. Computing accurately the forces requires very accurate 3D hand poses; otherwise they can become very unstable.

Very recently, Oculus has introduced real-time interaction with virtual objects using depth-based hand tracking on the Quest 2 headset [65]. No technical details are available, but the system allows for relatively realistic grasping-based interactions.

4.11 3D Object+Hand Pose Estimation

We finally turn to the problem of estimating a 3D pose for both a hand and an object, when the hand directly manipulates the object. Even if the existing methods are still preliminary, the potential application to AR is very appealing, as this could offer tangible interfaces by manipulating objects or even the possibility to bring real objects into the virtual world.

This problem still remains very challenging, as the close contact between the hand and the object results in mutual occlusions that can be large. Dealing with egocentric views, which are more relevant for AR applications, is often even more challenging. Fortunately, there are also physical constraints between the hand and the object, such as impossibility of penetration but also natural grasping poses, which may help solving this problem.

Pioneered approaches for joint hand+object pose estimation [5, 66, 109] typically relied on frame-by-frame tracking and, in the case of [48, 106], also on a physics simulator to exploit the constraints between the hand and object. However, frame-to-frame tracking, when used alone, may require careful initialization from the user and may drift over time. The ability to estimate the 3D poses from a single image without prior from previous frames or from the user makes tracking more robust. Given the difficulty of the task, this has been possible only with the use of Deep Learning.

4.11.1 ObMan and HOPS-Net

ObMan [32] and HOPS-Net [46] almost simultaneously proposed methods with some similarities for estimating the 3D pose *and shape* of both a hand and of the object it manipulates from a single RGB image. Creating realistic training data is the first challenge, and they both created a dataset of synthetic images of hands and objects using the *GraspIt!* simulation environment [57] for generating realistic grasps given 3D models for the hand and for the object.

Hasson et al. [32] train their model on their synthetic dataset before fine-tuning it on [20], which provide annotated real images. To make the synthetic images more realistic, [46] use "augmented CycleGAN" [1]: The motivation to use augmented CycleGAN rather than CycleGAN is that the former can deal with many-to-many mappings, while the latter considers only one-to-one mappings and that a synthetic image can correspond to multiple real images.

Hasson et al. [32] does not go through an explicit 2D detection of the object or the hand, by contrast with the other methods we already discussed. Instead, the method directly predicts from the input image the pose and shape parameters for the hand (using the MANO model) and the shape for the object. To predict the object shape, they use AtlasNet [27], a method that can predict a set of 3D points from a color image. This shape prediction does not require to know the object in advance and predicts the shape in the camera coordinates system; there is therefore no notion of object pose. The loss function to train the architecture includes a term that prevents intersection between the hand and the object and a term that penalizes cases in which the hand is close to the surface of the object without being in contact.

Kokic et al. [46] focuses on objects from specific categories (bottle, mug, knife, and bowl) and uses Mask R-CNN [34] to detect and segment the object: The object does not have to be known as long as it belongs to a known category. The method also relies on [119] to predict the 3D hand pose. A network is trained to predict, from the bounding box predicted by Mask-RCNN and the hand pose (to provide additional constraints), the 3D object pose and a description vector for the shape. The 3D pose representation depends on the category to handle symmetries in the object shape, and the description vector is matched against a database of 3D models to retrieve a 3D model with a similar shape (see [26] discussed in Sect. 4.8). A final refinement step optimizes the object pose so that it is in contact with the hand.

4.11.2 H+O

H+O [99] is a method that can predict the 3D poses of a hand and an object from an image, but it also learns a temporal model to recognize the performed actions (e.g., pouring, opening, or cleaning) and to introduce temporal constraints to improve the pose estimation. Despite this, the architecture remains relatively simple. It is trained on the First-Person Hand Action dataset (see Sect. 4.6.3).

The 3D pose of the object is predicted as the 3D locations of the corners of the 3D bounding box – from this information, it is possible to compute a 3D rotation and translation, using [24], for example. Since the 3D hand pose can also be predicted as a set of 3D locations, this makes the two outputs of the network consistent and avoids a weight to tune between the loss term for the hand and the loss term for the object. More exactly, the 3D locations are not predicted directly. The 3D space is split into cells, and for each cell and each 3D location, it is predicted if the 3D location is within this cell, plus an offset between a reference corner of the cell and the 3D location – this offset ensures that the 3D location can be predicted accurately, even with the space discretization into cells. A confidence is also predicted for each 3D location prediction. The 3D hand and object poses are computed by taking the 3D points with the highest confidences.

To enforce both temporal constraints and recognize the actions, the method relies on a recurrent neural network and more exactly on a long short-term memory (LSTM) [37]. Such network can propagate information from the predicted 3D locations over time, and it is used here to predict the action and the nature of the object, in addition to be a way to learn constraints between the hand and the object.

4.11.3 HOnnotate

Hampali et al. [29] proposed a method called HOnnotate to automatically annotate real images of hands grasping objects with their 3D poses (see Fig. 4.6f), which works with a single RGB-D camera but can exploit more cameras if available for better robustness and accuracy. The main idea is to optimize jointly all the 3D poses of the hand and the object over the sequence to exploit temporal consistency. The method is automated, which means that it can be used easily to labeled new sequences.

The authors use the resulting dataset called HO-3D to learn to predict from a single RGB image the 3D pose of a hand manipulating an object. This is done by training a deep network to predict the 2D joint locations of the hand along with the joint direction vectors and lift them to 3D by fitting a MANO model to these predictions. This reaches good accuracy despite occlusions by the object, even when the object was not seen during training.

4.12 The Future of 3D Object and Hand Pose Estimation

This chapter aimed at demonstrating and explaining the impressive development of 3D pose estimation in computer vision since the early pioneer works and in the context of potential applications to Augmented Reality. Methods are becoming more robust and accurate while benefiting from fast implementations on GPUs. We focused on some popular works, but they are only entries to many other works we could not present here and that the reader can explore on their own.

What is the Best Method? However, as we pointed out at the end of Sect. 4.7, it is still too early to draw conclusions on what is the best methodology for 3D pose estimation. Quantitative results not only reflect the contributions of the methods and also depend on how much effort the authors put in their implementation. As a result, there are sometimes contradictory conclusions between papers. For example, is it important to first detect the object or the hand in 2D first or not? Are 2D heatmaps important for accuracy or not? What is the best pose representation for a 3D pose: A 3D rotation and translation or a set of 3D points? Also, performance is not the only aspect here: for example, using a set of 3D points for both the object and hand poses as in [99] relaxes for tuning the weights in the loss function, which is also interesting in practice.

The focus on some public benchmarks may also bias the conclusions on the performance of the methods in the real world: How well do they perform under poor lighting? How well do they perform on different cameras or internal parameters without retraining?

About the Need for Training Data Another aspect that needs to be solved is the dependence on training sets, which may be complex to create, and on a long training stage each time a new object should be considered. The problem does not really occur for hands, as a large enough training set for hands would ensure generalization to unseen hands. For objects, variability in terms of shape and appearance is of course much more important.

Using domain randomization [100] seems to be key for using synthetic images for training a deep model, as the remarkable performance of the recent CosyPose method showed [49]. With domain randomization, large, unrealistic variations of the objects' appearances are used instead of trying to generate realistic synthetic images: it may be more difficult to train a network on such images, but it is easy to create the training set under these conditions, and a network trained in this way is likely to generalize to new images.

Self-supervised training is another direction to avoid annotating a training set. Self-supervised approaches exploit temporal and geometrical constraints to automatically find 3D poses that explain images from video sequences or multiple viewpoints while training a network predicting the poses [47, 89, 108]. A few-shot approach may also reduce the need for the number of images needed for training [114].

About the Need for a Training Stage Some methods have shown some generalization power within known categories [26, 46]. Xiang et al. [113] has shown some generalization to new objects but only on simple synthetic images and for known 3D models. A few other approaches aim at skipping the training stage even for object from unknown categories [73, 93]. Being able to consider instantly (without retraining nor capturing a 3D model) any object in an AR system is a dream that will probably become accessible in the next years.

References

1. Almahairi, A., Rajeswar, S., Sordoni, A., Bachman, P., Courville, A.: Augmented CycleGAN: Learning many-to-many mappings from unpaired data. In: International Conference on Machine Learning (2018)
2. Armagan, A., Garcia-Hernando, G., Baek, S., Hampali, S., Rad, M., Zhang, Z., Xie, S., Chen, M., Zhang, B., Xiong, F., Xiao, Y., Cao, Z., Yuan, J., Ren, P., Huang, W., Sun, H., Hrúz, M., Kanis, J., Krňoul, Z., Wan, Q., Li, S., Yang, L., Lee, D., Yao, A., Zhou, W., Mei, S., Liu, Y., Spurr, A., Iqbal, U., Molchanov, P., Weinzaepfel, P., Brégier, R., Rogez, G., Lepetit, V., Kim, T.-K.: Measuring generalisation to unseen viewpoints, articulations, shapes and objects for 3D hand pose estimation under hand-object interaction. In: European Conference on Computer Vision. Springer, Berlin (2020)
3. ARToolKit. http://www.hitl.washington.edu/artoolkit/
4. Aubry, M., Russell, B.: Understanding deep features with computer-generated imagery. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
5. Ballan, L., Taneja, A., Gall, J., Van Gool, L., Pollefeys, M.: Motion capture of hands in action using discriminative salient points. In: European Conference on Computer Vision. Springer, Berlin (2012)
6. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Computer Vision and Image Understanding, pp. 346–359 (2008)
7. Bazarevsky, V., Zhang, F.: Google Project on Hand Pose Prediction (2019)
8. Billinghurst, M., Kato, H., Poupyrev, I.: The magicbook: a transitional AR interface. Comput. Graphics **25**, 745–753 (2001)
9. Boukhayma, A., De Bem, R., Torr, P.H.S.: 3D hand shape and pose from images in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)
10. Brachmann, E., Krull, A., Michel, F., Gumhold, S., Shotton, J., Rother, C.: Learning 6D object pose estimation using 3D object coordinates. In: European Conference on Computer Vision. Springer, Berlin (2014)
11. Brachmann, E., Michel, F., Krull, A., Yang, M.M., Gumhold, S., Rother, C.: Uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)
12. Brégier, R., Devernay, F., Leyrit, L., Crowley, J.L.: Defining the pose of any 3D rigid object and an associated distance. Int. J. Comput. Vis. **126**(6), 571–596 (2018)
13. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
14. Calli, B., Singh, A., Bruce, J., Walsman, A., Konolige, K., Srinivasa, S., Abbeel, P., Dollar, A.M.: Yale-CMU-Berkeley dataset for robotic manipulation research. Int. J. Rob. Res. **36**(3), 261–268 (2017)
15. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet: An Information-Rich 3D Model Repository. In: arXiv Preprint (2015)
16. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in Neural Information Processing Systems (2016)
17. Drummond, T., Cipolla, R.: Real-time visual tracking of complex structures. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 932–946 (2002)
18. Eldan, R., Shamir, O.: The power of depth for feedforward neural networks. In: Conference on Learning Theory (2016)
19. Gao, X., Hou, X., Tang, J., Cheng, H.: Complete solution classification for the perspective-three-point problem. IEEE Trans. Pattern Anal. Mach. Intell. **25**(8), 930–943 (2003)
20. Garcia-Hernando, G., Yuan, S., Baek, S., Kim, T.-K.: First-person hand action benchmark with RGB-D videos and 3D hand pose annotations. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–419 (2018)
21. Ge, L., Ren, Z., Li, Y., Xue, Z., Wang, Y., Cai, J., Yuan, J.: 3D hand shape and pose estimation from a single RGB Image. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)
22. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, A.Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (2014)
23. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge, MA 02142 (2016)
24. Gower, J.C.: Generalized procrustes analysis. Psychometrika **40**(1), 33–51 (1975)
25. Grabner, A., Roth, P.M., Lepetit, V.: 3D pose estimation and 3D model retrieval for objects in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)

26. Grabner, A., Roth, P.M., Lepetit, V.: Location field descriptors: single image 3D model retrieval in the wild. In: IEEE International Conference on 3D Vision (2019)
27. Groueix, T., Fisher, M., Kim, V., Russell, B., Aubry, M.: Atlasnet: A Papier-Mâché approach to learning 3D surface generation. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
28. Gustus, A., Stillfried, G., Visser, J., Jörntell, H., Van Der Smagt, P.: Human hand modelling: kinematics, dynamics, applications. Comput. Anim. **106**(11), 741–755 (2012)
29. Hampali, S., Rad, M., Oberweger, M., Lepetit, A.V.: HOnnotate: A method for 3D annotation of hand and object poses. In: IEEE Conference on Computer Vision and Pattern Recognition (2020)
30. Harris, C., Stennett, C.: RAPiD-A video rate object tracker. In: British Machine Vision Conference (1990)
31. Harris, C.G., Stephens, M.J.: A combined corner and edge detector. In: Fourth Alvey Vision Conference (1988)
32. Hasson, Y., Varol, G., Tzionas, D., Kalevatykh, I., Black, M.J., Laptev, I., Schmid, C.: Learning joint reconstruction of hands and manipulated objects. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)
33. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)
34. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: IEEE International Conference on Computer Vision (2017)
35. Hinterstoisser, S., Cagniart, C., Ilic, S., Sturm, P.F., Navab, N., Fua, P., Lepetit, V.: Gradient response maps for real-time detection of textureless objects. IEEE Trans. Pattern Anal. Mach. Intell. **34**(5) 876–888 (2012)
36. Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G.R., Konolige, K., Navab, N.: Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In: Asian Conference on Computer Vision. Springer, Berlin (2012)
37. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
38. Hodan, T., Haluza, P., Obdrzalek, S., Matas, J., Lourakis, M., Zabulis, X.: On evaluation of 6D object pose estimation. In: European Conference on Computer Vision. Springer, Berlin (2016)
39. Hoell, M., Oberweger, M., Arth, C., Lepetit, A.V.: Efficient physics-based implementation for realistic hand-object interaction in virtual reality. In: IEEE Conference on Virtual Reality (2018)
40. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks **2**(5), 359–366 (1989)
41. Hu, Y., Hugonot, J., Fua, P., Salzmann, M.: Segmentation-driven 6D object pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)
42. Iqbal, U., Molchanov, P., Breuel, T., Gall, J., Kautz, J.: Hand pose estimation via Latent 2.5D heatmap regression. In: European Conference on Computer Vision. Springer, Berlin (2018)
43. Izadinia, H., Shan, Q., Seitz, S.: Im2cad. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
44. Kehl, W., Manhardt, F., Tombari, F., Ilic, S., Navab, N.: SSD-6D: making RGB-Based 3D detection and 6D pose estimation great again. In: IEEE International Conference on Computer Vision (2017)
45. Kim, K., Grundmann, M., Shamir, A., Matthews, I., Hodgins, J., Essa, I.: Motion fields to predict play evolution in dynamic sport scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (2010)
46. Kokic, M., Kragic, D., Bohg, J.: Learning to estimate pose and shape of hand-held objects from RGB images. In: IEEE International Conference on Intelligent Robots and Systems (2019)
47. Kulon, D., Guler, R.A., Kokkinos, I., Bronstein, M., Zafeiriou, S.: Weakly-supervised mesh-convolutional hand reconstruction in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (2020)
48. Kyriazis, N., Argyros, A.A.: Physically plausible 3D scene tracking: the single actor hypothesis. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
49. Labbé, Y., Carpentier, J., Aubry, M., Sivic, J.: CosyPose: consistent multi-view multi-object 6D pose estimation. In: European Conference on Computer Vision. Springer, Berlin (2020)
50. Lepetit, V., Lagger, P., Fua, P.: Randomized trees for real-time keypoint recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (2005)
51. Li, Y., Fu, J.L., Pollard, N.S.: Data-driven grasp synthesis using shape matching and task-based pruning. IEEE Trans. Visual Comput. Graphics **13**(4), 732–747 (2007)
52. Li, Y., Wang, G., Ji, X., Xiang, Y., Fox, D.: DeepIM: deep iterative matching for 6D pose estimation. In: European Conference on Computer Vision. Springer, Berlin (2018)
53. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: common objects in context. In: European Conference on Computer Vision, pp. 740–755. Springer, Berlin (2014)
54. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C.-Y., Berg, A.C.: SSD: single shot multibox detector. In: European Conference on Computer Vision. Springer, Berlin (2016)
55. Lowe, D.G.: Fitting parameterized three-dimensional models to images. IEEE Trans. Pattern Anal. Mach. Intell. **13**(5), 441–450 (1991)
56. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **20**(2), 91–110 (2004)
57. Miller, A.T., Allen, P.K.: Graspit! a versatile simulator for robotic grasping. IEEE Rob. Autom. Mag. **11**(4), 110–122 (2004)
58. Miller, A.T., Knoop, S., Christensen, H.I., Allen, P.K.: Automatic grasp planning using shape primitives. In: International Conference on Robotics and Automation (2003)
59. Moon, G., Chang, J.Y., Lee, K.M.: V2V-PoseNet: Voxel-To-Voxel Prediction network for accurate 3D hand and human pose estimation from a single depth map. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
60. Mueller, F., Bernard, F., Sotnychenko, O., Mehta, D., Sridhar, S., Casas, D., Theobalt, C.: GANerated hands for real-time 3D hand tracking from monocular RGB. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
61. Mueller, F., Bernard, F., Sotnychenko, O., Mehta, D., Sridhar, S., Casas, D., Theobalt, C.: GANerated hands for real-time 3D hand tracking from monocular RGB. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
62. Oberweger, M., Wohlhart, P., Lepetit, V.: Training a feedback loop for hand pose estimation. In: IEEE International Conference on Computer Vision (2015)
63. Oberweger, M., Wohlhart, P., Lepeti, V.: DeepPrior++: improving fast and accurate 3D hand pose estimation. In: IEEE International Conference on Computer Vision, p. 2 (2017)
64. Oberweger, M., Rad, M., Lepetit, V.: Making deep heatmaps robust to partial occlusions for 3D object pose estimation. In: European Conference on Computer Vision. Springer, Berlin (2018)
65. Oculus. https://www.oculus.com/
66. Oikonomidis, I., Kyriazis, N., Argyros, A.A.: Full DoF tracking of a hand interacting with an object by modeling occlusions and physical constraints. In: IEEE International Conference on Computer Vision (2011)
67. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, A.S.: Deepsdf: learning continuous signed distance functions for shape representation. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)

68. Park, K., Patten, T., Vincze, M.: Pix2pose: pixel-wise coordinate regression of objects for 6D pose estimation. In: IEEE International Conference on Computer Vision (2019)
69. Pavlakos, G., Zhou, X., Chan, A., Derpanis, K., Daniilidis, K.: 6-DoF object pose from semantic keypoints. In: International Conference on Robotics and Automation, pp. 2011–2018 (2017)
70. Peng, S., Liu, Y., Huang, Q., Bao, H., Zhou, X.: PVNet: pixel-wise voting network for 6DoF pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)
71. Pepik, B., Stark, M., Gehler, P., Ritschel, T., Schiele, B.: 3D object class detection in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
72. Pinkus, A.: Approximation theory of the MLP model in neural networks. Comput. Anim **8**, 143–195 (1999)
73. Pitteri, G., Bugeau, A., Ilic, S., Lepetit, V.: 3D object detection and pose estimation of unseen objects in color images with local surface EMbeddings. In: Asian Conference on Computer Vision Springer, Berlin (2020)
74. Poirson, P., Ammirato, P., Fu, C.-Y., Liu, W., Kosecka, J., Berg, A.C.: Fast single shot detection and pose estimation. In: IEEE International Conference on 3D Vision (2016)
75. Pumarola, A., Agudo, A., Porzi, L., Sanfeliu, A., Lepetit, V., Moreno-Noguer, F.: Geometry-aware network for non-rigid shape prediction from a single view. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
76. Rad, M., Lepetit, V.: BB8: a scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth. In: IEEE International Conference on Computer Vision (2017)
77. Rad, M., Oberweger, M., Lepetit, V.: Feature mapping for learning fast and accurate 3D pose inference from synthetic images. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
78. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: IEEE conference on computer vision and pattern recognition (2017)
79. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)
80. Rijpkema, H., Girard, M.: Computer animation of knowledge-based human grasping. In: ACM SIGGRAPH, pp. 339–348 (1991)
81. Romero, J., Tzionas, D., Black, M.J.: Embodied hands: modeling and capturing hands and bodies together. ACM Trans. Graphics **36**, 245:1 – 245:17 (2017)
82. Rublee, E., Rabaud, V., Konolidge, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision (2011)
83. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)
84. Salzmann, M., Fua, P.: Deformable surface 3D reconstruction from monocular images. Morgan-Claypool Publishers, San Rafael (2010)
85. Sharp, T., Keskin, C., Robertson, D.P., Taylor, J., Shotton, J., Kim, D., Rhemann, C., Leichter, I., Vinnikov, A., Wei, Y., Freedman, D., Kohli, P., Krupka, E., Fitzgibbon, A.W., Izadi, A.S.: Accurate, robust, and flexible real-time hand tracking. In: Association for Computing Machinery (2015)
86. Shin, D., Fowlkes, C.C., Hoiem, D.: Pixels, voxels, and views: a study of shape representations for single view 3D object shape prediction. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
87. Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A., Blake, A.: Efficient human pose estimation from single depth images. IEEE Trans. Pattern Anal. Mach. Intell. **35**(12), 2821–2840 (2013)
88. Simon, T., Joo, H., Matthews, I.A., Sheikh, A.Y.: Hand keypoint detection in single images using multiview bootstrapping. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
89. Sock, J., Garcia-Hernando, G., Armagan, A., Kim, T.-K.: Introducing pose consistency and warp-alignment for self-supervised 6D object pose estimation in color images. In: IEEE International Conference on 3D Vision (2020)
90. Sun, X., Wei, Y., Liang, S., Tang, X., Sun, J.: Cascaded hand pose regression. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
91. Sun, X., Shang, J., Liang, S., Wei, Y.: Compositional human pose regression. In: IEEE International Conference on Computer Vision (2017)
92. Sundermeyer, M., Marton, Z.-C., Durner, M., Triebel, R.: Augmented autoencoders: implicit 3D orientation learning for 6D object detection. Int. J. Comput. Vis. **128**(3) 714–729 (2019)
93. Sundermeyer, M., Durner, M., Puang, E.Y., Vaskevicius, N., Marton, Z.C., Arras, K.O., Triebel, R.: Multi-path learning for object pose estimation across domains. In: IEEE Conference on Computer Vision and Pattern Recognition (2020)
94. Supancic, J.S., Rogez, G., Yang, Y., Shotton, J., Ramanan, D.: Depth-based hand pose estimation: data, methods, and challenges. In: IEEE International Conference on Computer Vision (2015)
95. Talvas, A., Marchal, M., Duriez, C., Otaduy, M.A.: Aggregate constraints for virtual manipulation with soft fingers. IEEE Trans. Visual Comput. Graphics **21**(4), 452–461 (2015)
96. Tang, D., Chang, H.J., Tejani, A., Kim, T.-K.: Latent regression forest: structured estimation of 3D articulated hand posture. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)
97. Taylor, J., Shotton, J., Sharp, T., Fitzgibbon, A.: The vitruvian manifold: inferring dense correspondences for one-shot human pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
98. Tekin, B., Sinha, S.N., Fua, P.: Real-time seamless single shot 6D object pose prediction. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
99. Tekin, B., Bogo, F., Pollefeys, M.: H+O: unified egocentric recognition of 3D hand-object poses and interactions. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)
100. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P.: Domain randomization for transferring deep neural networks from simulation to the real world. In: IEEE International Conference on Intelligent Robots and Systems (2017)
101. Tolani, D., Goswami, A., Badler, N.I.: Real-time inverse kinematics techniques for anthropomorphic limbs. Comput. Anim. **62**(5), 353–388 (2000)
102. Tompson, J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: Advances in Neural Information Processing Systems (2014)
103. Toshev, A., Szegedy, C.: DeepPose: human pose estimation via deep neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)
104. Tulsiani, S., Malik, J.: Viewpoints and keypoints. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
105. Tulsiani, S., Carreira, J., Malik, J.: Pose induction for novel object categories. In: IEEE International Conference on Computer Vision (2015)
106. Tzionas, D., Ballan, L., Srikantha, A., Aponte, P., Pollefeys, M., Gall, J.: Capturing hands in action using discriminative salient

points and physics simulation. Int. J. Comput. Vis. **118**(2), 172–193 (2016)
107. Vacchetti, L., Lepetit, V., Fua, P.: Stable real-time 3D tracking using online and offline information IEEE Trans. Pattern Anal. Mach. Intell. **26**(10), 1385–1391 (2004)
108. Wan, C., Probst, T., Van Gool, L., Yao, A.: Self-supervised 3D hand pose estimation through training by fitting. In: IEEE Conference on Computer Vision and Pattern Recognition (2019)
109. Wang, R., Paris, S., Popović, J.: 6D Hands: markerless hand-tracking for computer aided design. In: Association for Computing Machinery, pp. 549–558 (2011)
110. Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: IEEE Conference on Computer Vision and Pattern Recognition (2016)
111. Xiang, Y., Kim, W., Chen, W., Ji, J., Choy, C., Su, H., Mottaghi, R., Guibas, L., Savarese, S.: ObjectNet3D: a large scale database for 3D object recognition. In: European Conference on Computer Vision. Springer, Berlin (2016)
112. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes. In: Robotics: Science and Systems Conference (2018)
113. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes. In: Robotics: Science and Systems Conference (2018)
114. Xiao, Y., Marlet, R.: Few-shot object detection and viewpoint estimation for objects in the wild. In: European Conference on Computer Vision. Springer, Berlin (2020)
115. Xiong, F., Zhanga, B., Xiao, Y., Cao, Z., Yu, T., Zhou, J.T., Yuan, J.: A2J: anchor-to-joint regression network for 3D articulated pose estimation from a single depth image. In: IEEE International Conference on Computer Vision (2019)
116. Yuan, S., Ye, Q., Stenger, B., Jain, S., Kim, T.-K.: BigHand2.2M benchmark: hand pose dataset and state of the art analysis. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
117. Zakharov, S., Shugurov, I., Ilic, S.: DPOD: 6D pose object detector and refiner. In: IEEE International Conference on Computer Vision (2019)
118. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision (2017)
119. Zimmermann, C., Brox, T.: Learning to estimate 3D hand pose from single RGB images. In: IEEE International Conference on Computer Vision (2017)
120. Zimmermann, C., Ceylan, D., Yang, J., Russell, B.C., Argus, M.J., Brox, T.: FreiHAND: a dataset for markerless capture of hand pose and shape from single RGB images. In: IEEE International Conference on Computer Vision (2019)

Vincent Lepetit is a Director of Research at ENPC ParisTech. Before that, he was a full professor at the Institute for Computer Graphics and Vision (ICG) at TU Graz, Austria, where he still leads a research group as an associate professor. He is an editor for the main journals (PAMI, IJCV, CVIU) and a frequent chair for the main conferences on computer vision (CVPR, ICCV, ECCV). He has been working on 3D pose estimation from images from almost two decades, always with a focus on Augmented Reality.

Mixed Reality Interaction Techniques

5

Jens Grubert

Contents

5.1 Introduction 109
5.2 Tangible and Surface-Based Interaction 110
5.3 Gesture-Based Interaction 112
5.4 Pen-Based Interaction 113
5.5 Gaze-Based Interaction 114
5.6 Haptic Interaction 114
5.7 Multimodal Interaction 115
5.8 Multi-Display Interaction 116
5.9 Interaction Using Keyboard and Mouse 117
5.10 Virtual Agents 118
5.11 Summary and Outlook 118
References 119

Abstract

This chapter gives an overview of the interaction techniques for mixed reality with its variations of augmented and virtual reality (AR/VR). Various modalities for input and output are discussed. Specifically, techniques for tangible and surface-based interaction, gesture-based, pen-based, gaze-based, keyboard and mouse-based, as well as haptic interaction are discussed. Furthermore, the combinations of multiple modalities in multisensory and multimodal interaction as well as interaction using multiple physical or virtual displays are presented. Finally, interactions with intelligent virtual agents are considered.

Keywords

Tangible interaction · Augmented surfaces · Gestures · Magic lens · Eye gaze · Pen-based interaction · Keyboard and mouse · Haptics · Multimodal · Multi-display interaction · Intelligent virtual agents

5.1 Introduction

This chapter gives an overview of interaction techniques for mixed reality (MR) with its variations of augmented and virtual reality (AR/VR). Early research in the field of MR interaction techniques focused on the use of surface-based, tangible, and gesture-based interaction, which will be presented at the beginning of this chapter. Further modalities, such as pen-based, gaze-based, or haptic interaction, have recently gained attention and are presented next. Further, with the move toward productivity-oriented use cases, interaction with established input devices such as keyboard and mouse has gained interest from the research community. Finally, inspired by the popularity of conversational agents, interaction with intelligent virtual agents is discussed.

The development of interaction techniques is closely related to the advancements in input devices. Hence, the reader is invited to study the according book chapter as well.

While this chapter follows the abovementioned structure, further possibilities of structure interaction techniques include organization according to interaction tasks [1] such as object selection [2–4] and object manipulation [5], navigation [6], symbolic input [7], or system control [8]. Further, interaction techniques for specific application domains have been proposed, such as music [9] games [10] or immersive analytics [11].

Interested readers are also referred to further surveys and books in areas such as 3D interaction techniques, [12] or interaction with smart glasses [13].

J. Grubert (✉)
Coburg University of Applied Sciences, Department of Electrical Engineering and Computer Science, Coburg, Germany
e-mail: jens.grubert@hs-coburg.de

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_5

5.2 Tangible and Surface-Based Interaction

This section presents the concepts of tangible user interfaces (TUIs) and their applicability in AR. It covers the effects of output media, spatial registration approaches for TUIs, tangible magic lenses, augmenting large surfaces like walls and whole rooms, the combination of AR with shape-changing displays, and the role of TUIs for VR-based interaction. Figure 5.1 depicts an overview about output and input devices typically found in TUI-based interaction for MR.

TUIs are concerned with the use of physical objects as medium for interaction with computers [14] and has gained substantial interest in human–computer interaction [15]. Early prototypes utilized tabletop setting on which physical objects were placed to change properties of digital media. For example, Underkoffler and Ishii introduced a simulation of an optical workbench using tangible objects on a tabletop [16] as well as an application for architectural planning [17].

In AR, this concept was introduced by Kato et al. [18] as tangible augmented reality (TAR). They used a paddle as prop, equipped with a fiducial, to place furniture inside a house model. Fjeld et al. [19] introduced further tangibles such as a booklet and a cube for interaction within an educational application for chemistry.

TAR AR is used for visualizing digital information on physical objects while using those physical objects as interaction devices. Billinghurst et al. [20] stated that the TAR characteristics have a spatial registration between virtual and physical objects and the ability of users to interact with those virtual objects by manipulating the physical ones. Regenbrecht et al. [21] utilized a rotary plate to allow multiple co-located users to manipulate the orientation of a shared virtual object.

This way, the gap between digital output (e.g., on a flat screen) and physical input (e.g., using a rotary knob) can be reduced as the digital information is directly overlaid over the physical content.

Lee et al. [22] described the common interaction themes in the TAR application such as static and dynamic mappings between physical and digital objects. They describe a space-multiplexed approach, where each physical tool is mapped to a single virtual tool or function as well as a time-multiplexed approach in which the physical object is mapped to different digital tools dependent on the context of use.

However, the effect of this overlay is also dependent on the output medium used. For example, when using projection-based systems [23] or video see-through (VST) head-mounted displays (HMDs) (c.f. chapter 10 in [24]), the distance between the observer and the physical and virtual objects is the same. In contrast, when using commodity optical see-through (OST) HMDs with a fixed focal plane, there can be an substantial cost of perceiving virtual and physical objects at the same time. Specifically, Eiberger et al. [25] demonstrated that when processing visual information jointly from objects within arms' reach (in this case, a handheld display) and information presented on a OST HMD at a different distance, the task completion times increases by approximately 50%, and the error rate increased by approximately 100% compared with processing this visual information solely on the OST HMD.

For spatially registering physical and virtual objects, early works on TAR often relied on fiducial markers, such as that provided by ARToolKit [26] or ARUCO [27]. While easy to prototype (i.e., simply, fiducials have to be printed out and attached to objects), these markers can inhibit interaction due to their susceptibility to occlusions (typically through hand and finger interaction). Hence, it is advised to use modern approaches for hand-based interaction [28, 29] with spatially tracked rigid and non-rigid objects [30–32].

A specific kind of TAR can be seen in *tangible magic lenses*, which evolved through a combination from the magic lens [33] and tangible interaction concepts [14]. Tangible magic lenses allow for access to and manipulation of otherwise hidden data in interactive spatial environments.

Evolving from the magic lens [33] and tangible interaction concepts [14], tangible magic lenses allow for access to and manipulation of otherwise hidden data in interactive spatial environments. A wide variety of interaction concepts for interactive magic lenses have been proposed within the scope of information visualization (see surveys [34, 35]).

Within AR, various rigid shapes have been explored. Examples include rectangular lenses for tabletop interaction [36] or circular lenses [37]. Flexible shapes (e.g., [38]) have been utilized as well as multiple sheets of paper [39]. In their pioneering work, Szalavári and Gervautz [40] introduced the personal-interaction panel in AR. The two-handed and pen-operated tablet allowed for the selection and manipulation of virtual object as well as for system control. Additionally, transparent props have been explored (e.g., a piece of plexiglass) both for tabletop AR [41–43] and VR [44]. Purely virtual tangible lenses have been proposed as well [45]. Brown et al. [46] introduced a cubic shape which could either perspectively correct, render, and manipulate 3D objects or text. This idea was later revisited by Issartel et al. [47] in a mobile setting.

Often, projection-based AR has been used to realize tangible magic lenses, in which a top-mounted projector illuminates a prop such as a piece of cardboard or other reflective materials [36, 48] and (typically RGB or depth) cameras process user input.

Mobile devices such as smartphones and tablets are also commonly used as a tangible magic lens [49, 50], and can be used in conjunction with posters [49], books [51], digital screens [50], or maps [52, 53].

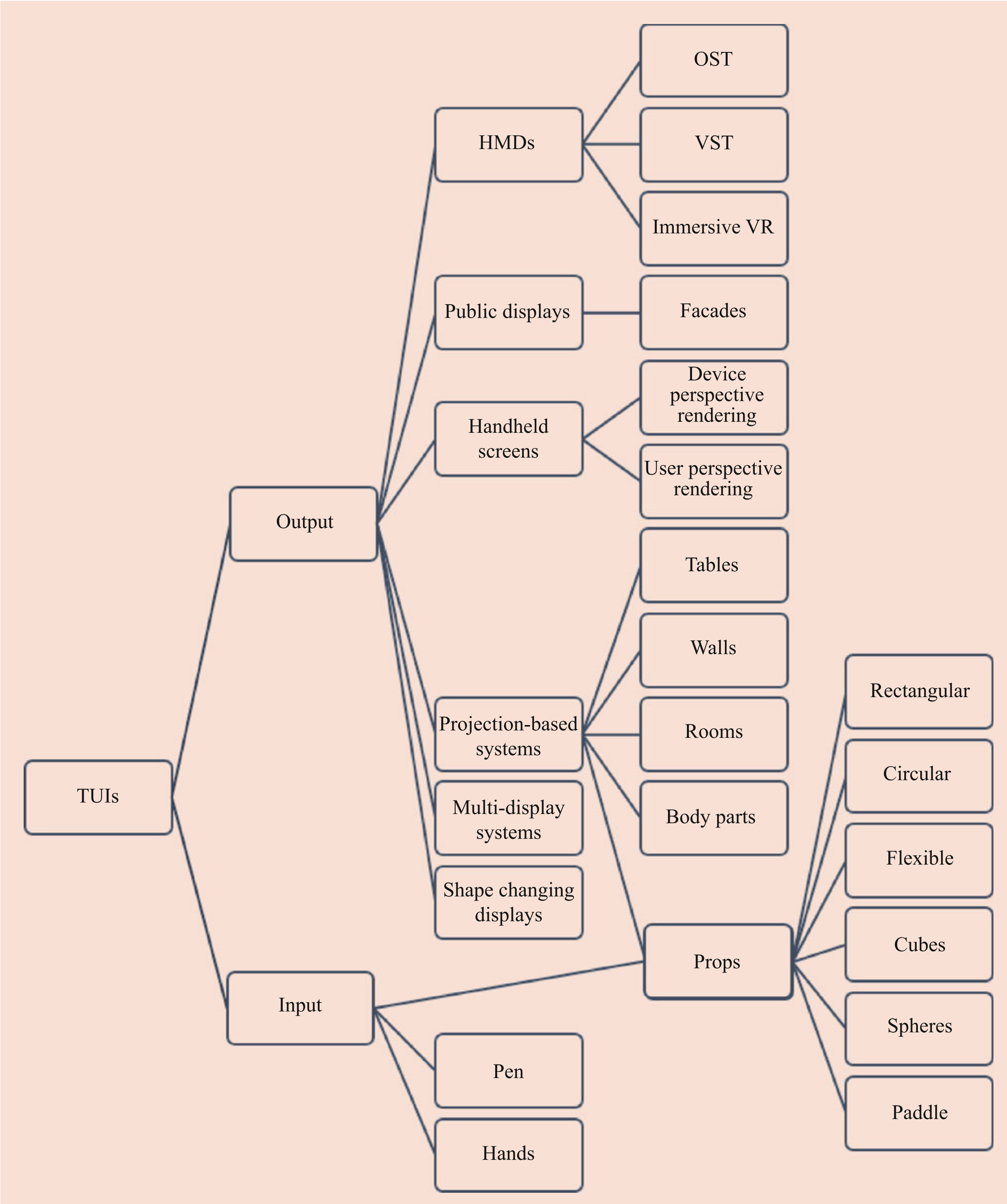


Fig. 5.1 A classification of input and output devices used in tangible user interfaces for AR and VR. OST: Optical See-Through. VST: Video See-Through. HMDs: Head-Mounted Displays

When using the tangible magic lens metaphor in public space, one should be aware about the social acceptability, specifically due to the visibility of spatial gestures and postures [54, 55]. For example, in a series of studies across gaming and touristic use cases, Grubert et al. [56, 57] explored benefits and drawbacks of smartphone-based tangible lens interfaces in public settings and compared them with traditional static peephole interaction, commonly used in mobile map applications. They found that user acceptance is largely dependent on the social and physical setting. In a public bus stop in a large open space used at a transit area in a public transportation stop, participants favored the magic lens over a static peephole interface despite tracking errors, fatigue, and potentially conspicuous gestures. Also, most passersby did not pay attention to the participants and vice versa. However, when deploying the same experience in a different public transportation stop with other spatial and social contexts (waiting area, less space to avoid physical proximity to others), participants used and preferred the magic lens interface significantly less compared with a static peephole interface.

Further, when using smartphones or tablets as magic lenses, the default user's view is based on the position of the physical camera attached to the handheld device. However, this can potentially negatively affect the user's experience [58, 59]. Hence, it can be advisable to incorporate user-perspective rendering to render the scene from the point of view of the user's head. In this domain, Hill et al. [60] introduced user-perspective rendering as *virtual transparency* for VST AR. Baričević et al. [61] compared user- vs. device-perspective rendering in a VR simulation. Tomioka et al. [62] presented approximated user-perspective rendering using homographies. Grubert et al. [63] proposed a framework for enabling user-perspective rendering to augment public displays. Čopič et al. [58, 59], quantified the performance differences between device- and user-perspective rendering in map-related tasks. Mohr et al. [64] developed techniques for an efficient computation of head-tracking techniques needed for user-perspective rendering.

Beyond handheld solutions, whole surfaces such as tables, walls, or body parts can be augmented and interacted with. Often projector-camera systems are used for processing input and creating output on surfaces. Early works included augmenting desks using projectors to support office work of single users [65–67] or in collaborative settings [68]. Later the Microsoft Kinect and further commodity depth sensors gave rise to a series of explorations with projector-camera systems.

For example, Xiao et al. [69] introduced WorldKit to allow users to sketch and operate user interface elements on everyday surfaces. Corsten et al. [70] proposed a pipeline for repurposing everyday objects as input devices. Henderson and Feiner also proposed the use of passive haptic feedback from everyday objects to interact with virtual control elements such as virtual buttons [71].

Mistry and Maes [72] utilized a necklace-mounted projector-camera system to sense finger interactions and project content on hands or the environment. Following suite, Harrison et al. [73] introduced OmniTouch, a wearable projector-depth-camera system that allowed for project user interface elements on body parts, such as the hand (e.g., a virtual dial pad), or for augmenting paper using touch.

Further, the idea of interacting with augmented surfaces was later expanded to cover bend surfaces [74], walls [75], and complete living rooms [76] or even urban facades [77, 78]. For example, in IllumiRoom [75], the area around a television was augmented using a projector, after initially scanning it with a depth camera. Possible augmentations included extending the field of view of on-screen content, selectively rendering scene elements of a game, or changing the appearance of the whole environment using non-photorealistic renderings (e.g., cartoon style or a wobble effect). In RoomAlive, multiple projector-depth camera systems were used to create a 3D scan of a living room as well as to spatially track the user's movement within that room. Users are able to interact with digital elements projected in the room using touch and in-air gestures. Apart from entertainment purposes, this idea was also investigated in productivity scenarios such as collaborative content sharing in meetings [79]. Finally, the augmentation of shape-changing interfaces was also explored [80–82]. For example, in Sublimate [82] an actuated pin display was combined with stereoscopic see-through screen to achieve a close coupling between physical and virtual object properties, e.g., for height fields or NURBS surface modeling. InForm [81] expanded this idea to allow both for user input on its pins (e.g., utilizing them as buttons or handles) as well as manipulation of external objects (such as moving a ball across its surface).

In VR, tangible interaction has been explored using various props. The benefit of using tangibles in VR is that a single physical object can be used to represent multiple virtual objects [83], even if they show a certain extend of discrepancy. Simeone et al. [84] presented a model of potential substitutions based on physical objects, such as mugs, bottles, umbrellas, and a torch. Hettiarachchi et al. [85] transferred this idea to AR. Harley et al. [86] proposed a system for authoring narrative experiences in VR using tangible objects.

5.3 Gesture-Based Interaction

Touch and in-air Gestures and postures make up a large part of interpersonal communication and have also been explored in depth in mixed reality. A driver for gesture-based interaction was the desire for "natural" user interaction, i.e., interaction without the need to explicitly handle artificial control devices but to rely on easy-to-learn interaction with (to the user) invisible input devices. While many gesture sets have been explored by researchers or users [87], it can be debated how "natural" those gesture-based interfaces really are [88], e.g., due to the poor affordances of lectures.

Still, the prevalence of small sensors such as RGB and depth cameras, inertial measurement units, radars or magnetic sensors in mobile devices and AR as well as VR HMDs, and continuing advances in hand [28, 29], head [89], and body pose estimation [90–96] gave rise to a wide variety of gesture-based interaction techniques being explored for mixed reality.

For mobile devices research began investigating options for interaction next to [97], above [98, 99], behind [100, 101], across [102–105], or around [106, 107] the device.

The additional modalities are either substituting or complementing the devices' capabilities. These approaches typically relied on modifying existing devices using a variety of sensing techniques, which can limit their deployment to mass audiences. Hence, researchers started to investigate the use of unmodified devices. Nandakumar et al. [108] proposed the use of the internal microphones of mobiles to determine the location of finger movements on surfaces but cannot support mid-air interaction. Song et al. [109] enabled in-air gestures using the front and back facing cameras of unmodified mobile devices. With Surround See, Yang et al. [110] modified the front-facing camera of a mobile phone with an omnidirectional lens, extending its field of view to 360° horizontally. They showcased different application areas, including peripheral environment, object, and activity detection, including hand gestures and pointing, but did not comment on the recognition accuracy. In *GlassHands*, it was demonstrated how the input space around a device can be extended by using a built-in front-facing camera of an unmodified handheld device and some reflective glasses, like sunglasses, ski goggles, or visors [111]. This work was later extended to work with eye reflections [112, 113].

While being explored since the mid-1990s in tabletop-based AR [114–116], for handheld AR, vision-based finger and hand tracking became popular since the mid-2000s [117–120]. Yusof et al. [121] provide a survey on the various flavors of gesture-based interaction in handheld AR, including marker-based and marker-less tracking of fingers or whole hands.

An early example of in-air interaction using AR HMDs is presented by Kolsch et al. [122], who demonstrated finger tracking with a head-mounted camera. Xiao et al. [123] showed how to incorporate touch gestures on everyday surfaces in to the Microsoft HoloLens. Beyond hand and finger tracking, full-body tracking using head-mounted cameras was also explored [124]. Also, reconstruction of facial gestures, e.g., for reenactment purposes, when wearing HMDs has gained increased interest [125–128].

Further solutions for freehand interaction were also proposed, including a wrist-worn gloveless sensor [129], swept frequency capacitive sensing [130], an optical mouse sensor attached to a finger [131], or radar-based sensing [132].

Most AR and VR in-air interactions typically aim at using unsupported hands. Hence, to enable reliable selection, targets are designed to be sufficiently large and spaced apart [133]. Also, while the addition of hand tracking to modern AR and VR HMDs allows for easy access to in-air gestures, the accuracy of those spatial tracking solutions is still significantly lower than dedicated lab-based external tracking systems [134].

Besides interaction with handheld or head-worn devices, also whole environments such as rooms can be equipped with sensors to facilitate gesture-based interaction [135–137].

In VR, off-the-shelf controllers were also appropriated to reconstruct human poses in real time [138, 139].

5.4 Pen-Based Interaction

In-air interactions in AR and VR typically make use of unsupported hands or controllers designed for gaming. In addition, pens (often in combination with tablets as supporting surface) have also been explored as input devices. Szalavári and Gervautz [40] as well as Billinghurst et al. [140] utilized pens for input on physical tablets in AR respectively VR. Watsen et al. [141] used a handheld Personal Digital Assistant (PDA) for operating menus in VR. In the *Studierstube* frameworks, pens were used to control 2D user interface elements on a PDA in AR. Poupyrev et al. [142] used a pen for notetaking in VR. Gesslein et al. [143] used a pen for supporting spreadsheet interaction in Mobile VR.

Many researches also investigated the use of pens for drawing and modeling. Sachs et al. [144] used an early system of 3D CAD modeling using a pen. Deering [145] used a pen for in-air sketching in a fishtank VR environment. Keeve et al. [146] utilized a brush for expressive painting in a Cave Automatic Virtual Environment (CAVE). Encarnacao [147] used a pen and pad for sketching in VR on top of an interactive table. Fiorentino et al. [148] explored the use of pens in mid-air for CAD applications in VR. Xin et al. [149] enabled the creation of 3D sketches using the pen and tablet interaction in handheld AR. Yee et al. [150] used a pen-line device along a VST HMD for in situ sketching in AR. Gasquez et al. [151, 152], Arora et al. [153], and Drey et al. [154] noted the benefits of supporting both free-form in-air sketching and on a supporting 2D surface in AR and VR. Suzuki et al. [155] expanded previous sketching applications for AR with dynamic and responsive graphics, e.g., to support physical simulations.

The performance of pen-based input was also investigated in VR. Bowman and Wingrave [156] compared pen and tablet input for menu selection against floating menus and a pinch-based menu system and found that the pen and tablet interaction was significantly faster. Teather and Stuerzlinger [157] compared pen-based input with mouse input for target selection in a fishtank VR environment and found that 3D pointing was inferior to 2D pointing when targets where rendered stereoscopically. Arora et al. [158] compared pen-based mid-air painting with surface-supported painting and found supporting evidence that accuracy improved using a physical drawing surface. Pham et al. [159] indicated that pens significantly outperform controllers for input in AR and VR and is comparable to mouse-based input for target selection. Batmaz et al. explored different pen grip styles for target selection in VR [160].

5.5 Gaze-Based Interaction

Besides input using touch input gestures or handhold input devices, gaze has also been explored as input modality in mixed reality.

Duchowski [161] presents a review of 30 years of gaze-based interaction, in which gaze-based interaction is categorized within a taxonomy that splits interaction into four forms, namely, diagnostic (off-line measurement), active (selection, look to shoot), passive (foveated rendering, a.k.a. gaze-contingent displays), and expressive (gaze synthesis).

For VR, Mine [162] proposed the use of gaze-directed steering and look-at menus in 1995. Tanriverdi and Jacob [163] highlighted that VR can benefit from gaze tracking. They stated that physical effort can be minimized through gaze, and user's natural eye movement can be employed to perform interactions in VR (e.g., with distant objects). They also show that a proposed heuristic gaze selection technique outperforms virtual hand-based interaction in terms of task-completion time. Cournia et al. [164] found that dwell-time-based selection was slower than manual ray-pointing. Duchowski et al. [165] presented software techniques for binocular eye tracking within VR as well as their application to aircraft inspection training. Specifically, they presented means for integrating eye trackers into a VR framework, novel 3D calibration techniques, and techniques for eye-movement analysis in 3D space. In 2020, Burova et al. [166] also utilized eye-gaze analysis in industrial tasks. They used VR to develop AR solutions for maintenance tasks and collected gaze data to elicit comments from industry experts on the usefulness of the AR simulation. Zeleznik et al. [167] investigated gaze interaction for 3D pointing, movement, menu selection, and navigation (orbiting and flying) in VR. They introduced "Lazy" interactions that minimize hand movements, "Helping Hand" techniques in which gaze augments hand-based techniques, as well as "Hands Down" techniques in which the hand can operate a separate input device. Piumsomboon et al. [168] presented three novel eye-gaze-based interaction techniques for VR: *Duo-Reticles*, an eye-gaze selection technique based on eye-gaze and inertial reticles; *Radial Pursuit*, a smooth pursuit-based technique for cluttered object; and *Nod and Roll*, a head-gesture-based interaction based on the vestibulo-ocular reflex.

5.6 Haptic Interaction

Auditory and visual channels are widely addressed sensory channels in AR and VR systems. Still, human experiences can be enriched greatly through touch and physical motions. Haptic devices enable the interaction between humans and computers by rendering mechanical signals to stimulate human touch and kinesthetic channels. The field of haptics has a long standing tradition and incorporates expertise from various fields such as robotics, psychology, biology, and computer science. They also play a role in diverse application domains such as gaming [169], industry [170], education [171], and medicine [172–174]. Haptic interactions are based on cutaneous/tactile (i.e., skin-related) and kinesthetic/proprioceptive (i.e., related to the body pose) sensations. Various devices have been proposed for both sensory channels, varying in form factor, weight, mobility, comfort as well as the fidelity, duration, and intensity of haptic feedback. For recent surveys, we refer to [175, 176].

In VR, the use of haptic feedback has a long tradition [177]. A commonly used active haptic device for stationary VR environment with a limited movement range of the users' hands is the PHANToM, which is a grounded system (or manipulandum) offering a high fidelity but low portability. Hence, over time substantial research efforts have been made in creating mobile haptic devices for VR [176].

In AR, the challenge in using haptics is that the display typically occludes real objects the user might want to interact with. Also, in OST displays, the haptic device is still visible behind virtual objects rendered on the display. When using VST displays, the haptic device might be removed by inpainting [178].

Besides *active* haptic systems, researchers have also investigated the use of low-fidelity physical objects to augment virtual environments in *passive* haptics. An early example of this type of haptic feedback is presented by Insko [179], who showed that passive haptics can improve both sense of presence and spatial knowledge training transfer in a virtual environment.

A challenge when using passive haptic feedback, besides a mismatch in surface fidelity, is that the objects used for feedback are typically static. To mitigate this problem, two strategies can be employed. First, the objects themselves can be moved during interaction by mounting them on robotic platforms such as robots [180, 181] or by human operators [182,183]. Second, the movements of the user themselves can be redirected to a certain extent by decoupling the physical motion of a user from the perceived visual motion. This can be done with individual body parts such as hands [184, 185] or the whole body using redirected walking techniques [186, 187].

5.7 Multimodal Interaction

While, often, AR and VR systems offer single input channels along with audio-visual output, rich interaction opportunities arise when considering the combination of further input and output modalities. Complementing the strengths of multiple channels can lead to increased user experiences. While multimodal (or multisensory) output is typically concerned with increasing the immersion and sense of presence in a scene, multimodal input typically tries to increase the efficiency of user interaction with a AR or VR system. For overviews about multimodal interaction beyond AR and VR, we refer to [188, 189]. Nizam et al. also provided a recent overview about multimodal interaction for specifically for AR [190].

The use of multisensory output such as the combination of audiovisual output with smell and touch has been shown to increase presence and perceived realism in VR [191, 192] and has been employed as early as in the 1960s [193]. Gallace et al. discussed both benefits and challenges when utilizing multiple output modes in VR [194]. Extrasensory experiences [195, 196] (such as making temperature visible through infrared cameras) have also been explored [197].

In AR, Narumi et al. [198] showed that increasing the perceived size of a real cookie using AR also increased the feeling of satiety. Narumi et al. [199] also created a multisensory eating experience in AR by changing the apparent look and smell of cookies. Koizumi et al. [200] could modulate the perceived food texture using a bone-conducting speaker. Ban et al. [201] showed that it is possible to influence fatigue while handling physical objects by affecting their perceived weight by modulating their size in AR.

Regarding multimodal input in VR, the combination of speech and gestures is a commonly used input combination. In 1980, Bolt [202] introduced *put-that-there*. Users could immerse themselves in a *Media Room* to place objects within that environment through a combination of gestures and speech. In 1989, Hauptmann [203] showed that users preferred a combination of speech and gestures for the spatial manipulation of 3D object. Cohen et al. [204] used a handheld computer along with speech and gesture for supporting map-based tasks on a virtual workbench. LaViola [205] used hand-based interaction (sensed through a data glove) along with speech for interior design in VR. Ciger et al. [206] combined speech with pointing of a magic wand on an immersive wall to create "magical" experiences. Burdea et al. [207] presented an early survey on VR input and output devices as well as an overview about studies that quantify the potentials of several modalities on simulation realism and immersion. Prange et al. [208] studied the use of speech and pen-based interaction in a medical setting.

In AR, Olwal et al. [209] combined speech and gestures for object selection. Kaiser et al. [210] extended that work by introducing mutual disambiguation to improve selection robustness. Similarly, Heidemann et al. [211] presented an AR system for online acquisition of visual knowledge and retrieval of memorized objects using speech and deictic (pointing) gestures. Kolsch et al. [122] combined speech input with gestures in an outdoor AR environment. Piumsomboon [212] studied the use of gestures and speech vs gestures only for object manipulation in AR. They found that the multimodal was not substantially better than gesture-only-based interaction for most tasks (but object scaling). This indicates that multimodality per se is not always beneficial for interaction but needs to be carefully designed to suit the task at hand. Rosa et al. [213] discussed different notions of AR and Mixed Reality as well as the role of multimodality. Wilson et al. [214] used a projector-camera system mounted on a pan-tilt platform for multimodal interaction in a physical room using a combination of speech and gestures.

The combination of touch and 3D movements has also been explored in VR and AR. Tsang et al. [215] introduced the Boom Chameleon, touch display mounted on a tracked mechanical boom, and used joint gesture, speech, and viewpoint input in a 3D annotation application. Benko et al. [216] combined on-surface and in-air gestures for content transfer between a 2D screen and 3D space. Mossel et al. [217] and Marzo et al. [218] combined touch input and handheld device movement for 3D object manipulations in mobile AR. Polvi et al. [219] utilized touch and the pose of a handheld touchscreen for reminded object positioning in mobile AR. Grandi et al. [220] studied the use of touch and the orientation of a smartphone for collaborative object manipulation in VR. Surale et al. [221] explored the use of touch input on a spatially tracked tablet for object manipulations in VR. In VR, Menzner et al. [222] utilized combined in-air and touch movements on and above smartphones for efficient navigation of multiscale information spaces. Several authors combined pen input both in mid-air as well as on touch surfaces to enhance sketching in VR [154] and AR [151–153].

Also, the combination of eye-gaze with other modalities such as mid-air gestures and head movements has seen recent interest for interaction in AR and VR. For example, Pfeuffer et al. [223] investigated the combination of gaze and gestures in VR. They described *Gaze + Pinch*, which integrates eye gaze to select 3D objects, and indirect freehand gestures to manipulate those objects. They explored this technique for object selection, manipulation, scene navigation, menu interaction, and image zooming. Similarly, Ryu et al. [224] introduced a combined grasp eye-pointing technique for 3D object selection. Kyto et al. [225] combined head and eye gaze for improving target selection in AR. Sidenmark and Gellersen [226, 227] studied different techniques combining eye and head pointing in VR. Gesslein et al. [143] combined pen-based input with gaze tracking for efficient interaction across multiple spreadsheets. Biener et al. [228] utilized gaze and touch interaction to navigate virtual multi-display environments.

5.8 Multi-Display Interaction

Traditionally, output of interactive systems is often limited to a single display, ranging from smartwatches to gigapixel displays. However, multi-display environments from the desktop to gigapixel displays are also increasingly common for knowledge work and complex tasks such as financial trading or factory management as well as for social applications such as second screen TV experiences [229]. Surveys about multi-display systems and distributed user interfaces have been presented by Elmqvist [230], Grubert et al. [229, 231, 232], and Brudy et al. [233].

Augmented reality has the potential to enhance interaction with both small and large displays by adding an unlimited virtual screen space or other complementing characteristics like mobility. However, this typically comes at the cost of a lower-display fidelity compared with a physical panel display (such as lower resolution, lower contrast, or a smaller physical field of view in OST HMDs).

In 1991, Feiner et al. [234] proposed a hybrid display combining a traditional desktop monitor with an OST HMD and explored a window manager application. Butz et al. [235] combined multiple physical displays ranging from handheld to wall-sized ones with OST HMDs in a multi-user collaborative environment. Baudisch et al. [236] used a lower-resolution projector to facilitate focus and context interaction on a desktop computer. MacWilliams et al. [237] proposed a multi-user game in which players could interact with a tabletop, laptop, and handheld displays. Serrano et al. [238] proposed to use an OST HMD to facilitate content transfer between multiple physical displays on a desktop. Boring et al. [239] used a smartphone to facilitate content transfer between multiple stationary displays. They later extended the work to manipulate screen content on stationary displays [240] and interactive facades [241] using smartphones. Raedle et al. [104] supported interaction across multiple mobile displays through a top-mounted depth camera. Grubert et al. [105, 242] used face tracking to allow user interaction across multiple mobile devices, which could be dynamically repositioned. They also proposed to utilize face tracking [242, 243] to create a cubic VR display with user-perspective rendering. Butscher et al. [244] explored the combination of VST HMDs with a tabletop displays for information visualization. Reipschläger et al. [245, 246] combined a high-resolution horizontal desktop display with an OST HMD for design activities. Gugenheimer et al. [247] introduced face touch, which allows interacting with display-fixed user interfaces (using direct touch) and world-fixed content (using raycasting). This work was later extended to utilize three touch displays around the user's head [248]. Gugenheimer et al. also introduced *ShareVR* [249], which enabled multi-user and multi-display interactions across users inside and outside of VR.

A number of systems also concentrated on the combination of HMDs and handheld as well body-worn displays, such as smartwatches, smartphones, and tablets in mobile contexts. Here, typically the head-mounted display extends the field of view of the handheld display to provide a larger virtual field of view. In MultiFi [250], an OST HMD provides contextual information for higher-resolution touch-enabled displays (smartwatch and smartphone). The authors explored different spatial reference systems such as body-aligned, device-aligned, and side-by-side modes. Similar explorations have followed suit using video-see-through HMDs [251], an extended set of interaction techniques [252], using smartwatches [253–255], or with a focus on understanding smartphone-driven window management techniques for HMDs [256].

Purely virtual multi-display environments have also been explored in AR and VR. In 1993, Feiner et al. [257] introduced head-surrounding and world reference frames for positioning 3D windows in VR. In 1998, Billinghurst et al. [258] introduced the spatial display metaphor, in which information windows are arranged on a virtual cylinder around the user. Since then, virtual information displays have been explored in various reference systems, such as world-, object-, head-, body-, or device-referenced [259]. Specifically, interacting with windows in body-centered reference systems [260] has attracted attention, for instance, to allow fast access to virtual items [261, 262], mobile multi-tasking [263, 264], and visual analytics [265]. Lee et al. [266] investigated positioning a window in 3D space using a continuous hand gesture. Petford et al. [267] compared the selection performance of mouse and raycast pointing in full coverage displays (not in VR). Jetter et al. [268] proposed to interactively design a space with various display form factors in VR.

5.9 Interaction Using Keyboard and Mouse

Being the de facto standard for human-computer interaction in personal computing environments for decades, standard input peripherals such as keyboard and mouse, while initially used in projection-based CAVE environments, were soon replaced by special-purpose input devices and associated interaction techniques for AR and VR (see previous sections). This was partly due to the constraints of those input devices, making them challenging to use for spatial input with six degrees of freedom. Physical keyboards typically support solely symbolic input. Standard computer mice are restricted to two-dimensional pointing (along with button clicks and a scroll-wheel). However, with modern knowledge workers still relying on the efficiency of those physical input devices, researchers revisited how to use them within AR and VR.

With increasing interest in supporting knowledge work using AR and VR HMDs [269–272], keyboard and mouse interaction drew the attention of several researchers.

The keyboard was designed for the rapid entrance of symbolic information, and although it may not be the best mechanism developed for the task, its familiarity that enabled good performance by users without considerable learning efforts kept it almost unchanged for many years. However, when interacting with spatial data, they are perceived as falling short of providing efficient input capabilities [273], even though they are successfully used in many 3D environments (such as CAD or gaming [274]), can be modified to allow 3D interaction [275, 276], or can outperform 3D input devices in specific tasks such as 3D object placement [277, 278]. Also for 3D object manipulation in AR and VR, they were found to be not significantly slower than a dedicated 3D input device [279].

In VR, a number of works investigated the costs of using physical keyboards for standard text entry tasks. Grubert et al. [280, 281], Knierim et al. [282], and McGill et al. [283] found physical keyboards to be mostly usable for text entry in immersive head-mounted display-based VR but varied in their observations about the performance loss when transferring text entry from the physical to the virtual world. Pham et al. [284] deployed a physical keyboard on a tray to facilitate mobile text entry. Apart from standard QWERTY keyboards, a variety of further text entry input devices and techniques have been proposed for VR; see [7].

Besides using unmodified physical keyboards, there have been several approaches in extending the basic input capabilities of physical keyboard beyond individual button presses. Specifically, input on, above, and around the keyboard surface have been proposed using acoustic [285, 286], pressure [287–289], proximity [290], and capacitive sensors [291–296], cameras [297–299], body-worn orientation sensors [300], or even unmodified physical keyboards [301, 302]. Besides sensing, actuation of keys has also been explored [303]. Embedding capacitive sensing into keyboards has been studied by various researchers. It lends itself to detect finger events on and slightly above keys and can be integrated into mass-manufacturing processes. Rekimoto et al. [294] investigated capacitive sensing on a keypad, but not a full keyboard. Habib et al. [292] and Tung et al. [293] proposed to use capacitive sensing embedded into a full physical keyboard to allow touchpad operation on the keyboard surface. Tung et al. [293] developed a classifier to automatically distinguish between text entry and touchpad mode on the keyboard. Shi et al. developed microgestures on capacitive sensing keys [295, 304]. Similarly, Zheng et al. [305, 306] explored various interaction mappings for finger and hand postures. Sekimoro et al. focused on exploring gestural interactions on the space bar [307]. Extending the idea of LCD-programmable keyboards [308], Block et al. extended the output capabilities of touch-sensitive, capacitive-sensing keyboard by using a top-mounted projector [296]. Several commercial products have also augmented physical keyboards with additional, partly interactive, displays (e.g., Apple Touch Bar, Logitech G19 [309], Razer DeathStalker Ultimate [310]).

Maiti et al. [311] explored the use of randomized keyboard layouts on physical keyboards using an OST display. Wang et al. [312] explored the use of an augmented reality extension to a desktop-based analytics environment. Specifically, they added a stereoscopic data view using a HoloLens to a traditional 2D desktop environment and interacted with keyboard and mouse across both the HoloLens and the desktop.

Schneider et al. [313] explored a rich design space of using physical keyboards in VR beyond text entry. Specifically, they proposed three different input mappings: 1 key to 1 action (standard mode of interaction using keyboards), multiple keys to a single action (e.g., mapping a large virtual button to several physical buttons), as well as mapping a physical key to a coordinate in a two-dimensional input space. Similarly, they proposed three different output mappings: augmenting individual keys (e.g., showing an emoji on a key), augmenting on and around the keyboard (e.g., adding user-interface elements on top of the keyboard such as virtual sliders), as well as transforming the keyboard geometry itself (e.g., only displaying single buttons or replacing the keyboard by other visuals). Those ideas were later also considered in the domain of immersive analytics [314].

Mouse-based pointing has been studied in depth outside of AR and VR for pointing on single monitors [315] as well as multi-display environments [316–318]. However, it has been found that stand 2D mouse devices do not adapt well to multi-display interaction [319], an issue which is also relevant for AR and VR. Consequently, standard mice have been modified in various ways to add degrees of freedom. For example, Villar et al. [320] explored multiple form factors for multi-touch-enabled mice. Other researchers have added additional mouse sensors to support yawing [321, 322], pressure sensors for discrete selection [323, 324] to allow for three instead of two degrees of freedom. Three-dimensional interaction was enabled using *Rockin'Mouse* [325] and the *VideoMouse* [326]. Both works added a dome below the device to facilitate 3D interaction. Steed and Slater [327] proposed to add a dome on top of the mouse rather than below. Further form factors have also been proposed to facilitate pointing-based interaction in 3D [328, 329]. Recently, researchers also worked on unifying efficient input both in 2D and 3D [276, 330].

Standard mice using a scroll wheel can also be efficiently used for 3D object selection when being combined with gaze-tracking in virtual multi-display environments [228]. For example, in the Windows Mixed Reality Toolkit [331], the x and y movements of the mouse can be mapped to the x and y movements on a proxy shape, such as a cylinder (or any object on that cylinder, like a window). The scroll wheel is used for changing the pointer depth (in discrete steps). The x and y movements can be limited to the current field of view of the user to allow for acceptable control to display ratios. The user gaze can then be used to change the view on different regions of the proxy shape.

5.10 Virtual Agents

Virtual agents can be considered as "intelligent" software programs performing tasks on behalf of users based on questions or commands. While it can be argued what "intelligent" really means in this context, a widely accepted characteristic of this "intelligence" is context-aware behavior [332, 333]. This allows an agent to interact with the user and environment through sensing and acting in an independent and dynamic way. The behavior is typically well defined and allows to trigger actions based on a set of conditions [334].

The rise of voice assistants (or conversational agents) [335], which interact with users through natural language, has brought media attention and a prevalence in various areas, such as home automation, in-car operation, automation of call centers, education, and training [336].

In AR and VR, virtual agents often use more than a single modality for input and output. Complementary to voice in an output, virtual agents in AR and VR can typically react to body gestures or postures or even facial expressions of the users. Due to their graphical representations, those agents are embodied in the virtual world. The level of embodiment of a virtual agent has been studied for decades [337, 338]. For example it has been shown that the effect of adding a face was larger than the effect of visual realism (both photo-realism and behavioral realism of the avatar). In VR, the level of visual realism of the virtual agent is typically matched to the visual realism of the environment. In contrast, in AR, there is often a noticeable difference between the agent representation and the physical scene, and those effects are still underexplored [339]. Hantono et al. reviewed the use of virtual agents in AR in educational settings. Norouzi et al. provided review of the convergence between AR and virtual agents [340].

Specifically for AR, Maes et al. [341] introduced a magic mirror AR system, in which humans could interact with a dog through both voice and gestures. Similarly, Cavazza et al. [342] allowed participants to interact with virtual agents in an interactive storytelling environment. MacIntyre et al. [343] used pre-recorded videos of physical actors to let users interact with them using OST HMDs. Anabuki et al. [344] highlighted that having virtual agents and users share the same physical environment is the most distinguishing aspect of virtual agents in AR. They introduced Welbo, an animated virtual agent, which is aware of its physical environment and can avoid standing in the user's way. Barakony et al. [345] presented "AR Puppet" as system that explored the context-aware animated agents within AR in investigated aspects like visualization, appearance, or behaviors. They investigated AR-specific aspects such as the ability of the agent to avoid physical obstacles or its ability to interact with physical objects. Based on this initial research, the authors explored various applications [346, 347]. Chekhlov et al. [348] presented a system based on Simultanteous Localization and Mapping (SLAM) [349], in which the virtual agent had to move in a physical environment. Blum et al. [350] introduced an outdoor AR game which included virtual agents. Kotranza et al. [351, 352] used a tangible physical representation of a human that could be touched, along with a virtual visual representation in a medical education context. They called this dual representation *mixed reality humans* and argued that affording touch between a human and a virtual agent enables interpersonal scenarios.

5.11 Summary and Outlook

This chapter served as an overview of a wide variety of interaction techniques MR, covering both device- and prop-based input such as tangible interaction and pen and keyboard input as well as utilizing human effector-based input such as spatial gestures, gaze, or speech.

The historical development of the presented techniques was closely coupled to the available sensing capabilities. For example, in order to recognize props such as paddles [18], they had to be large enough in order to let fiducials be recognized by low-resolution cameras. With the advancement of computer vision-based sensing, fiducials could become smaller, change their appearance to natural-looking images, or be omitted altogether (e.g., for hand and finger tracking). Further, the combination of more than one modality became possible by increasing computational capabilities of MR systems.

In the future, we expect an ongoing trend of both minimizing the size and price of sensors, as well as the ubiquitous availability of those sensors, in dedicated computing devices, in everyday objects [353], on [354] or even in the human body itself [355]. Hence, MR interaction techniques will play a central role on shaping the future of both pervasive computing [333] as well as augmenting humans with (potentially) superhuman capabilities (e.g., motor capabilities [356, 357], cognitive and perceptual capabilities [358]. Besides technological and interaction challenges along the way, the field of MR interaction will greatly benefit from including both social and ethical implications when designing future interfaces.

References

1. Bowman, D., Kruijff, E., LaViola Jr, J.J., Poupyrev, I.P.: 3D User Interfaces: Theory and Practice, CourseSmart eTextbook. Addison-Wesley, New York (2004)
2. Dang, N.-T.: A survey and classification of 3D pointing techniques. In: Proceedings of the 2007 IEEE International Conference on Research, Innovation and Vision for the Future, pp. 71–80 (2007)
3. Argelaguet, F., Andujar, C.: A survey of 3D object selection techniques for virtual environments. Comput. Graph. **37**(3), 121–136 (2013)
4. Singh, H., Singh, J.: Object acquisition and selection in human computer interaction systems: a review. Int. J. Intell. Syst. Appl. Eng. **7**(1), 19–29 (2019)
5. Mendes, D., Caputo, F.M., Giachetti, A., Ferreira, A., Jorge, J.: A survey on 3D virtual object manipulation: From the desktop to immersive virtual environments. In: Computer Graphics Forum, vol. 38, pp. 21–45 (2019)
6. Jankowski, J., Hachet, M.: A Survey of Interaction Techniques for Interactive 3D Environments (2013)
7. Dube, T.J., Arif, A.S.: Text entry in virtual reality: A comprehensive review of the literature. In: International Conference on Human-Computer Interaction, pp. 419–437 (2019)
8. Dachselt, R., Hübner, A.: Three-dimensional menus: A survey and taxonomy. Comput. Graph. **31**(1), 53–65 (2007)
9. Berthaut, F.: 3D interaction techniques for musical expression. J. New Music Res. **49**(1), 1–13 (2019)
10. Riecke, B.E., LaViola Jr, J.J., Kruijff, E.: 3D user interfaces for virtual reality and games: 3D selection, manipulation, and spatial navigation. In: ACM SIGGRAPH 2018 Courses, pp. 1–94 (2018)
11. Büschel, W., Chen, J., Dachselt, R., Drucker, S., Dwyer, T., Görg, C., Isenberg, T., Kerren, A., North, C., Stuerzlinger, W.: Interaction for immersive analytics. In: Immersive Analytics, pp. 95–138. Springer, Berlin (2018)
12. Hand, C.: A survey of 3D interaction techniques. In: Computer Graphics Forum, vol. 16, pp. 269–281 (1997)
13. Lee, L.-H., Hui, P.: Interaction methods for smart glasses: A survey. IEEE Access **6**, 28712–28732 (2018)
14. Ullmer, B., Ishii, H.: The metaDESK: models and prototypes for tangible user interfaces. In: Proceedings of Symposium on User Interface Software and Technology (UIST 97). ACM, New York (1997)
15. Shaer, O., Hornecker, E.: Tangible user interfaces: past, present, and future directions. Found. Trends Hum. Comput. Interact. **3**(1–2), 1–137 (2010)
16. Underkoffler, J., Ishii, H.: Illuminating light: an optical design tool with a luminous-tangible interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 542–549 (1998)
17. Underkoffler, J., Ishii, H.: URP: a luminous-tangible workbench for urban planning and design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 386–393 (1999)
18. Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., Tachibana, K.: Virtual object manipulation on a table-top AR environment. In: Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), pp. 111–119 (2000)
19. Fjeld, M., Voegtli, B.M.: Augmented chemistry: An interactive educational workbench. In: Proceedings. International Symposium on Mixed and Augmented Reality, pp. 259–321 (2002)
20. Billinghurst, M., Grasset, R., Looser, J.: Designing augmented reality interfaces. ACM Siggraph Comput. Graphics **39**(1), 17–22 (2005)
21. Regenbrecht, H.T., Wagner, M., Baratoff, G.: Magicmeeting: A collaborative tangible augmented reality system. Virtual Reality **6**(3), 151–166 (2002)
22. Lee, G.A., Kim, G.J., Billinghurst, M.: Interaction design for tangible augmented reality applications. In: Emerging Technologies of Augmented Reality: Interfaces and Design, pp. 261–282. IGI Global, New York (2007)
23. Bimber, O., Raskar, R.: Spatial Augmented Reality: Merging Real and Virtual Worlds. CRC Press, New York (2005)
24. Hainich, R.R., Bimber, O.: Displays: Fundamentals and Applications. CRC Press, New York (2016)
25. Eiberger, A., Kristensson, P.O., Mayr, S., Kranz, M., Grubert, J.: Effects of depth layer switching between an optical see-through head-mounted display and a body-proximate display. In: Symposium on Spatial User Interaction, pp. 1–9 (2019)
26. Kato, H.: Inside ARToolKit. In: Proceedings of the 1st IEEE International Workshop on Augmented Reality Toolkit (2007)
27. Salinas, R.M.: ArUco: A Minimal Library for Augmented Reality Applications Based on OpenCV (2012)
28. Barsoum, E.: Articulated Hand Pose Estimation Review. arXiv preprint arXiv:1604.06195 (2016)
29. Li, R., Liu, Z., Tan, J.: A survey on 3D hand pose estimation: Cameras, methods, and datasets. Pattern Recogn. **93**, 251–272 (2019)
30. Rahman, M.M., Tan, Y., Xue, J., Lu, K.: Recent advances in 3D object detection in the era of deep neural networks: A survey. IEEE Trans. Image Process. **29**, 2947–2962 (2019)
31. Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M.: Deep Learning for 3D Point Clouds: A Survey. arXiv preprint arXiv:1912.12033 (2019)
32. Wu, Y.-C., Chan, L., Lin, W.-C.: Tangible and visible 3D object reconstruction in augmented reality. In: Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 26–36 (2019)
33. Bier, E.A., Stone, M.C., Pier, K., Buxton, W., DeRose, T.D.: Toolglass and magic lenses: the see-through interface. In: Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, pp. 73–80 (1993)
34. Tominski, C., Gladisch, S., Kister, U., Dachselt, R., Schumann, H.: A survey on interactive lenses in visualization. In: EuroVis (STARs) (2014)

35. Tominski, C., Gladisch, S., Kister, U., Dachselt, R., Schumann, H.: Interactive lenses for visualization: An extended survey. In: Computer Graphics Forum, vol. 36, pp. 173–200 (2017)
36. Spindler, M., Dachselt, R.: PaperLens: advanced magic lens interaction above the tabletop. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, p. 7 (2009)
37. Spindler, M., Tominski, C., Schumann, H., Dachselt, R.: Tangible views for information visualization. In: ACM International Conference on Interactive Tabletops and Surfaces, pp. 157–166 (2010)
38. Steimle, J., Jordt, A., Maes, P.: Flexpad: highly flexible bending interactions for projected handheld displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 237–246 (2013)
39. Holman, D., Vertegaal, R., Altosaar, M., Troje, N., Johns, D.: Paper windows: interaction techniques for digital paper. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 591–599 (2005)
40. Szalavári, Z., Gervautz, M.: The personal interaction Panel–a two-handed interface for augmented reality. In: Computer Graphics Forum, vol. 16, pp. C335–C346 (1997)
41. Rekimoto, J., Ullmer, B., Oba, H.: DataTiles: a modular platform for mixed physical and graphical interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 269–276 (2001)
42. Brown, L.D., Hua, H.: Magic lenses for augmented virtual environments. IEEE Comput. Graphics Appl. **26**(4), 64–73 (2006)
43. Oh, J.-y., Hua, H.: User evaluations on form factors of tangible magic lenses. In: Proceedings of the 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 23–32 (2006)
44. Schmalstieg, D., Encarnação, L.M., Szalavári, Z.: Using transparent props for interaction with the virtual table. SI3D **99**, 147–153 (1999)
45. Looser, J., Grasset, R., Billinghurst, M.: A 3D flexible and tangible magic lens in augmented reality. In: Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 51–54 (2007)
46. Brown, L.D., Hua, H., Gao, C.: A widget framework for augmented interaction in SCAPE. In: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, pp. 1–10 (2003)
47. Issartel, P., Besançon, L., Isenberg, T., Ammi, M.: A tangible volume for portable 3d interaction. In: Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 215–220 (2016)
48. Chan, L.K., Lau, H.Y.: MagicPad: the projection based 3D user interface. Int. J. Interact. Des. Manuf. (IJIDeM) **6**(2), 75–81 (2012)
49. Grubert, J., Pahud, M., Grasset, R., Schmalstieg, D., Seichter, H.: The utility of magic lens interfaces on handheld devices for touristic map navigation. Pervasive Mob. Comput. **18**, 88–103 (2015)
50. Leigh, S.-w., Schoessler, P., Heibeck, F., Maes, P., Ishii, H.: THAW: tangible interaction with see-through augmentation for smartphones on computer screens. In: Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 89–96 (2015)
51. Kljun, M., Pucihar, K.Č., Alexander, J., Weerasinghe, M., Campos, C., Ducasse, J., Kopacin, B., Grubert, J., Coulton, P., Čelar, M.: Augmentation not duplication: considerations for the design of digitally-augmented comic books. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2019)
52. Reitmayr, G., Eade, E., Drummond, T.: Localisation and interaction for augmented maps. In: Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05), pp. 120–129 (2005)
53. Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., Näsänen, J., Juustila, A.: Like bees around the hive: a comparative study of a mobile augmented reality map. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1889–1898 (2009)
54. Reeves, S., Benford, S., O'Malley, C., Fraser, M.: Designing the spectator experience. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 741–750 (2005)
55. Rico, J., Brewster, S.: Usable gestures for mobile interfaces: evaluating social acceptability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 887–896 (2010)
56. Grubert, J., Morrison, A., Munz, H., Reitmayr, G.: Playing it real: magic lens and static peephole interfaces for games in a public space. In: Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 231–240 (2012)
57. Grubert, J., Schmalstieg, D.: Playing it real again: a repeated evaluation of magic lens and static peephole interfaces in public space. In: Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 99–102 (2013)
58. Čopič Pucihar, K., Coulton, P., Alexander, J.: Evaluating dual-view perceptual issues in handheld augmented reality: device vs. user perspective rendering. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction, pp. 381–388 (2013)
59. Čopič Pucihar, K., Coulton, P., Alexander, J.: The use of surrounding visual context in handheld AR: device vs. user perspective rendering. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 197–206 (2014)
60. Hill, A., Schiefer, J., Wilson, J., Davidson, B., Gandy, M., MacIntyre, B.: Virtual transparency: Introducing parallax view into video see-through AR. In: Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 239–240 (2011)
61. Baričević, D., Lee, C., Turk, M., Höllerer, T., Bowman, D.A.: A hand-held AR magic lens with user-perspective rendering. In: Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 197–206 (2012)
62. Tomioka, M., Ikeda, S., Sato, K.: Approximated user-perspective rendering in tablet-based augmented reality. In: Proceedings of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 21–28 (2013)
63. Grubert, J., Seichter, H., Schmalstieg, D.: Towards user perspective augmented reality for public displays. In: Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 339–340 (2014)
64. Mohr, P., Tatzgern, M., Grubert, J., Schmalstieg, D., Kalkofen, D.: Adaptive user perspective rendering for handheld augmented reality. In: Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI), pp. 176–181 (2017)
65. Wellner, P.: The DigitalDesk calculator: tangible manipulation on a desk top display. In: Proceedings of the 4th annual ACM symposium on User Interface Software and Technology, pp. 27–33 (1991)
66. Wellner, P.: Interacting with paper on the DigitalDesk. Commun. ACM **36**(7), 87–96 (1993)
67. Mackay, W.E., Fayard, A.-L.: Designing interactive paper: lessons from three augmented reality projects. In: Proceedings of the International Workshop on Augmented Reality: Placing Artificial Objects in Real Scenes: Placing Artificial Objects in Real Scenes, pp. 81–90 (1999)

68. Rekimoto, J., Saitoh, M.: Augmented surfaces: a spatially continuous work space for hybrid computing environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 378–385 (1999)
69. Xiao, R., Harrison, C., Hudson, S.E.: WorldKit: rapid and easy creation of ad-hoc interactive applications on everyday surfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 879–888 (2013)
70. Corsten, C., Avellino, I., Möllers, M., Borchers, J.: Instant user interfaces: repurposing everyday objects as input devices. In: Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, pp. 71–80 (2013)
71. Henderson, S., Feiner, S.: Opportunistic tangible user interfaces for augmented reality. IEEE Trans. Vis. Comput. Graph. **16**(1), 4–16 (2010)
72. Mistry, P., Maes, P.: SixthSense: a wearable gestural interface. In: ACM SIGGRAPH ASIA 2009 Art Gallery and Emerging Technologies: Adaptation, pp. 85–85 (2009)
73. Harrison, C., Benko, H., Wilson, A.D.: OmniTouch: wearable multitouch interaction everywhere. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, pp. 441–450 (2011)
74. Benko, H., Jota, R., Wilson, A.: MirageTable: freehand interaction on a projected augmented reality tabletop. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 199–208 (2012)
75. Jones, B.R., Benko, H., Ofek, E., Wilson, A.D.: IllumiRoom: peripheral projected illusions for interactive experiences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 869–878 (2013)
76. Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., Ofek, E., MacIntyre, B., Raghuvanshi, N., Shapira, L.: RoomAlive: magical experiences enabled by scalable, adaptive projector-camera units. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, pp. 637–644 (2014)
77. Boring, S., Gehring, S., Wiethoff, A., Blöckner, A.M., Schöning, J., Butz, A.: Multi-user interaction on media facades through live video on mobile devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2721–2724 (2011)
78. Fischer, P.T., Hornecker, E.: Urban HCI: spatial aspects in the design of shared encounters for media facades. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 307–316 (2012)
79. Fender, A.R., Benko, H., Wilson, A.: Meetalive: Room-scale omni-directional display system for multi-user content and control sharing. In: Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, pp. 106–115 (2017)
80. Rasmussen, M.K., Pedersen, E.W., Petersen, M.G., Hornbæk, K.: Shape-changing interfaces: a review of the design space and open research questions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 735–744 (2012)
81. Follmer, S., Leithinger, D., Olwal, A., Hogge, A., Ishii, H.: inFORM: dynamic physical affordances and constraints through shape and object actuation.. In: Uist, vol. 13, pp. 2501988–2502032 (2013)
82. Leithinger, D., Follmer, S., Olwal, A., Luescher, S., Hogge, A., Lee, J., Ishii, H.: Sublimate: state-changing virtual and physical rendering to augment interaction with shape displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1441–1450 (2013)
83. Araujo, B., Jota, R., Perumal, V., Yao, J.X., Singh, K., Wigdor, D.: Snake Charmer: physically enabling virtual objects. In: Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 218–226 (2016)
84. Simeone, A.L., Velloso, E., Gellersen, H.: Substitutional reality: Using the physical environment to design virtual reality experiences. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3307–3316 (2015)
85. Hettiarachchi, A., Wigdor, D.: Annexing reality: Enabling opportunistic use of everyday objects as tangible proxies in augmented reality. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 1957–1967 (2016)
86. Harley, D., Tarun, A.P., Germinario, D., Mazalek, A.: Tangible vr: Diegetic tangible objects for virtual reality narratives. In: Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 1253–1263 (2017)
87. Piumsomboon, T., Clark, A., Billinghurst, M., Cockburn, A.: User-defined gestures for augmented reality. In: IFIP Conference on Human-Computer Interaction, pp. 282–299 (2013)
88. Norman, D.A.: Natural user interfaces are not natural. Interactions **17**(3), 6–10 (2010)
89. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: A survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–626 (2008)
90. Chen, L., Wei, H., Ferryman, J.: A survey of human motion analysis using depth imagery. Pattern Recogn. Lett. **34**(15), 1995–2006 (2013)
91. Liu, Z., Zhu, J., Bu, J., Chen, C.: A survey of human pose estimation: the body parts parsing based methods. J. Vis. Commun. Image Represent. **32**, 10–19 (2015)
92. Gong, W., Zhang, X., Gonzàlez, J., Sobral, A., Bouwmans, T., Tu, C., Zahzah, E.-h.: Human pose estimation from monocular images: A comprehensive survey. Sensors **16**(12), 1966 (2016)
93. Sarafianos, N., Boteanu, B., Ionescu, B., Kakadiaris, I.A.: 3d human pose estimation: A review of the literature and analysis of covariates. Comput. Vis. Image Underst. **152**, 1–20 (2016)
94. Dang, Q., Yin, J., Wang, B., Zheng, W.: Deep learning based 2d human pose estimation: A survey. Tsinghua Sci. Technol. **24**(6), 663–676 (2019)
95. Caserman, P., Garcia-Agundez, A., Goebel, S.: A Survey of Full-Body Motion Reconstruction in Immersive Virtual Reality Applications. IEEE Trans. Vis. Comput. Graph. **26**(10), 3089–3108 (2019)
96. Chen, Y., Tian, Y., He, M.: Monocular human pose estimation: A survey of deep learning-based methods. Comput. Vis. Image Underst. **192**, 102897 (2020)
97. Oakley, I., Lee, D.: Interaction on the edge: offset sensing for small devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14), pp. 169–178. ACM, New York (2014). https://doi.org/10.1145/2556288.2557138
98. Kratz, S., Rohs, M.: Hoverflow: exploring around-device interaction with IR distance sensors. In: Proceedings of the MobileHCI '09, p. 42 (2009)
99. Freeman, E., Brewster, S., Lantz, V.: Towards usable and acceptable above-device interactions. In: Proceedings of the MobileHCI '14, pp. 459–464. ACM, New York (2014). https://doi.org/10.1145/2628363.2634215
100. De Luca, A., von Zezschwitz, E., Nguyen, N.D.H., Maurer, M.-E., Rubegni, E., Scipioni, M.P., Langheinrich, M.: Back-of-device authentication on smartphones. In: Proceedings of the CHI '13, pp. 2389–2398. ACM, New York (2013). https://doi.org/10.1145/2470654.2481330
101. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., Shen, C.: Lucid touch: a see-through mobile device. In: Proceedings of the UIST '07, pp. 269–278. ACM, New York (2007). https://doi.org/10.1145/1294211.1294259
102. Schmidt, D., Seifert, J., Rukzio, E., Gellersen, H.: A cross-device interaction style for mobiles and surfaces. In: Proceedings of the DIS '12 (2012)

103. Chen, X., Grossman, T., Wigdor, D.J., Fitzmaurice, G.: Duet: exploring joint interactions on a smart phone and a smart watch. In: Proceedings of the CHI '14, pp. 159–168 (2014)
104. Rädle, R., Jetter, H.-C., Marquardt, N., Reiterer, H., Rogers, Y.: Huddlelamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. In: Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, pp. 45–54 (2014)
105. Grubert, J., Kränz, M.: Towards ad hoc mobile multi-display environments on commodity mobile devices. In: Proceedings of the 2017 IEEE Virtual Reality (VR), pp. 461–462 (2017)
106. Zhao, C., Chen, K.-Y., Aumi, M.T.I., Patel, S., Reynolds, M.S.: SideSwipe: Detecting In-air Gestures Around Mobile Devices Using Actual GSM Signal. In: Proceedings of the UIST '14, pp. 527–534. ACM, New York (2014). https://doi.org/10.1145/2642918.2647380
107. Xiao, R., Lew, G., Marsanico, J., Hariharan, D., Hudson, S., Harrison, C.: Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In: Proceedings of the MobileHCI '14, pp. 67–76 (2014)
108. Nandakumar, R., Iyer, V., Tan, D., Gollakota, S.: Fingerio: Using active sonar for fine-grained finger tracking. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 1515–1525 (2016)
109. Song, J., Sörös, G., Pece, F., Fanello, S.R., Izadi, S., Keskin, C., Hilliges, O.: In-air gestures around unmodified mobile devices. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, pp. 319–329 (2014)
110. Yang, X.-D., Hasan, K., Bruce, N., Irani, P.: Surround-see: enabling peripheral vision on smartphones during active use. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, pp. 291–300 (2013)
111. Grubert, J., Ofek, E., Pahud, M., Kranz, M., Schmalstieg, D.: Glasshands: Interaction around unmodified mobile devices using sunglasses. In: Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, pp. 215–224 (2016)
112. Schneider, D., Grubert, J.: [POSTER] Feasibility of corneal imaging for handheld augmented reality. In: Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 44–45 (2017)
113. Schneider, D., Grubert, J.: Towards around-device interaction using corneal imaging. In: Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, pp. 287–293 (2017)
114. Crowley, J., Berard, F., Coutaz , J.: Finger tracking as an input device for augmented reality. In: International Workshop on Gesture and Face Recognition, pp. 195–200 (1995)
115. Brown, T., Thomas, R.C.: Finger tracking for the digital desk. In: Proceedings First Australasian User Interface Conference. AUIC 2000 (Cat. No. PR00515), pp. 11–16 (2000)
116. Dorfmuller-Ulhaas, K., Schmalstieg, D.: Finger tracking for interaction in augmented environments. In: Proceedings IEEE and ACM International Symposium on Augmented Reality, pp. 55–64 (2001)
117. Lee, T., Hollerer, T.: Handy AR: Markerless inspection of augmented reality objects using fingertip tracking. In: Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers, pp. 83–90 (2007)
118. Lee, M., Green, R., Billinghurst, M.: 3D natural hand interaction for AR applications. In: Proceedings of the 2008 23rd International Conference Image and Vision Computing New Zealand, pp. 1–6 (2008)
119. Shen, Y., Ong, S.-K., Nee, A.Y.: Vision-based hand interaction in augmented reality environment. Int. J. Hum. Comput. Interact. **27**(6), 523–544 (2011)
120. Hürst, W., Van Wezel, C.: Gesture-based interaction via finger tracking for mobile augmented reality. Multimed. Tools Appl. **62**(1), 233–258 (2013)
121. Yusof, C.S., Bai, H., Billinghurst, M., Sunar, M.S.: A review of 3D gesture interaction for handheld augmented reality. Jurnal Teknologi **78**(2-2) (2016). https://journals.utm.my/jurnalteknologi/article/view/6923
122. Kolsch, M., Turk, M., Hollerer, T.: Vision-based interfaces for mobility. In: The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004 (MOBIQUITOUS 2004), pp. 86–94 (2004)
123. Xiao, R., Schwarz, J., Throm, N., Wilson, A.D., Benko, H.: MRTouch: adding touch input to head-mounted mixed reality. IEEE Trans. Vis. Comput. Graph. **24**(4), 1653–1660 (2018)
124. Cha, Y.-W., Price, T., Wei, Z., Lu, X., Rewkowski, N., Chabra, R., Qin, Z., Kim, H., Su, Z., Liu, Y. : Towards fully mobile 3D face, body, and environment capture using only head-worn cameras. IEEE Trans. Vis. Comput. Graph. **24**(11), 2993–3004 (2018)
125. Chen, S.-Y., Gao, L., Lai, Y.-K., Rosin, P.L., Xia, S.: Real-time 3d face reconstruction and gaze tracking for virtual reality. In: Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 525–526 (2018)
126. Elgharib, M., BR, M., Tewari, A., Kim, H., Liu, W., Seidel, H.-P., Theobalt, C.: EgoFace: Egocentric Face Performance Capture and Videorealistic Reenactment. arXiv preprint arXiv:1905.10822 (2019)
127. Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M.: FaceVR: Real-time gaze-aware facial reenactment in virtual reality. ACM Trans. Graph. (TOG) **37**(2), 1–15 (2018)
128. Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., Stamminger, M., Nießner, M., Theobalt, C.: State of the art on monocular 3D face reconstruction, tracking, and applications. In: Computer Graphics Forum, vol. 37, pp. 523–550 (2018)
129. Kim, D., Hilliges, O., Izadi, S., Butler, A.D., Chen, J., Oikonomidis, I., Olivier, P.: Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, pp. 167–176 (2012)
130. Sato, M., Poupyrev, I., Harrison, C.: Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 483–492 (2012)
131. Yang, X.-D., Grossman, T., Wigdor, D., Fitzmaurice, G.: Magic finger: always-available input through finger instrumentation. In: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, pp. 147–156 (2012)
132. Wang, S., Song, J., Lien, J., Poupyrev, I., Hilliges, O.: Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 851–860 (2016)
133. Speicher, M., Feit, A.M., Ziegler, P., Krüger, A.: Selection-Based Text Entry in Virtual Reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), pp. 1–13. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3173574.3174221
134. Schneider, D., Otte, A., Kublin, A.S., Kristensson, P.O., Ofek, E., Pahud, M., Martschenko, A., Grubert, J.: Accuracy of commodity finger tracking systems for virtual reality head-mounted displays. In: IEEE VR 2020. IEEE, New York (2020)
135. Bränzel, A., Holz, C., Hoffmann, D., Schmidt, D., Knaust, M., Lühne, P., Meusel, R., Richter, S., Baudisch, P.: GravitySpace: tracking users and their poses in a smart room using a pressure-sensing floor. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 725–734 (2013)

136. Nabil, S., Plötz, T., Kirk, D.S.: Interactive architecture: Exploring and unwrapping the potentials of organic user interfaces. In: Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction, pp. 89–100 (2017)
137. Zhang, Y., Yang, C., Hudson, S.E., Harrison, C., Sample, A.: Wall++ Room-Scale Interactive and Context-Aware Sensing. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2018)
138. Jiang, F., Yang, X., Feng, L.: Real-time full-body motion reconstruction and recognition for off-the-shelf VR devices. In: Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1, pp. 309–318 (2016)
139. Caserman, P., Garcia-Agundez, A., Konrad, R., Göbel, S., Steinmetz, R.: Real-time body tracking in virtual reality using a Vive tracker. Virtual Reality **23**(2), 155–168 (2019)
140. Billinghurst, M., Baldis, S., Matheson, L., Philips, M.: 3D palette: a virtual reality content creation tool. In: Proceedings of the ACM symposium on Virtual reality software and technology, pp. 155–156 (1997)
141. Watsen, K., Darken, R., Capps, M.: A handheld computer as an interaction device to a virtual environment. In: Proceedings of the Third Immersive Projection Technology Workshop (1999)
142. Poupyrev, I., Tomokazu, N., Weghorst, S.: Virtual Notepad: handwriting in immersive VR. In: Proceedings of the IEEE 1998 Virtual Reality Annual International Symposium (Cat. No. 98CB36180), pp. 126–132 (1998)
143. Gesslein, T., Biener, V., Gagel, P., Schneider, D., Ofek, E., Pahud, M., Kristensson, P.O., Grubert, J.: Pen-based interaction with spreadsheets in mobile virtual reality. In: Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (2020)
144. Sachs, E., Roberts, A., Stoops, D.: 3-Draw: A tool for designing 3D shapes. IEEE Comput. Graphics Appl. **11**(6), 18–26 (1991)
145. Deering, M.F.: HoloSketch: a virtual reality sketching/animation tool. ACM Trans. Comput. Human Interact. (TOCHI) **2**(3), 220–238 (1995)
146. Keefe, D.F., Feliz, D.A., Moscovich, T., Laidlaw, D.H., LaViola Jr, J.J.: CavePainting: a fully immersive 3D artistic medium and interactive experience. In: Proceedings of the 2001 Symposium on Interactive 3D Graphics, pp. 85–93 (2001)
147. Encarnaĉão, L., Bimber, O., Schmalstieg, D., Chandler, S.: A translucent sketchpad for the virtual table exploring motion-based gesture recognition. In: Computer Graphics Forum, vol. 18, pp. 277–286 (1999)
148. Fiorentino, M., Uva, A.E., Monno, G.: The Senstylus: A Novel Rumble-Feedback Pen Device for CAD Application in Virtual Reality (2005)
149. Xin, M., Sharlin, E., Sousa, M.C.: Napkin sketch: handheld mixed reality 3D sketching. In: Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology, pp. 223–226 (2008)
150. Yee, B., Ning, Y., Lipson, H.: Augmented reality in-situ 3D sketching of physical objects. In: Intelligent UI Workshop on Sketch Recognition, vol. 1 (2009)
151. Gasques, D., Johnson, J.G., Sharkey, T., Weibel, N.: PintAR: sketching spatial experiences in augmented reality. In: Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion, pp. 17–20 (2019)
152. Gasques, D., Johnson, J.G., Sharkey, T., Weibel, N.: What you sketch is what you get: Quick and easy augmented reality prototyping with pintar. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–6 (2019)
153. Arora, R., Habib Kazi, R., Grossman, T., Fitzmaurice, G., Singh, K.: Symbiosissketch: Combining 2d & 3d sketching for designing detailed 3d objects in situ. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 185 (2018)
154. Drey, T., Gugenheimer, J., Karlbauer, J., Milo, M., Rukzio, E.: VRSketchIn: exploring the design space of pen and tablet interaction for 3D sketching in virtual reality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (2020)
155. Suzuki, R., Kazi, R.H., Wei, L.-Y., DiVerdi, S., Li, W., Leithinger, D.: RealitySketch: embedding responsive graphics and visualizations in AR with dynamic sketching. In: Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology, pp. 135–138 (2020)
156. Bowman, D.A., Wingrave, C.A.: Design and evaluation of menu systems for immersive virtual environments. In: Proceedings of the IEEE Virtual Reality 2001, pp. 149–156 (2001)
157. Teather, R.J., Stuerzlinger, W.: Pointing at 3D targets in a stereo head-tracked virtual environment. In: Proceedings of the 2011 IEEE Symposium on 3D User Interfaces (3DUI), pp. 87–94 (2011)
158. Arora, R., Kazi, R.H., Anderson, F., Grossman, T., Singh, K., Fitzmaurice, G.W.: Experimental evaluation of sketching on surfaces in VR.. In: CHI, vol. 17, pp. 5643–5654 (2017)
159. Pham, D.-M., Stuerzlinger, W.: Is the pen mightier than the controller? A comparison of input devices for selection in virtual and augmented reality. In: Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology, pp. 1–11 (2019)
160. Batmaz, A.U., Mutasim, A.K., Stuerzlinger, W.: Precision vs. Power grip: a comparison of pen grip styles for selection in virtual reality. In: Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, New York (2020)
161. Duchowski, A.T.: Gaze-based interaction: A 30 year retrospective. Comput. Graph. **73**, 59–69 (2018)
162. Mine, M.R.: Virtual Environment Interaction Techniques. UNC Chapel Hill CS Department, New York (1995)
163. Tanriverdi, V., Jacob, R.J.: Interacting with eye movements in virtual environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 265–272 (2000)
164. Cournia, N., Smith, J.D., Duchowski, A.T.: Gaze-vs. hand-based pointing in virtual environments. In: CHI'03 Extended Abstracts on Human Factors in Computing Systems, pp. 772–773 (2003)
165. Duchowski, A.T., Medlin, E., Gramopadhye, A., Melloy, B., Nair, S.: Binocular eye tracking in VR for visual inspection training. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, pp. 1–8 (2001)
166. Burova, A., Mäkelä, J., Hakulinen, J., Keskinen, T., Heinonen, H., Siltanen, S., Turunen, M.: Utilizing VR and Gaze tracking to develop AR solutions for industrial maintenance. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2020)
167. Zeleznik, R.C., Forsberg, A.S., Schulze, J.P.: Look-that-there: Exploiting gaze in virtual reality interactions. Technical Report CS-05 (2005)
168. Piumsomboon, T., Lee, G., Lindeman, R.W., Billinghurst, M.: Exploring natural eye-gaze-based interaction for immersive virtual reality. In: Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI), pp. 36–39 (2017)
169. Deng, S., Chang, J., Zhang, J.J.: A survey of haptics in serious gaming. In: International Conference on Games and Learning Alliance, pp. 130–144 (2013)
170. Xia, P., Lopes, A.M., Restivo, M.T.: A review of virtual reality and haptics for product assembly (part 1): rigid parts. In: Assembly Automation (2013)
171. Minogue, J., Jones, M.G.: Haptics in education: Exploring an untapped sensory modality. Rev. Educ. Res. **76**(3), 317–348 (2006)

172. Westebring-van der Putten, E.P., Goossens, R.H., Jakimowicz, J.J., Dankelman, J.: Haptics in minimally invasive surgery–a review. Minim. Invasive Ther. Allied Technol. **17**(1), 3–16 (2008)
173. Coles, T.R., Meglan, D., John, N.W.: The role of haptics in medical training simulators: a survey of the state of the art. IEEE Trans. Haptic **4**(1), 51–66 (2010)
174. Hamza-Lup, F.G., Bogdan, C.M., Popovici, D.M., Costea, O.D.: A survey of visuo-haptic simulation in surgical training. arXiv preprint arXiv:1903.03272 (2019)
175. Bermejo, C., Hui, P.: A survey on haptic technologies for mobile augmented reality. arXiv preprint arXiv:1709.00698 (2017)
176. Pacchierotti, C., Sinclair, S., Solazzi, M., Frisoli, A., Hayward, V., Prattichizzo, D.: Wearable haptic systems for the fingertip and the hand: taxonomy, review, and perspectives. IEEE Trans. Haptic **10**(4), 580–600 (2017)
177. McNeely, W.A.: Robotic graphics: a new approach to force feedback for virtual reality. In: Proceedings of the IEEE Virtual Reality Annual International Symposium, pp. 336–341 (1993)
178. Sandor, C., Uchiyama, S., Yamamoto, H.: Visuo-haptic systems: Half-mirrors considered harmful. In: Second Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07), pp. 292–297 (2007)
179. Insko, B.E., Meehan, M., Whitton, M., Brooks, F.: Passive Haptics Significantly Enhances Virtual Environments, Ph.D. Thesis (2001)
180. Suzuki, R., Hedayati, H., Zheng, C., Bohn, J.L., Szafir, D., Do, E.Y.-L., Gross, M.D., Leithinger, D.: RoomShift: room-scale dynamic haptics for VR with furniture-moving Swarm Robots. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–11 (2020)
181. Wang, Y., Chen, Z., Li, H., Cao, Z., Luo, H., Zhang, T., Ou, K., Raiti, J., Yu, C., Patel, S.: MoveVR: Enabling multiform force feedback in virtual reality using household cleaning robot. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2020)
182. Cheng, L.-P., Lühne, P., Lopes, P., Sterz, C., Baudisch, P.: Haptic turk: a motion platform based on people. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3463–3472 (2014)
183. Cheng, L.-P., Roumen, T., Rantzsch, H., Köhler, S., Schmidt, P., Kovacs, R., Jasper, J., Kemper, J., Baudisch, P.: Turkdeck: Physical virtual reality based on people. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, pp. 417–426 (2015)
184. Azmandian, M., Hancock, M., Benko, H., Ofek, E., Wilson, A.D.: Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In: Proceedings of the 2016 Chi Conference on Human Factors in Computing Systems, pp. 1968–1979 (2016)
185. Cheng, L.-P., Ofek, E., Holz, C., Benko, H., Wilson, A.D.: Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3718–3728 (2017)
186. Kohli, L., Burns, E., Miller, D., Fuchs, H.: Combining passive haptics with redirected walking. In: Proceedings of the 2005 International Conference on Augmented Tele-Existence, pp. 253–254 (2005)
187. Nilsson, N.C., Peck, T., Bruder, G., Hodgson, E., Serafin, S., Whitton, M., Steinicke, F., Rosenberg, E.S.: 15 years of research on redirected walking in immersive virtual environments. IEEE Comput. Graphics Appl. **38**(2), 44–56 (2018)
188. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: A survey. Comput. Vis. Image Underst. **108**(1–2), 116–134 (2007)
189. Turk, M.: Multimodal interaction: A review. Pattern Recogn. Lett. **36**, 189–195 (2014)
190. Nizam, S.S.M., Abidin, R.Z., Hashim, N.C., Lam, M.C., Arshad, H., Majid, N.A.A.: A review of multimodal interaction technique in augmented reality environment. Int. J. Adv. Sci. Eng. Inf. Technol **8**(4-2), 8–4 (2018)
191. Cater, J.P.: Smell/taste: odors in reality. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, vol. 2, pp. 1781–vol. IEEE, New York (1994)
192. Hoffman, H.G., Hollander, A., Schroder, K., Rousseau, S., Furness, T.: Physically touching and tasting virtual objects enhances the realism of virtual experiences. Virtual Reality **3**(4), 226–234 (1998)
193. Heilig, M.L.: Sensorama Simulator, vol. 28 (1962)
194. Gallace, A., Ngo, M.K., Sulaitis, J., Spence, C.: Multisensory presence in virtual reality: possibilities and limitations. In: Multiple Sensorial Media Advances and Applications: New Developments in MulSeMedia. IGI Global, New York (2012), pp. 1–38
195. Lemmon, V.W.: Extra-sensory perception. J. Psychol. **4**(1), 227–238 (1937)
196. Dublon, G., Paradiso, J.A.: Extra sensory perception. Sci. Am. **311**(1), 36–41 (2014)
197. Knierim, P., Kiss, F., Schmidt, A.: Look inside: understanding thermal flux through augmented reality. In: Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 170–171 (2018)
198. Narumi, T., Ban, Y., Kajinami, T., Tanikawa, T., Hirose, M.: Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 109–118 (2012)
199. Narumi, T., Nishizaka, S., Kajinami, T., Tanikawa, T., Hirose, M.: Augmented reality flavors: gustatory display based on edible marker and cross-modal interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 93–102 (2011)
200. Koizumi, N., Tanaka, H., Uema, Y., Inami, M.: Chewing jockey: augmented food texture by using sound based on the cross-modal effect. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, pp. 1–4 (2011)
201. Ban, Y., Narumi, T., Fujii, T., Sakurai, S., Imura, J., Tanikawa, T., Hirose, M.: Augmented endurance: controlling fatigue while handling objects by affecting weight perception using augmented reality. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 69–78 (2013)
202. Bolt, R.A.: "Put-that-there" Voice and gesture at the graphics interface. In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, pp. 262–270 (1980)
203. Hauptmann, A.G.: Speech and gestures for graphic image manipulation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 241–245 (1989)
204. Cohen, P., McGee, D., Oviatt, S., Wu, L., Clow, J., King, R., Julier, S., Rosenblum, L.: Multimodal interaction for 2D and 3D environments [virtual reality]. IEEE Comput. Graphics Appl. **19**(4), 10–13 (1999)
205. LaViola, J.: Whole-hand and speech input in virtual environments. In: Unpublished master's thesis, Department of Computer Science, Brown University, CS-99-15 (1999)
206. Ciger, J., Gutierrez, M., Vexo, F., Thalmann, D.: The magic wand. In: Proceedings of the 19th Spring Conference on Computer Graphics, pp. 119–124 (2003)
207. Burdea, G., Richard, P., Coiffet, P.: Multimodal virtual reality: Input-output devices, system integration, and human factors. Int. J. Hum. Comput. Interact. **8**(1), 5–24 (1996)
208. Prange, A., Barz, M., Sonntag, D.: Medical 3d images in multimodal virtual reality. In: Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion, pp. 1–2 (2018)

209. Olwal, A., Benko, H., Feiner, S.: Senseshapes: Using statistical geometry for object selection in a multimodal augmented reality. In: Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003, pp. 300–301 (2003)
210. Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., Feiner, S.: Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In: Proceedings of the 5th International Conference on Multimodal Interfaces, pp. 12–19 (2003)
211. Heidemann, G., Bax, I., Bekel, H.: Multimodal interaction in an augmented reality scenario. In: Proceedings of the 6th International Conference on Multimodal Interfaces, pp. 53–60 (2004)
212. Piumsomboon, T., Altimira, D., Kim, H., Clark, A., Lee, G., Billinghurst, M.: Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In: Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 73–82 (2014)
213. Rosa, N., Werkhoven, P., Hürst, W.: (Re-) examination of multimodal augmented reality. In: Proceedings of the 2016 Workshop on Multimodal Virtual and Augmented Reality, pp. 1–5 (2016)
214. Wilson, A., Benko, H., Izadi, S., Hilliges, O.: Steerable augmented reality with the beamatron. In: Proceedings of the 25th annual ACM symposium on User interface software and technology, pp. 413–422 (2012)
215. Tsang, M., Fitzmaurice, G.W., Kurtenbach, G., Khan, A., Buxton, B.: Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. In: Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology, pp. 111–120 (2002)
216. Benko, H., Ishak, E.W., Feiner, S.: Cross-dimensional gestural interaction techniques for hybrid immersive environments. In: IEEE Proceedings of Virtual Reality 2005 (VR 2005), pp. 209–216 (2005)
217. Mossel, A., Venditti, B., Kaufmann, H.: 3DTouch and HOMER-S: intuitive manipulation techniques for one-handed handheld augmented reality. In: Proceedings of the Virtual Reality International Conference: Laval Virtual, p. 12 (2013)
218. Marzo, A., Bossavit, B., Hachet, M.: Combining multi-touch input and device movement for 3D manipulations in mobile augmented reality environments. In: Proceedings of the 2nd ACM Symposium on Spatial User Interaction, pp. 13–16 (2014)
219. Polvi, J., Taketomi, T., Yamamoto, G., Dey, A., Sandor, C., Kato, H.: SlidAR: A 3D positioning method for SLAM-based handheld augmented reality. Comput. Graph. **55**, 33–43 (2016)
220. Grandi, J.G., Debarba, H.G., Nedel, L., Maciel, A.: Design and evaluation of a handheld-based 3D user interface for collaborative object manipulation. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5881–5891 (2017)
221. Surale, H.B., Gupta, A., Hancock, M., Vogel, D.: TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, p. 13 (2019)
222. Menzner, T., Gesslein, T., Otte, A., Grubert, J.: Above surface interaction for multiscale navigation in mobile virtual reality. In: Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 372–381 (2020)
223. Pfeuffer, K., Mayer, B., Mardanbegi, D., Gellersen, H.: Gaze+ pinch interaction in virtual reality. In: Proceedings of the 5th Symposium on Spatial User Interaction, pp. 99–108 (2017)
224. Ryu, K., Lee, J.-J., Park, J.-M.: GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection. J. Multimodal User Interfaces **13**(4), 1–11 (2019)
225. Kytö, M., Ens, B., Piumsomboon, T., Lee, G.A., Billinghurst, M.: Pinpointing: Precise head-and eye-based target selection for augmented reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 81 (2018)
226. Sidenmark, L., Gellersen, H.: Eye&Head: Synergetic eye and head movement for gaze pointing and selection. In: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, pp. 1161–1174 (2019)
227. Sidenmark, L., Mardanbegi, D., Ramirez Gomez, A., Clarke, C., Gellersen, H.: BimodalGaze: Seamlessly refined pointing with gaze and filtered gestural head movement. In: Proceedings of Eye Tracking Research and Applications (2020)
228. Biener, V., Schneider, D., Gesslein, T., Otte, A., Kuth, B., Kristensson, P.O., Ofek, E., Pahud, M., Grubert, J.: Breaking the screen: interaction across touchscreen boundaries in virtual reality for mobile knowledge workers. IEEE Trans. Vis. Comput. Graph.(01), 1–1 (2020). https://doi.org/10.1109/TVCG.2020.3023567
229. Grubert, J., Kranz, M., Quigley, A.: Challenges in mobile multidevice ecosystems. mUX: J. Mobile User Exp. **5**(1), 1–22 (2016)
230. Elmqvist, N.: Distributed user interfaces: State of the art. In: Distributed User Interfaces, pp. 1–12. Springer, New York (2011)
231. Grubert, J., Kranz, M., Quigley, A.: Design and technology challenges for body proximate display ecosystems. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 951–954 (2015)
232. Quigley, A., Grubert, J.: Perceptual and social challenges in body proximate display ecosystems. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 1168–1174 (2015)
233. Brudy, F., Holz, C., Rädle, R., Wu, C.-J., Houben, S., Klokmose, C.N., Marquardt, N.: Cross-device taxonomy: survey, opportunities and challenges of interactions spanning across multiple devices. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–28 (2019)
234. Feiner, S., Shamash, A.: Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers. In: Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology, pp. 9–17 (1991)
235. Butz, A., Hollerer, T., Feiner, S., MacIntyre, B., Beshers, C.: Enveloping users and computers in a collaborative 3D augmented reality. In: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99), pp. 35–44 (1999)
236. Baudisch, P., Good, N., Stewart, P.: Focus plus context screens: combining display technology with visualization techniques. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, pp. 31–40 (2001)
237. MacWilliams, A., Sandor, C., Wagner, M., Bauer, M., Klinker, G., Bruegge, B.: Herding sheep: live system development for distributed augmented reality. In: Proceedings of the 2nd IEEE ACM International Symposium on Mixed and Augmented Reality (ISMAR'03), p. 123 (2003)
238. Serrano, M., Ens, B., Yang, X.-D., Irani, P.: Gluey: Developing a head-worn display interface to unify the interaction experience in distributed display environments. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 161–171 (2015)
239. Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P.: Touch projector: mobile interaction through video. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2287–2296 (2010)
240. Baur, D., Boring, S., Feiner, S.: Virtual projection: exploring optical projection as a metaphor for multi-device interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1693–1702 (2012)
241. Boring, S., Baur, D.: Making public displays interactive everywhere. IEEE Comput. Graphics Appl. **33**(2), 28–36 (2012)

242. Grubert, J., Kranz, M.: Headphones: Ad hoc mobile multi-display environments through head tracking. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3966–3971 (2017)
243. Grubert, J., Kranz, M.: mpCubee: Towards a mobile perspective cubic display using mobile phones. In: Proceedings of the 2017 IEEE Virtual Reality (VR), pp. 459–460 (2017)
244. Butscher, S., Hubenschmid, S., Müller, J., Fuchs, J., Reiterer, H.: Clusters, trends, and outliers: How immersive technologies can facilitate the collaborative analysis of multidimensional data. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2018)
245. Reipschläger, P., Dachselt, R.: DesignAR: immersive 3D-modeling combining augmented reality with interactive displays. In: Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces, pp. 29–41 (2019)
246. Reipschläger, P., Engert, S., Dachselt, R.: Augmented Displays: seamlessly extending interactive surfaces with head-mounted augmented reality. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–4 (2020)
247. Gugenheimer, J., Dobbelstein, D., Winkler, C., Haas, G., Rukzio, E.: Facetouch: Enabling touch interaction in display fixed uis for mobile virtual reality. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 49–60 (2016)
248. Gugenheimer, J., Stemasov, E., Sareen, H., Rukzio, E.: FaceDisplay: enabling multi-user interaction for mobile virtual reality. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 369–372 (2017)
249. Gugenheimer, J., Stemasov, E., Frommel, J., Rukzio, E.: Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 4021–4033 (2017)
250. Grubert, J., Heinisch, M., Quigley, A., Schmalstieg, D.: Multifi: Multi fidelity interaction with displays on and around the body. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3933–3942 (2015)
251. Normand, E., McGuffin, M.J.: Enlarging a smartphone with AR to create a handheld VESAD (Virtually Extended Screen-Aligned Display). In: Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 123–133 (2018)
252. Zhu, F., Grossman, T.: BISHARE: Exploring bidirectional interactions between smartphones and head-mounted augmented reality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (2020)
253. Wenig, D., Schöning, J., Olwal, A., Oben, M., Malaka, R.: WatchThru: Expanding smartwatch displays with mid-air visuals and wrist-worn augmented reality. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 716–721 (2017)
254. Lu, Z., Hu, Y., Dai, J.: WatchAR: 6-DoF tracked watch for AR interaction. In: IEEE ISMAR 2019 Demonstrations (2019)
255. Wolf, D., Dudley, J.J., Kristensson, P.O.: Performance envelopes of in-air direct and smartwatch indirect control for head-mounted augmented reality. In: Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 347–354 (2018)
256. Ren, J., Weng, Y., Zhou, C., Yu, C., Shi, Y.: Understanding window management interactions in AR Headset+ smartphone interface. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts, pp. 1–8 (2020)
257. Feiner, S., MacIntyre, B., Haupt, M., Solomon, E.: Windows on the world: 2 D windows for 3 D augmented reality. In: ACM Symposium on User Interface Software and Technology, pp. 145–155 (1993)
258. Billinghurst, M., Bowskill, J., Jessop, M., Morphett, J.: A wearable spatial conferencing space. In: Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215), pp. 76–83 (1998)
259. LaViola Jr, J.J., Kruijff, E., McMahan, R.P., Bowman, D., Poupyrev, I.P.: 3D User Interfaces: Theory and Practice. Addison-Wesley Professional, New York (2017)
260. Wagner, J., Nancel, M., Gustafson, S.G., Huot, S., Mackay, W.E.: Body-centric design space for multi-surface interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1299–1308 (2013)
261. Li, F.C.Y., Dearman, D., Truong, K.N.: Virtual shelves: interactions with orientation aware devices. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 125–128 (2009)
262. Chen, X.A., Marquardt, N., Tang, A., Boring, S., Greenberg, S.: Extending a mobile devices interaction space through body-centric interaction. In: Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 151–160 (2012)
263. Ens, B.M., Finnegan, R., Irani, P.P.: The personal cockpit: a spatial interface for effective task switching on head-worn displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3171–3180 (2014)
264. Ens, B., Hincapié-Ramos, J.D., Irani, P.: Ethereal planes: a design framework for 2D information space in 3D mixed reality environments. In: Proceedings of the 2nd ACM Symposium on Spatial User Interaction, pp. 2–12 (2014)
265. Ens, B., Irani, P.: Spatial analytic interfaces: Spatial user interfaces for in situ visual analytics. IEEE Comput. Graphics Appl. **37**(2), 66–79 (2016)
266. Lee, J.H., An, S.-G., Kim, Y., Bae, S.-H.: Projective Windows: bringing windows in space to the fingertip. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 218 (2018)
267. Petford, J., Nacenta, M.A., Gutwin, C.: Pointing all around you: selection performance of mouse and ray-cast pointing in full-coverage displays. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 533 (2018)
268. Jetter, H.-C., Rädle, R., Feuchtner, T., Anthes, C., Friedl, J., Klokmose, C.N.: "In VR, everything is possible!": Sketching and simulating spatially-aware interactive spaces in virtual reality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–16 (2020)
269. Grubert, J., Ofek, E., Pahud, M., Kristensson, P.O., Steinicke, F., Sandor, C.: The office of the future: virtual, portable, and global. IEEE Comput. Graphics Appl. **38**(6), 125–133 (2018)
270. Ruvimova, A., Kim, J., Fritz, T., Hancock, M., Shepherd, D.C.: "Transport me away": Fostering flow in open offices through virtual reality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (2020)
271. Guo, J., Weng, D., Zhang, Z., Jiang, H., Liu, Y., Wang, Y., Duh, H.B.-L.: Mixed reality office system based on Maslow's hierarchy of needs: towards the long-term immersion in virtual environments. In: Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 224–235 (2019)
272. Li, Z., Annett, M., Hinckley, K., Singh, K., Wigdor, D.: HoloDoc: enabling mixed reality workspaces that harness physical and digital content. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, p. 687 (2019)
273. Besançon, L., Issartel, P., Ammi, M., Isenberg, T.: Mouse, tactile, and tangible input for 3D manipulation. In: Proceedings of the

2017 CHI Conference on Human Factors in Computing Systems, pp. 4727–4740 (2017)
274. Stuerzlinger, W., Wingrave, C.A.: The value of constraints for 3D user interfaces. In: Virtual Realities. Springer, Berlin (2011), pp. 203–223
275. Ware, C., Lowther, K.: Selection using a one-eyed cursor in a fish tank VR environment. ACM Trans. Comput. Human Interact. (TOCHI) **4**(4), 309–322 (1997)
276. Perelman, G., Serrano, M., Raynal, M., Picard, C., Derras, M., Dubois, E.: The roly-poly mouse: Designing a rolling input device unifying 2d and 3d interaction. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 327–336 (2015)
277. Bérard, F., Ip, J., Benovoy, M., El-Shimy, D., Blum, J.R., Cooperstock, J.R.: Did "Minority Report" get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task. In: IFIP Conference on Human-Computer Interaction, pp. 400–414 (2009)
278. Sun, J., Stuerzlinger, W., Riecke, B.E.: Comparing input methods and cursors for 3D positioning with head-mounted displays. In: Proceedings of the 15th ACM Symposium on Applied Perception, pp. 1–8 (2018)
279. Krichenbauer, M., Yamamoto, G., Taketom, T., Sandor, C., Kato, H.: Augmented reality versus virtual reality for 3d object manipulation. IEEE Trans. Vis. Comput. Graph. **24**(2), 1038–1048 (2017)
280. Grubert, J., Witzani, L., Ofek, E., Pahud, M., Kranz, M., Kristensson, P.O.: Text entry in immersive head-mounted display-based virtual reality using standard keyboards. In: Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 159–166 (2018)
281. Grubert, J., Witzani, L., Ofek, E., Pahud, M., Kranz, M., Kristensson, P.O.: Effects of hand representations for typing in virtual reality. In: Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 151–158 (2018)
282. Knierim, P., Schwind, V., Feit, A.M., Nieuwenhuizen, F., Henze, N.: Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–9 (2018)
283. McGill, M., Boland, D., Murray-Smith, R., Brewster, S.: A dose of reality: Overcoming usability challenges in vr head-mounted displays. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 2143–2152 (2015)
284. Pham, D.-M., Stuerzlinger, W.: HawKEY: Efficient and versatile text entry for virtual reality. In: Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology, pp. 1–11 (2019)
285. Kurosawa, T., Shizuki, B., Tanaka, J.: Keyboard Clawing: input method by clawing key tops. In: International Conference on Human-Computer Interaction, pp. 272–280 (2013)
286. Kato, J., Sakamoto, D., Igarashi, T.: Surfboard: keyboard with microphone as a low-cost interactive surface. In: Adjunct Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, pp. 387–388 (2010)
287. Dietz, P.H., Eidelson, B., Westhues, J., Bathiche, S.: A practical pressure sensitive computer keyboard. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 55–58 (2009)
288. Zagler, W.L., Beck, C., Seisenbacher, G.: FASTY-Faster and Easier Text Generation for Disabled People (2003)
289. Loy, C.C., Lai, W., Lim, C.: Development of a pressure-based typing biometrics user authentication system. In: ASEAN Virtual Instrumentation Applications Contest Submission (2005)
290. Taylor, S., Keskin, C., Hilliges, O., Izadi, S., Helmes, J.: Type-hover-swipe in 96 bytes: a motion sensing mechanical keyboard. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 1695–1704 (2014)
291. Fallot-Burghardt, W., Fjeld, M., Speirs, C., Ziegenspeck, S., Krueger, H., Läubli, T.: Touch&Type: a novel pointing device for notebook computers. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, pp. 465–468 (2006)
292. Habib, I., Berggren, N., Rehn, E., Josefsson, G., Kunz, A., Fjeld, M.: DGTS: Integrated typing and pointing. In: IFIP Conference on Human-Computer Interaction, pp. 232–235 (2009)
293. Tung, Y.-C., Cheng, T.Y., Yu, N.-H., Wang, C., Chen, M.Y.: FlickBoard: Enabling trackpad interaction with automatic mode switching on a capacitive-sensing keyboard. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1847–1850 (2015)
294. Rekimoto, J., Ishizawa, T., Schwesig, C., Oba, H.: PreSense: interaction techniques for finger sensing input devices. In: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, pp. 203–212 (2003)
295. Shi, Y., Zhang, H., Rajapakse, H., Perera, N.T., Vega Gálvez, T., Nanayakkara, S.: GestAKey: touch interaction on individual keycaps. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 596 (2018)
296. Block, F., Gellersen, H., Villar, N.: Touch-display keyboards: transforming keyboards into interactive surfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1145–1154 (2010)
297. Wilson, A.D.: Robust computer vision-based detection of pinching for one and two-handed gesture input. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, pp. 255–258 (2006)
298. Kim, D., Izadi, S., Dostal, J., Rhemann, C., Keskin, C., Zach, C., Shotton, J., Large, T., Bathiche, S., Nießner, M. : RetroDepth: 3D silhouette sensing for high-precision input on and above physical surfaces. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 1377–1386 (2014)
299. Ramos, J., Li, Z., Rosas, J., Banovic, N., Mankoff, J., Dey, A.: Keyboard Surface Interaction: Making the Keyboard into a Pointing Device. arXiv preprint arXiv:1601.04029 (2016)
300. Buschek, D., Roppelt, B., Alt, F.: Extending keyboard shortcuts with arm and wrist rotation gestures. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 21 (2018)
301. Lee, B., Park, H., Bang, H.: Multidirectional pointing input using a hardware keyboard. ETRI J. **35**(6), 1160–1163 (2013)
302. Zhang, H., Li, Y.: GestKeyboard: enabling gesture-based interaction on ordinary physical keyboard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1675–1684 (2014)
303. Bailly, G., Pietrzak, T., Deber, J., Wigdor, D.J.: Métamorphe: augmenting hotkey usage with actuated keys. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 563–572 (2013)
304. Shi, Y., Vega Gálvez, T., Zhang, H., Nanayakkara, S.: Gestakey: Get more done with just-a-key on a keyboard. In: Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 73–75 (2017)
305. Zheng, J., Vogel, D.: Finger-aware shortcuts. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4274–4285 (2016)
306. Zheng, J., Lewis, B., Avery, J., Vogel, D.: Fingerarc and fingerchord: Supporting novice to expert transitions with guided finger-aware shortcuts. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, pp. 347–363 (2018)

307. Sekimori, K., Yamasaki, Y., Takagi, Y., Murata, K., Shizuki, B., Takahashi, S.: Ex-space: Expanded space key by sliding thumb on home position. In: Proceedings of the International Conference on Human-Computer Interaction, pp. 68–78 (2018)
308. I. U. T. Inc.: A Brief History of the LCD Key Technology. https://visualled.com/en/led-screens/brief-history-of-lcd-screens/. Last access: 25.06.2022. https://support.apple.com/guide/mac-help/use-the-touch-bar-mchlbfd5b039/mac. Last access: 25.06.2022
309. Logitech: Logitech G19 Keyboard for Gaming
310. Razor: Razor Deathstalker Ultimate Keyboard
311. Maiti, A., Jadliwala, M., Weber, C.: Preventing shoulder surfing using randomized augmented reality keyboards. In: Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 630–635 (2017). https://mysupport.razer.com/app/answers/detail/a_id/3624/?/razer-deathstalker-ultimate-%7C-rz03-00790-support. Last access: 25.06.2022
312. Wang, X., Besançon, L., Rousseau, D., Sereno, M., Ammi, M., Isenberg, T.: Towards an understanding of augmented reality extensions for existing 3D data analysis tools. In: ACM Conference on Human Factors in Computing Systems(2020)
313. Schneider, D., Otte, A., Gesslein, T., Gagel, P., Kuth, B., Damlakhi, M.S., Dietz, O., Ofek, E., Pahud, M., Kristensson, P.O. : ReconViguRation: reconfiguring physical keyboards in virtual reality. IEEE Trans. Vis. Comput. Graph. **25**(11), 3190–3201 (2019)
314. Grubert, J., Ofek, E., Pahud, M., Kristensson, b.y.o., Per Ola: Back to the future: revisiting mouse and keyboard interaction for HMD-based immersive analytics. arXiv preprint arXiv:2009.02927 (2020)
315. Casiez, G., Vogel, D., Balakrishnan, R., Cockburn, A.: The impact of control-display gain on user performance in pointing tasks. Hum. Comput. Interact. **23**(3), 215–250 (2008)
316. Baudisch, P., Cutrell, E., Hinckley, K., Gruen, R.: Mouse ether: accelerating the acquisition of targets across multi-monitor displays. In: CHI'04 Extended Abstracts on Human Factors in Computing Systems, pp. 1379–1382 (2004)
317. Ashdown, M., Oka, K., Sato, Y.: Combining head tracking and mouse input for a GUI on multiple monitors. In: CHI'05 Extended Abstracts on Human Factors in Computing Systems, pp. 1188–1191 (2005)
318. Benko, H., Feiner, S.: Multi-monitor mouse. In: CHI'05 Extended Abstracts on Human Factors in Computing Systems, pp. 1208–1211 (2005)
319. Waldner, M., Kruijff, E., Schmalstieg, D.: Bridging gaps with pointer warping in multi-display environments. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp. 813–816 (2010)
320. Villar, N., Izadi, S., Rosenfeld, D., Benko, H., Helmes, J., Westhues, J., Hodges, S., Ofek, E., Butler, A., Cao, X. : Mouse 2.0: multi-touch meets the mouse. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 33–42 (2009)
321. MacKenzie, I.S., Soukoreff, R.W., Pal, C.: A two-ball mouse affords three degrees of freedom. In: CHI'97 Extended Abstracts on Human Factors in Computing Systems, pp. 303–304 (1997)
322. Olwal, A., Feiner, S.: Unit: modular development of distributed interaction techniques for highly interactive user interfaces. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 131–138 (2004)
323. Cechanowicz, J., Irani, P., Subramanian, S.: Augmenting the mouse with pressure sensitive input. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1385–1394 (2007)
324. Kim, S., Kim, H., Lee, B., Nam, T.-J., Lee, W.: Inflatable mouse: volume-adjustable mouse with air-pressure-sensitive input and haptic feedback. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 211–224 (2008)
325. Balakrishnan, R., Baudel, T., Kurtenbach, G., Fitzmaurice, G.: The Rockin'Mouse: integral 3D manipulation on a plane. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 311–318 (1997)
326. Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., Conway, M.: The videomouse: a camera-based multi-degree-of-freedom input device. In: Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology, pp. 103–112 (1999)
327. Steed, A., Slater, M.: 3d interaction with the desktop bat. In: Computer Graphics Forum, vol. 14, pp. 97–104 (1995)
328. Fröhlich, B., Plate, J.: The cubic mouse: a new device for three-dimensional input. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 526–531 (2000)
329. Froehlich, B., Hochstrate, J., Skuk, V., Huckauf, A.: The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 191–199 (2006)
330. Saidi, H., Serrano, M., Irani, P., Dubois, E.: TDome: a touch-enabled 6DOF interactive device for multi-display environments. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5892–5904 (2017)
331. Microsoft: Getting Started with the Windows Mixed Reality Toolkit (2019)
332. Völkel, S.T., Schneegass, C., Eiband, M., Buschek, D.: What is "intelligent" in intelligent user interfaces? a meta-analysis of 25 years of IUI. In: Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 477–487 (2020)
333. Grubert, J., Langlotz, T., Zollmann, S., Regenbrecht, H.: Towards pervasive augmented reality: Context-awareness in augmented reality. IEEE Trans. Vis. Comput. Graph. **23**(6), 1706–1724 (2016)
334. Nwana, H.S.: Software agents: An overview. Knowl. Eng. Rev. **11**(3), 205–244 (1996)
335. Hoy, M.B.: Alexa, Siri, Cortana, and more: an introduction to voice assistants. Med. Ref. Serv. Q. **37**(1), 81–88 (2018)
336. Norouzi, N., Kim, K., Hochreiter, J., Lee, M., Daher, S., Bruder, G., Welch, G.: A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents, pp. 17–22 (2018)
337. Dehn, D.M., Van Mulken, S.: The impact of animated interface agents: a review of empirical research. Int. J. Hum. Comput. Stud. **52**(1), 1–22 (2000)
338. Yee, N., Bailenson, J.N., Rickertsen, K.: A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1–10 (2007)
339. Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., Welch, G.F.: Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In: Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 105–114 (2018)
340. Norouzi, N., Bruder, G., Belna, B., Mutter, S., Turgut, D., Welch, G.: A systematic review of the convergence of augmented reality, intelligent virtual agents, and the internet of things. In: Artificial Intelligence in IoT. Springer, Berlin (2019), pp. 1–24
341. Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The ALIVE system: Wireless, full-body interaction with autonomous agents. Multimedia Systems **5**(2), 105–112 (1997)

342. Cavazza, M., Martin, O., Charles, F., Mead, S.J., Marichal, X.: Interacting with virtual agents in mixed reality interactive storytelling. In: International Workshop on Intelligent Virtual Agents, pp. 231–235 (2003)
343. MacIntyre, B., Bolter, J.D., Moreno, E., Hannigan, B.: Augmented reality as a new media experience. In: Proceedings of the IEEE and ACM International Symposium on Augmented Reality, pp. 197–206 (2001)
344. Anabuki, M., Kakuta, H., Yamamoto, H., Tamura, H.: Welbo: An embodied conversational agent living in mixed reality space. In: CHI'00 Extended Abstracts on Human Factors in Computing Systems, pp. 10–11 (2000)
345. Barakonyi, I., Psik, T., Schmalstieg, D.: Agents that talk and hit back: Animated agents in augmented reality. In: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 141–150 (2004)
346. Barakonyi, I., Schmalstieg, D.: Augmented reality agents in the development pipeline of computer entertainment. In: International Conference on Entertainment Computing, pp. 345–356 (2005)
347. Barakonyi, I., Schmalstieg, D.: Ubiquitous animated agents for augmented reality. In: Proceedings of the 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 145–154 (2006)
348. Chekhlov, D., Gee, A.P., Calway, A., Mayol-Cuevas, W.: Ninja on a plane: Automatic discovery of physical planes for augmented reality using visual slam. In: Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 153–156 (2007)
349. Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: part I. IEEE Robot. Autom. Mag. **13**(2), 99–110 (2006)
350. Blum, L., Wetzel, R., McCall, R., Oppermann, L., Broll, W.: The final TimeWarp: using form and content to support player experience and presence when designing location-aware mobile augmented reality games. In: Proceedings of the Designing Interactive Systems Conference, pp. 711–720 (2012)
351. Kotranza, A., Lok, B.: Virtual human+ tangible interface= mixed reality human an initial exploration with a virtual breast exam patient. In: Proceedings of the 2008 IEEE Virtual Reality Conference, pp. 99–106 (2008)
352. Kotranza, A., Lok, B., Deladisma, A., Pugh, C.M., Lind, D.S.: Mixed reality humans: Evaluating behavior, usability, and acceptability. IEEE Trans. Vis. Comput. Graph. **15**(3),369–382 (2009)
353. Weiser, M.: The computer for the 21st century. ACM SIGMOBILE Mob. Comput. Compens. Rev. **3**(3), 3–11 (1999)
354. Schmidt, A.: Biosignals in human-computer interaction. Interactions **23**(1), 76–79 (2015)
355. Schraefel, M.: in5: a Model for Inbodied Interaction. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–6 (2019)
356. Kazerooni, H.: A review of the exoskeleton and human augmentation technology. In: Dynamic Systems and Control Conference, vol. 43352, pp. 1539–1547 (2008)
357. Kunze, K., Minamizawa, K., Lukosch, S., Inami, M., Rekimoto, J.: Superhuman sports: Applying human augmentation to physical exercise. IEEE Pervasive Comput. **16**(2), 14–17 (2017)
358. Schmidt, A.: Augmenting human intellect and amplifying perception and cognition. IEEE Pervasive Comput. **16**(1), 6–10 (2017)
359. Benko, H., Holz, C., Sinclair, M., Ofek, E.: Normaltouch and texturetouch: High-fidelity 3d haptic shape rendering on handheld virtual reality controllers. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 717–728 (2016)
360. Schmidt, S., Bruder, G., Steinicke, F.: Effects of virtual agent and object representation on experiencing exhibited artifacts. Comput. Graph. **83**, 1–10 (2019)

Jens Grubert is Professor of Human-Computer Interaction in the Internet of Things at Coburg University. He has more than 10 years of academic and industrial experience in fields such as human-computer interaction, augmented and virtual reality (AR/VR), computer graphics, and computer vision. Jens Grubert is author of more than 70 peer-reviewed publications and patents.

Interaction with AI-Controlled Characters in AR Worlds

6

Christoph Bichlmeier

Contents

6.1 **Populating AR Worlds with Virtual Creatures** 131
6.1.1 AI Characters in AR Literature 133

6.2 **Designing AI Characters for AI Worlds** 135
6.2.1 Categorization of AI Characters 136
6.2.2 Architecture Base Components 138
6.2.3 Appearance 140
6.2.4 Movement 145
6.2.5 Reasoning 147

6.3 **Conclusion** 153

References 153

Abstract

With the release of commercial head worn augmented reality (AR) devices, the AR community has grown exponentially. These AR devices come with robust computer vision algorithms to perform adequate object tracking and spatial localization of the user as well as objects in the AR environment. They allow for user input in various ways to interact with the scene. In addition, these devices finally became lightweight and comfortable enough to be used for real jobs in real professional environments. A new era has begun that allows the development of AR applications on a new level. It is no longer necessary to be an expert in computer vision or computer graphics to create meaningful solutions for application ideas that mostly have been proposed already many years ago. The usage of top-level game engines enables app developers to integrate well-designed and textured models, animations, physical effects, and illumination into their solutions, even without programming knowledge. In the past, the big majority of AR applications has focused on placing additional rigid or animated objects into the AR environment. Alternatively, tracked real objects have been superimposed by virtual supplementary information such as virtual structures inside a real object. However, another type of augmented content has almost been neglected so far. This chapter addresses the integration of and interaction with virtual characters in AR environments. It provides an overview of research that has been published by the AR community in the past and classifies different types of characters. In addition, we introduce the reader to Game AI, an R&D subdiscipline of game engineering that focusses on controlling non-player characters (NPCs) in video games. Furthermore, this chapter discusses peculiarities of the interaction with NPCs in AR environments and takes a closer look at the psychological subjects namely perception, communication, and behavioral psychology.

Keywords

Augmented reality · Game AI · Artificial intelligence · Game engineering · Social presence · Social perception · Decision-making · Movement · Non-player characters · Virtual humans · Embodied agents

6.1 Populating AR Worlds with Virtual Creatures

The augmented reality (AR) community has experienced a paradigm shift in the last few years and opened its door for developers, artists, psychologists, and many more disciplines beyond specialists in computer vision and computer graphics to create exiting AR applications.

When the first smartphones were released equipped with a color display and a camera, many AR researchers focused on this type of display hardware. Consequently, powerful software development kits (SDKs) for mobile devices such Google's ARCore or Apple's ARKit provide today core AR

C. Bichlmeier (✉)
Fakultät Informatik, Studiengänge Game Engineering (BA) und Game Engineering & Visual Computing (MA), Hochschule für angewandte Wissenschaften Kempten, Kempten, Germany
e-mail: christoph.bichlmeier@hs-kempten.de

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_6

features such as scanning the space geometry and performing robust inside-out tracking to deliver AR experiences with mobile devices. An even bigger step towards the realization of Sutherland's original idea of generating AR environments has been made with the commercial release of robust, lightweight, mobile head-worn AR devices. The availability of *ultimate displays* [1] presenting the AR space in stereo and from the natural point of view, i.e., the user's eyes, lifted the state of the art of AR to the next level. For the first time, see-through head-mounted displays (HMDs) became available not only for the staff of exclusive research laboratories but for all developers.

Both mobile AR SDKs and HMD providers identified already existing game engines as powerful development tools to push AR out of the scientific niche. For this reason, Microsoft, Magic Leap, and other display providers as well as the AR SDKs adapted their development interfaces strongly to game engines such as Unity 3D or Unreal Engine. The tight connection between SDKs coming with AR devices and state-of-the-art game engines was a clever decision and created a strong impulse. From this time on, AR developers had access to powerful tools of the game engineering world to create appealing virtual content for their AR environments including easy integration of illumination effects, shadows, texturing, animation, 3D graphical user interfaces (GUIs), physics simulation, and basic artificial intelligence (AI) control features.

This chapter addresses the integration of virtual characters into AR environments. When the AR research community describes the vision of the ultimate AR scenario, quite often movies such as Jurassic Park (1993) or Space Jam (1996) are referenced integrating virtual characters into real scenes. Already in 1982, the children's TV series and movie Pumuckel inserted an animated cartoon character into a real environment. Other examples combining real environments with virtual characters are Avatar (2009) as well as various iterations of Godzilla and Transformers. While these visual benchmark examples required tedious preproduction work, AR always targeted the combination of real and virtual in real time. Recent demo apps of head-worn AR devices showcase their capabilities and get closer to the early visions. For example, the games Dr. Grordbort's Invaders coming with the Magic Leap One as well as the game RoboRaid shipped with Microsoft's HoloLens show classic AI-controlled enemies in video games now living in and interacting with the AR space.

Intelligent, visual agents are subject of research for decades [2–6]. One of the most famous commercial 2D agents has been Microsoft's office assistant *Clippit*, although it was not very popular. Various names for these intelligent characters have been introduced in the literature depending on the application, type of character, and research community. In the game engineering world, the terms *non-player character* (NPC), bots, AIs, or simply *characters* are used to distinguish from the visible creatures being directly controlled by players usually called avatars. In gaming, the design of characters is not restricted to highly realistic, anthropomorphic representations of humans. The most popular characters in the gaming culture are rather carefully designed fictive creatures such as cartoons of stereotypical humans, monsters, zombies, dwarfs, and similar fellows rendered with different artistic comics styles. Other communities focus on virtual characters as part of an experimental setup or as a replacement for real humans in training scenarios and other serious applications. They rather create anthropomorphic virtual characters showing human-like behavior and use the term *virtual humans* [7–10]. More general terms, also applicable for non-anthropomorphic characters, have been introduced by the literature such as *embodied conversational agents* [11], *embodied autonomous agents* [12], and *intelligent virtual agents (IVA)* [13, 14].

With respect to the display technology presenting immersive applications, many AI techniques to design reasonable behavior of virtual characters similarly apply to monitor-based applications such as classic desktop and mobile games and VR or AR experiences. Some of the summarized guidelines in this chapter to integrate virtual characters into AR worlds are valid as well for VR experiences when the world is presented from the user's natural perspective, for example, using HMDs. AR has always been an interdisciplinary domain incorporating and combining findings from other disciplines. In the past, for example, computer vision, mobile hardware technology, optics, computer graphics, and human computer interaction provided major input to construct AR experiences. However, time has come to consider the integration of wisdom from further fields of research to enhance immersion of AR environments such as communication, perception, social presence, as well as AI techniques to design meaningful behaviors of characters. Only a small amount of research papers has been published so far targeting the integration of NPCs in AR experiences. Most of the described characters only show very little artificial intelligence sometimes limited to a set of animations only. More advanced techniques such as autonomously exploring the AR space or making decent decisions have been mostly neglected. Section 6.1.1 will start with an overview of related work published by the AR research community.

The integration of AI agents into AR environments requires the combination of multiple disciplines that have been studied separately in the past. Section 6.2 summarizes this knowledge to design complex characters for interactive AR experiences. Section 6.2.1 starts with a categorization of different types of NPCs with respect to typical AR application scenarios and key properties of characters for these scenarios. Section 6.2.2 will discuss the fundamental requirements to release virtual characters into their AR space. In addition, it will introduce basic AI components and a layer model to design the AI architecture. The first architecture layer *appearance* is discussed in Sect. 6.2.3 including an excursion to the psychological subject of verbal and non-verbal communication.

In the world of game engineering, such NPCs were already part of the very early game titles. For instance, the early ghosts of Pac Man are equipped with some artificial intelligence to sense roughly the position of the player and adapt their behavior according to the player's state (hunting or fleeing). State-of-the-art game engines are shipped today already with a basic set of AI features to control characters, e.g., movement of an agent or state machines to control animations of a character. However, there is also a small research community reutilizing classic and inventing new AI techniques for virtual characters and adapting them to the needs of computer games, simulations, and comparable interactive experiences. Sections 6.2.4 and 6.2.5 will introduce the most popular techniques to move characters in space and to control decision-making processes enabling responsive behavior [3–5].

6.1.1 AI Characters in AR Literature

This review is limited to AR literature only, i.e., referenced publications investigate the integration of virtual characters into AR worlds. Found research work is listed chronologically in order to keep the evolution of display technology in mind and better acknowledge the researchers' efforts to realize their AR scenarios.

In 1995, Maes et al. [15] reported about their ALIVE System, a *magic mirror* AR interface that allows for interaction with virtual characters presented in the mirrored-augmented scene. They describe a comprehensive approach to trigger context-dependent behavior of agents consisting of a sensor system and a reasoner system to create appealing, reactive AIs. With the help of computer vision techniques, user input such as user position, body pose (e.g., bending down), and basic gestures with 3D information can be measured. In addition, user's utterances and input from the agent sensing other virtual objects and AIs being part of the AR environment control the character's actions. Additional information is evaluated by their decision-making system analyzing internal needs and motivations in the context of the storyline and the current situation. The agent's actions can be perceived with audio output, animations, and steering behaviors. Authors conclude that "successful gestures are intuitive [. . .] and provide immediate feedback" such as motion, facial expression, gestures, or utterances. In addition, they find that "users are more tolerant of imperfections" when dealing with agents in contrast to interaction with non-intelligent objects. AIs may have their own way of thinking and behaving while an object such as a switch should work as expected.

Cassell et al. [16] proposed in 2000 a mixed reality (MR) user interface (UI) that consists of a projection wall serving as a portal to a virtual extension of a playing setup for children. The wall separates a virtual and a real part of a toy castle, while in the virtual world, a small virtual child and playmate called Sam can interact with the real user. Sam is a so called "embodied conversational playmate" whose "behavior is modeled on children's social and conversational skills." They incorporate a "model of story-listening" being used to animate the child user of the MR system to "tell stories that are more imaginative and more narratively complex." Sam uses a microphone and pressure-sensitive mats to determine the presences and location of the user. Sam is animated to look at the user and gives responsive audio feedback for instance backchannels like "uh-huh" or inviting prompts such as "And then what happened?"

Anabuki et al. [11] introduced in 2002 Welbo a small robot living in an AR environment presented through a see-through HMD. Welbo can perceive and output audio to conduct a simple conversation. In addition, it reacts on instructions given by the user such as moving virtual furniture in their *MR Living Room* scenario and further "user's actions and movement." Animations during conversation and waiting periods and changes of location within AR space let Welbo behave like a living being. Authors suggested to adjust the size of the agent that users "can see Welbo's body in their field of view." Also "people feel uncomfortable when . . . " 'Welbo floats over' the user or Welbo gets to close. Welbo has one of the most advanced AI architectures identified in this literature review to establish social presence.

In 2002, Cheok et al. [17] published about a system combining VR and AR as well as body capturing to create interactive theater experiences. Authors follow a different strategy than other works described in this review to integrate characters into their MR scenes. Here, a capturing system records 3D shapes of human actors in real time and projects their reconstructions into the AR environment. For this reason, characters are not controlled with AI techniques.

In 2002, Balcisoy et al. [18] describe a set of interaction techniques with virtual humans in MR environments. They distinguish guided virtual humans which are avatars of real humans and autonomous virtual humans having their own sensor system and responsive "behavioral mechanisms." Authors report on two case studies employing such autonomous virtual humans being controlled by a so-called Virtual Human Director (VHD) equipping the NPC with basic behavioral features, i.e., speech as well as facial and body animation. The goal of the first study is the application of a virtual human to design appearance and movement of virtual entities of the AR scene in a more natural way. A checkers game has been chosen as an example to use the animated virtual human as the opponent player moving virtual pieces of the game. This leads to a better perception of time needed by the AI

6

imitating the thinking process of the virtual opponent. Also, the movement of a piece from one place to another place and the strategy behind this step can be better understood by the real player. In addition, animations can be designed to introduce social components and emotions enhancing the gaming experience. The second study uses a virtual human to test ergonomic parameters when interacting with a device, here a PC tower. Using the virtual human, testing becomes an automatic process that reduces human efforts and provides rich evaluation data.

One of the most advanced AI systems to control virtual characters in MR environments has been proposed by Cavazza et al. in 2004 [19]. Their system, an AR magic mirror, is used as a virtual stage incorporating a real actor as part of the scene to implement a platform for interactive story telling. The human user can interact with objects of the virtual stage including virtual characters. The scene follows story line, here a James Bond scenario. However, the human user can influence the narrative through multimodal input such as speech and gestures. For this reason, authors have developed a library of 15 semiotic gestures that can be optically detected to be interpreted depending on the narrative context. The story line including any interactive entities is controlled by hierarchical task networks (HTN), a data structure, and planning algorithm for constructing sets of actions and stories in real time upon changing situations due to multimodal input through the human actor.

Barakonyi and Schmalstieg [12] report in 2005 on their research of integrating interactive agents into AR environments for entertaining applications. They emphasize the need of maintaining the agent's world model in order to react adequately to any situation. This is only practicable, if the agent's knowledge is restricted to information relevant to the context of its role. They present different scenarios how real or virtual objects influence behaviors of virtual characters. One application shows an animated character balancing on a physical object. Pose of the physical object is used as input to trigger animations such as spreading waving arms to keep balance or falling. In another scenario, the NCP reacts on the pose of a physical object, a small Lego Mindstorms robot, to keep a safety distance. A third scenario introduces a sandbox AR game consisting of virtual and real objects, influencing the behavior of NPCs. Here "a dedicated control logic or virtual brain" makes decisions on available actions such as movement via pathfinding, a selection of animations, and playing sounds. All examples use optical markers to detect and track planes such as the table surface, tangible interactive devices, or the Lego robot. Authors highlight the importance and effect of appropriate, contextual animations visualizing, e.g., success, failure, falling, and jumping.

Magnenat-Thalmann et al. [20] proposed in 2005 to use animated virtual characters to invigorate AR experiences targeting cultural heritages. They mainly focused on the correct animation of those NPCs, for instance, skeletal animation, skin deformation, cloth simulation, or facial speech. NPCs are scripted to enhance storytelling and handle basic interaction with virtual objects. In addition, occlusion of virtual objects with real objects can be handled.

Wager et al. [21] evaluated in 2006 how realistic a virtual character being part of a AR scene should be. In order to teach art history, they compared different types of representations of a supporting agent being displayed on an AR window using a mobile device. The visualization modes were only text overlay, 2D overlay of a non-animated projection of a 3D cartoon character, a 3D-animated character in screen space, and finally an animated 3D character in AR space. This last "AR character" is the only representation fulfilling the definition of an AR entity. However, it is still rather an animated object giving static information via "text, audio playback and animation" rather than being interactive and reactive to the user's input. The term realism is used here to differentiate the types of spaces of an agent rather than the realistic representation of a virtual human versus a comic-like character. Their user study shows that subjects prefer the presence of the AR character.

Miller and Weising [22] filed a patent in 2010 disclosing a mobile device AR setup that places a virtual character into an AR scene showing "awareness of an observer," e.g., his/her position in space by "adjusting a view direction of the virtual character" including "head and eyes." Next to the character's view direction, they also mention the movement and gesturing of the virtual character as designated signals exhibiting awareness of the AR user.

In 2013, Tomi and Rambli [23] published about an AR storybook to teach preschool children how to count. Their AR hardware setup uses a mobile device and a real book. Children can use their fingers to interact with the AR content. A small animated crow plays a major role in this scenario. It comes with audio output to support the counting task and makes the experience more playful and attractive for the target audience.

Arroyo-Palacios and Marks [24] combined in 2017 a PlayStation VR Device with a stereo camera to create a video see-through HMD. They introduced a comprehensive example application consisting of little virtual reactive and interactive robots. Those robots are equipped with a large set of rules to adequately react on user behavior such as looking at the robots or other objects in the room. To design behavior, authors reference to a cat that "shares a physical space with us." Even though high-level intelligent communication cannot be expected, the cat "would acknowledge our presence, respond to our actions, behave, act and navigate in the same physical space." Similar expectations of intelligence are projected onto the robots, and users "would overlook incorrect actions or behavior" to a certain extend.

Hartholt et al. [10] presented in 2019 a virtual human, which is introduced as an "interactive digital representation of humans who can perceive real humans and respond appropriately, both verbally and non-verbally," to train young adults suffering from autism spectrum disorder (ASD) in job interviews. As AR hardware, the magic leap device is used. The virtual human is designed to react on the user's interaction, e.g., starting a verbal conversation. One of the authors' findings was the importance of animating the virtual human in order to maintain eye contact. In monitor-based AR applications, characters looking into the virtual camera seem to look straight into the eyes of the user when the scene is rendered onto the screen. However, AR environments require a different approach. Hartholt et al. use a hierarchy of animated joints following the user's motion adapted from the work of Lee et al. [25]. While moving around the virtual human, a coordinated look is created animating "the eyes first, then the neck and finally the spine, to keep facing the user" [10]. Hartholt et al. further used the Virtual Human Toolkit, "audio recordings, lip sync data and pre-authored nonverbal behavior Markup Language (BML)" [10].

Kim et al. run a series of experiments from 2017 to 2019 to investigate effects being beneficial for NPC's social presence [8, 9, 13, 13, 26]. Using the Microsoft HoloLens as AR hardware, Kim et al. examined [26] different approaches to maintain the "virtual human's spatial and behavioral coherence with respect to the physical objects or events." In particular, the misleading perception of incorrect visual occlusion of virtual and real objects as well as real objects implausibly not blocking virtual dynamic objects has been addressed. Using the spatial map of the HoloLens as an indicator for occluding real object geometries is beneficial to enhance the social presence of NPCs. If such conflicts cannot be solved visually, also an intelligent virtual human, here a virtual person in a virtual wheelchair, proactively asking for help can solve perceptual conflicts in a plausible way. For instance, the virtual human may ask the user to remove a physical barrier such as a blocking chair or a closed door. They also show that human users respect the space virtually occupied by a virtual embodied agent and walk around the virtual human in case it is blocking the user's way.

Kim et al. [13] investigated in 2018 social effects of intelligent virtual agents (IVAs) with different levels of visual presence on the human user of an AR application. They compared three different agents. The first has a voice but no body comparable to a virtual AI assistant such as Amazon's Alexa. The second agent has an animated body and voice but however remains at one spot in space. The third IVA has all features of the first two agents and can move around in space. With a user study, authors show indications that the user's confidence and trust in the agent increases, when the agent is rather fully animated, it is aware of its physical space and it can change states of the AR world such as switching off a lamp. It is also beneficial, when the IVA respects privacy by visually leaving the shared space when demanded. Visual behavior such as gestures, locomotion, and the body by its own enhances "the users' sense of engagement, social richness, and social presence."

The group around Kim also reports on studies exploring the "effects of subtle multimodal interaction on social presence with a virtual human" [8, 9]. Events in AR space should influence both, real and virtual entities. In this case, researchers investigated the event of induced wind by a real fan that also influences virtual objects. In case the virtual human shows awareness of the wind, e.g., by animating its head to look for the wind source, social presence is increased. They show that this kind of responsive AR environment helps the user the get the feeling of sharing the same space as the virtual human.

One of the most recent publications is a study on social interaction in AR of Miller et al. being published in 2019 [27]. The work provides a comprehensive literature review on psychological aspects and work from the VR and AR community with respect to the topics being addressed in their studies. For this book chapter, the first two of overall three studies are of interest. First, they report on "social facilitation and inhibition," which is an effect that has been intensively studied by psychological research for decades. The effect social facilitation reveals that individuals perform a simple task more efficiently when a second person is present. In contrast, social inhibition leads to less efficient performance of complex tasks when being not alone. Miller et al. show that these effects also occur when a virtual character is present, and the AR user is asked to accomplish either complex or simple tasks. Their second experiment shows that AR users respect the presence of virtual characters, e.g., avoiding overlaps in space while real collisions are per se not possible. Respecting the presence of a virtual character even remains when an AR user stops the AR experience, i.e., takes of the AR HMD and does not see the NPC anymore.

6.2 Designing AI Characters for AI Worlds

The design of a character consists of its embodiment, animations, and AI architecture. These components define how the user socially perceives the NPC sometimes described as the *social presence* [7]. The following sections serve as a starter kit to design AI components of characters for different AR scenarios.

Any character exists for a certain reason and must meet the user's expectations. For instance, a zombie's job is to attack the user. It has abilities to move in AR space, but cognitive abilities may be limited. In an AR training scenario, an assisting AI nurse reacts on requests of the user playing the surgeon. She/he hands over surgical instruments while staying at her/his designated spot next to the patient. For this reason, he/she does not need to move around, however, but needs some responsive intelligence while following a clear script of the surgical procedure. A virtual training partner to practice sales conversations may not move in space. Instead, the agent needs to show verbal and non-verbal communication skills. Depending on their role in the AR experience, these types of agent need an individual AI configuration. For this reason, sometimes there is no need to put effort in providing agents with *real* intelligence imitating human intelligence [3]. They rather need to be clever enough to meet the user's expectations and avoid obviously stupid behavior. We begin this guideline with a categorization of different types of NPCs in AR applications in Sect. 6.2.1.

AI-controlled agents in AR spaces are presented with display technology and act in a designated AR living space. In addition, any AI architecture uses low-level components containing and managing information to be shared among AI modules. Such architectures are constructed in layers having their own specialized duties while communicating with each other. Section 6.2.2 reviews these key components of AR system and AI systems.

When using NPCs, the main difference between virtual worlds and AR worlds can be found in the appearance layer of the AI architecture described in Sect. 6.2.3. The section starts with a review of findings from psychology about human-human verbal and non-verbal communication. We continue with an overview of research results of human-AI interaction and communication to extract a set of behaviors for AR agents enhancing their social presence.

Section 6.2.4 gives an overview of basic algorithms handling movement beyond animation. The objective of these algorithms is to navigate characters from their current position to a new position determined through user input or decision-making.

Section 6.2.5 discusses decision-making algorithms usually called reasoners. All reasoners evaluate available knowledge to decide on what to do next. The reasoner layer controls the features of the communication and movement layer and is the basis for intelligent behavior.

6.2.1 Categorization of AI Characters

With respect to the AR scene considered as a user interface (UI), Milgram et al. defined in 1994 the well-known reality-virtuality (RV) continuum [28]. Here AR has been classified as a subdomain of mixed reality (MR). According to the amount of virtuality/reality of a scene, the continuum lines up the virtual environment, augmented virtuality, augmented reality, and real environment. However, Milgram's RV Continuum does not cover all perceptible sources that enhance the user's immersion into an MR/AR environment. Stapleton proposed an extension for the techno-centric view of Milgram's RV continuum [29, 30]. Beside composition of the real and virtual world, also the cognitive world must be considered, which gradually adds the user's imagination to perceive an AR environment. AI characters filling the AR space with life have a major contribution to trigger the user's ability of feeling present. They help to make the AR scene feel right. Microsoft's shooter game RoboRaid is only one example where characters play an important role to enhance the user's imagination and make the AR experience successful. As part of our augmented reality master course at Hochschule Kempten, we let students play this game to get aware of design options and hardware capabilities to create AR experiences. Anytime, students fight against virtual enemies; we directly see the impact of moving characters on immersion and presence. Students even react emotional when getting attacked and try to evade missiles and enemies by moving through space, ducking, or jumping.

Next to gaming applications, there are many training and simulation scenarios that can take advantage of AR spaces filled with interactive and responsive virtual characters. In these scenarios, usually real, cost-intensive training environments and human actors are required, e.g., emergency care in hospitals and sites of accidents, firefighters, job interviews, sales conversations, appraisal interviews, patient education, military simulations, planning evacuations for public buildings, etc. Here, AR agents might have the potential to reduce costs, enhance learning, and individualize training experiences.

Every AI character exists to fulfil a certain task. In most cases, agents and their tasks serve as a vehicle to drive storytelling. In addition, the character also needs to meet the expectations of the user, e.g., in contrast to a gamer, it is likely that a maintenance worker will not accept a flying robot giving instructions with a comic style voice. The following subsections try to categorize different roles of virtual characters and the expected AI-driven abilities.

Trainer, Trainee, and Coaches

This type of character can be found in serious games trying to solve a real-world hurt in a more efficient way than standard methods not using AR technology. The buyer and users usually have a clear economic interest in using this AR application. They have strong expectations towards robustness and effects of the application.

A typical method in many professional fields requiring communication skills as a core competence, playing role games is inherent part of training. Exemplary application fields are, e.g., training sales conversations with customers, patient/physician conversation, or classroom trainings for teachers [10, 11, 15, 16, 21, 23, 30]. All applications have the intention to improve the processes of social interaction rather than entertaining the users. Serious objectives include a better transmission of learning content, taking control of a social group, optimizing structured conversation to transfer knowledge, or training persuasion skills. Interestingly, one of the first interactive AI applications is Eliza who has been designed in 1966 by Joseph Weizenbaum to break the Turing test. In its original version, Eliza has no body, and interaction takes place with text input/output. She imitates a psychotherapist mostly posing open questions animating the user to speak freely about his/her problems, such as "Please tell me more about it!." All applications of this application category have the ultimate but ambitious goal to break the Turing test. The objective of these agents is to truly replace real humans.

A virtual character may take over the role of the trainer or the trainee. It needs to be equipped with abilities to communicate verbally, e.g., provide instructions, pose and answer questions, and non-verbally with adequate gestures and mimics. Its behavior and appearance should realistically imitate real humans. Communication and decision-making processes require an authoring tool, to design behavior and structure the story or workflow of the experience.

One example is a training scenario for doctors to exercise patient education prior to a surgery. At Kempten University of Applied Sciences, Germany, we develop in cooperation with the TUM Medical Education Center (TUM MEC), TU Munich, Germany, desktop serious games to potentially replace cost-intensive human actors with virtual patients. One goal of the conversation addresses legal aspects, i.e., the patient agrees on the treatment. In addition, the patient shall feel confident and well informed when making the decision to undergo the surgery. The duration of this conversation usually limited by the tight schedule of medical staff and the interlocution follows a semi-rigid structure. While the trainee, here the doctor, launches and leads through the main storyline, the virtual patient can pose questions evoking interruptions or even bypasses along this storyline.

On the other hand, in most situations, agents in these scenarios may not have the capabilities to move around in AR space. They rather occupy one spot during the experience. In addition, their perception is limited to the user's communication input in most cases.

Subject of an Examination

This category employs embodied agents as subjects for experiments or a protagonist of an interactive cultural experience [8, 9, 12, 13, 17–20, 26, 27]. Typical tasks of this type of agents are performing ergonomic tests for human-machine interfaces, evaluation of production pipeline in a factory requiring handwork, planning capacities in public buildings, serving as a mannequin to showcase other objects such as fashion, or tutorials to interact with devices or tools. Alternatively, they are part of a simulation scenario to examine group behavior and strategic plans or evaluate parallel processes. With respect to entertaining applications, characters may play story relevant roles used to enhance storytelling in an interactive experience. They may play a background role to fill a scenery with life with the intent to create atmosphere such as birds, flies, and other creatures.

Users expect agents to serve as a subject of an exanimation delivering higher level information in a more efficient and appealing way than traditional methods. They expect applications to deliver rich evaluation data while saving costs for real-world experiments. In entertainment applications, users expect an enrichment of the stage setting with dynamic, living protagonists.

One of the first showcases of the HoloLens came from Case Western Reserve University showing a true scale anatomy model for medical education. At Kempten University of Applied Sciences, Germany, we created a similar application showing a true scale woman with her unborn baby at different stages of her pregnancy. Users can interactively transverse through the anatomical layers to explore her body and anatomical deformations as a result of pregnancy.

The role of these agents is rather passive, and responsiveness to user interaction is limited. Agent's behavior is rather manipulated using clearly defined parameter sets, e.g., to control appearance due to a story line or simulation settings. Sensing communicative input of the user is less important. In simulation scenarios, these agents become highly dynamic objects interacting with and moving within AR space, i.e., sensing real and virtual objects to take them into account for decision-making.

Assistants and Companions

AIs of this section have a closer relationship to the user [11, 15, 21, 23, 24], which may be visualized with a closer interpersonal distance than previously described character types. This also affects their abilities of verbal and non-verbal communication, e.g., interpersonal distance, informal speech, and structure of conversations. These characters wait for sensory input indicating demands to support users when needed. Also, they must be aware of the user's position in space to follow him and stay within his/her perceptible range to keep in a standby mode.

One major application of the so-called smart glasses as Google Glass usually providing a monocular video or image projection onto the glass of head-worn device is remote assistance. The wearer of the smart glasses can connect with an expert providing instructions to get a certain task done. Like Microsoft's idea of Clippit, such assistants can also be virtual and even exist in AR spaces. These fellows guide through a checklist of tasks, answer FAQs with 3D annotations and additional media, or even sense the progress of the main task or users' difficulties to provide appropriate support.

Video games with rather complex game mechanics regularly use tutorials showcasing, e.g., how to use the controller for typical situations during gameplay. In the same manner, such tutorials are helpful for customers to introduce features of a newly bought complex product or a maintenance worker extending his/her service portfolio. In entertainment applications, these agents may be used to tell background stories or bridge from one narrative thread to another. They may be used as team members to fight against opponent AIs or other users. In serious games, such characters may also be used to guide through learning content, i.e., add a verbal description of the subject, recapitulate a learning lesson, or answer option questions.

Characters of this category can move in space with a high focus on the user's position and view direction. They must respond to virtual and real events from any source. Their verbal and non-verbal communication skills are not designed to test the user, but rather supplement information when needed. In gaming and further entertainment applications, they are responsible to guide through the experience . However, in industrial applications, they are one of several tools to accomplish efficiently a task. Respectively, their proactive component needs to be designed carefully, either to create narrative flow or to progress the workflow.

Enemies and Opponent

This character type exists exclusively in entertainment applications [18, 24]. Usually, these NPCs have a short lifetime, and their intelligence is very specialized to fulfill a very specific task. With respect to the game, they need to follow the rules of game mechanics and sense only those events triggering their available options. Their verbal and non-verbal communication skills can be limited in most cases. However, they are equipped with decision-making algorithms with the objective to create a challenging experience for the user. These algorithms use basic sensor information of the user, e.g., position, view direction, move direction, and speed. In addition, they analyze the AR space to find strategic points on their battlefield such as unoccupied coverage points. Exemplary commercial applications are Dr. Grordbort's Invaders for Magic Leap One and RoboRaid for HoloLens. As a student's team project at Kempten University of Applied Sciences, Germany, the game app Cthulhu Attacks has been created showing several types of enemies attacking the user in waves.

6.2.2 Architecture Base Components

The following section provides a summary of technical preconditions to successfully insert NPCs into AR spaces. In addition, several basic components of an AI architecture are introduced building the foundation of higher-level AI features namely movement, decision-making, planning, and learning.

The big majority of commercially available head-worn AR devices are optical see-through HMDs (OST), e.g., the Magic Leap One or both generations of the Microsoft HoloLens. OST provides a direct view onto the real portion of the AR scene, while virtual entities are projected onto a semi-transparent screen in front of the user's eyes. On the other hand, video see-through HMDs (VST) record reality with video cameras and compose the video images with virtual content to be presented on a video screen in front of the user's eyes [31, 32].

Each type of AR display hardware has its advantages and drawbacks. Using OST, virtual objects can only be drawn semi-transparent but never fully opaque. This drawback becomes particularly visible in very bright environments, i.e., illuminated with daylight. Also, there is always an uncertain transformation introduced within the tracking and registration matrix chain lying between the user's eyes and the screen showing the projection of virtual content. This becomes unacceptable, if the main quality criteria of the AR scene is defined by the accuracy of registering virtual objects with real-world objects, e.g., augmented reality-supported surgical navigation [33]. In addition, it is not possible to register the real and the virtual world over time. The dynamic view of the real environment is limited to the user's abilities of perception and cognition. However, the view of the superimposed virtual environment is restricted by the capabilities of the underlying display, tracking, and rendering hardware. Characters are in most cases dynamically moving entities, which do not require submillimeter registration with real locations. Users simply cannot perceive discrepancies in registration accuracy. For this reason, OST devices sufficiently fulfill the criteria to successfully integrate virtual characters in AR space.

VST devices still lack in screen resolution, which can be easily experienced with today's available VR HMDs. The visible pixels clearly affect the user's immersion. However, VST HMDs allow for the manipulation of the real part of the scene, e.g., the creation of views into real objects covering augmented internal objects such as the patient's anatomy [33]. When it comes to image composition of real and virtual with the intent to create a seamless transition between both worlds, the higher level of image control in VST systems is advantageous. Sometimes filters are used for the composed AR images to assimilate image quality of both worlds [24]. In addition, VST systems, for example, ease the analysis of real environment illumination to apply it to virtual objects [34].

One of the core features to allow virtual characters live in and interact with the same physical space as the user is the knowledge about both virtual and real geometries of the AR environment. Knowing the space geometry of the real world, sometimes called the spatial map, allows characters to respect and react on natural boundaries such as the floor or walls. It further allows real objects to correctly occlude virtual entities when appropriate. Also, real-time shadows as part of the AR illumination model can be cast by virtual objects onto real objects and vice versa. Some of the today's AR SDKs, e.g., HoloLens, are even able to compute higher level information of parts of the world geometry, e.g., the floor, walls, windows, or doors.

Figures 6.1 and 6.2 show the result of Lucien Scherer's bachelor theses at Kempten University of Applied Sciences, Germany [35]. The project uses Microsoft's HoloLens, the game engine Unity, as well as a life simulation algorithm called goal-oriented behavior (GOB) (see Sect. 6.2.5). Small minions behave according to internal needs that need to be satisfied by objects located in the AR environment, e.g., hunger can be satisfied by resources provided through user interaction. With the help of the user adding bridges or gondola lifts to the AR space, characters can overstep gabs, barriers, and obstacles of the real environment to reach their resources.

When setting up an architecture for AI characters, we recommend a generic approach using core components that become more and more specialized when required. The most important basic structures in such a system architecture are *options*, *sensors*, *knowledge*, *considerations*, and *reasoner*.

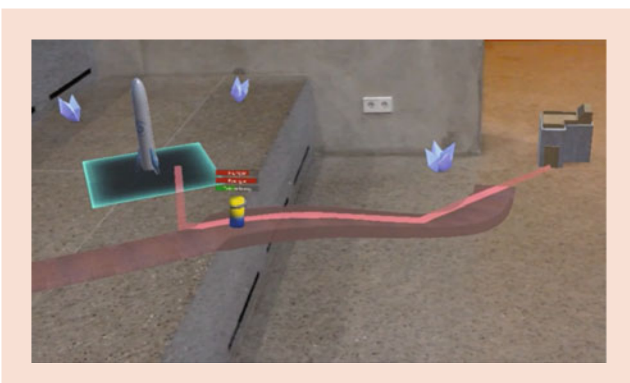


Fig. 6.1 Blue flowers provide resources to satisfy needs of yellow minions controlled with the GOB algorithm. (After [35] with permission of Lucien Scherer)

Fig. 6.2 Real barriers can be overcome with bridges or gondola lifts. (After [35] with permission of Lucien Scherer)

The component option describes what the character can do. Options can be perceivable to the user or invisible. Every character owns a set of specialized options reflecting its role in the application. Options may be an idle animation or starting an audio output such as greeting the player. Other options may change the state of the story, switching a light on, or change an internal parameter of the NPC. Options that can be perceived visually or acoustically by the user are often called actions. The set of available options may change, i.e., some options disappear and appear during runtime. Sometimes options are available through objects of the AR space. However, these objects may be created or destroyed during the AR experience.

The component knowledge describes the information available to the character. Knowledge can have any source in AR space and serves as the basis for decision-making processes. This may include sensor information about the player such as its position in space or its gaze direction. It may also include the character's own internal states such as an energy value. Usually a small set of specialized knowledge components such as the value of a float variable or a 3D position in space is enough to model the character's knowledge. Some more advanced options require knowledge, e.g., an action for the movement from a place A to a place B using a path finding algorithm (see Sect. 6.2.4) requires knowledge about the starting position, the destination, and the map of the AR space.

The component consideration handles the information to evaluate knowledge. An example would be a consideration that handles a threshold value to judge the knowledge about the NPC's distance to the player. Only if the character is close enough to the player, it may react on his/her gestures. The same knowledge may be evaluated by different considerations, e.g., an option idle animation is started once the distance to the player is greater than a certain threshold.

Considerations and options feed the so-called reasoners to decide on the next action/option. There are many different approaches how to design reasoners to allow for believable decision-making depending on the character type. In Sect. 6.2.5, a set of standard reasoners from the game AI world is described.

Knowledge of AI characters can be implemented as *hard-wired* to access the most relevant information without cost-intensive sensor algorithms. For example, it can be a good idea to provide the character with the information about its walkable living space as part of the initialization process of the application. In case the space changes, the character needs to sense appearing and disappearing obstacles or objects of interest, either real or virtual ones. In most game engines, visual sensor systems can be implemented by built-in ray cast systems sending a sensor ray from the character's position towards its view direction to determine information about objects within a certain distance range. These objects may be entirely virtual entities, or the spatial map representing

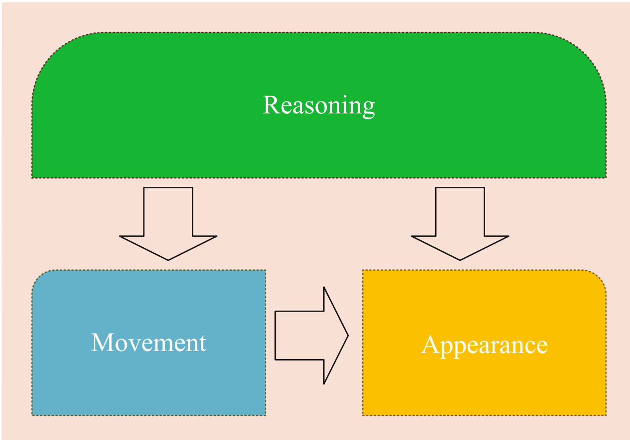


Fig. 6.3 A basic AI architecture consists of three layers interacting with each other while having their specialized responsibilities

the geometry of real entities. A second sensor system might handle acoustic input, e.g., determining when the AR user starts speaking or harrumphs. Sensors update knowledge when appropriate.

We propose an AI character architecture (Fig. 6.3) that consists of three layers having specific responsibilities but share information and interact with each other. The *appearance* layer handles the mimesis of the character via specialized options, including audio output, animation of body parts and face, and communication. This layer generates the impression of the character's personality, its social presence, or social perception during runtime. Any function of this layer can be perceived by the user. The *movement* layer consists of more complex options moving the character from one position in space to another using steering behaviors and/or path finding algorithms. Compared to the layer appearance, the layer movement has rather technical features than handling the personality of the character. The *reasoning* layer is responsible to evaluate available knowledge in order to make reasonable decisions. A decision generally leads to an option of the layer's *appearance* and *movement* changing the world state, e.g., moving the character to a new place, changing some internal parameters, starting an animation, or generating audio output.

6.2.3 Appearance

The following section reviews findings from human-human [36] and human-agent communication [7, 14]. Research targeting social presence and appearance factors of virtual characters [7, 14] is usually inspired by findings of human-human relation looking at effects of rapport, communication, and interaction. The book of Röhner and Schütz [36] on psychology of human-human communication summarizes the most relevant research literature in this field. Their findings are reviewed in the following subsection and human-human communication used to extract a set of guidelines beneficial for human-AI communication in AR applications.

Human-Human Communication

Communication takes place among two or more participants exchanging messages with various modalities. It depends on the context of the situation, here the subject of the AR application. In addition, it is designed as a process. While verbal communication shows usually a sequential process pattern, non-communication can take place also in parallel, e.g., mimics combined with gestures.

A communication process creates an impression of the communication partner, which is the social perception. Social perception is constructed by proximal and distal components. While proximal information can be directly perceived, e.g., clean clothes, a firm handshake, or a clear voice, distal information is an interpretation from a summary of proximal information, e.g., trustworthiness or authority. Research [36] has shown that attractive persons are rather attributed with intelligence and sincerity.

One of the most important goals of communication is influencing the communication partner's attitude towards a certain subject, e.g., in a product or service, politics , or learning subject. It has been shown [36] that the message should rather contain a contract, structured instruction than a general suggestion, e.g., "Make a project plan, read the literature, plan the software structure, create a prototype . . . " rather than "Try to give your best and don't fail!." Further efficient communication mechanisms to convince a communication partner, e.g., to participate in an AR experience, can be extracted from principles reported by Cialdini [37]. One may create a social dependency by giving free information, goods, and probes that causes the pressure of needing to return the favor. In addition, a message such as "Everyone enjoys doing . . . " giving the impression of a general social consent helps convincing. The efficiency of convincing another communication partner with a certain message depends on the cognitive abilities of the receiver, the context, and situation. The elaboration likelihood model introduced by Petty and Cacioppo [38] differentiates the central and the peripheral way of processing information. Using the central way, information is critically analyzed to judge its value and meaning. The peripheral way rather uses mechanisms known from classical and operand conditioning. Information is received here without further elaboration. Important factors influencing the way we process information are cognitive abilities; time pressure, i.e., how much time do we have to think about it; and personality, i.e., how much do we enjoy thinking.

While non-verbal communication describes *how* we say something, verbal communication describes *what* we say. Representatives of popular science often refer to the 7-38-55 rule claiming that human communication consist of 55% non-verbal, 38% paraverbal (accent, voice volume, and pitch of voice), and 7% verbal messages. Röhner and Schütz claim that this distribution is rather exaggerated [36]. However, if we are forced to make a choice, we rather trust in non-verbal messages than verbal messages. For instance, the content of the verbal message "I'm deadly serious about this!" is easily overwritten when the teller is grinning. Communication partners with higher emotional intelligence more likely identify the discrepancy between verbal and non-verbal messages, since they have learned from experience that verbal messages can be easier distorted.

Non-verbal Communication

Communication without the spoken word has been proven to be more powerful than verbal communication [36]. However, in an international context, non-verbal messages can be easily misunderstood. While head shaking indicates a refusal in the culture of the western world, in India a very similar looking head motion has rather the opposite meaning. Also, the hand gestures *OK* or the *heavy metal* sign has varying meanings in different cultural environments.

Touching the conversation partner can have both negative and positive effects depending on the subjective judgement of expertise and attractiveness as well as the status, the context, and sex of the communicators. Let us assume there is an established hierarchy between conversation partners such as boss and employee, senior and junior, master and student, and waiter and guest. The one having a lower status touches with conventional gestures such as handshake, while the person having a higher status tends to rather touch arms and shoulders conversation partner. The acceptance of touching another person as a form of communication strongly depends on the cultural background and applied rules. The handshake is the most common haptic signal, and its quality being evaluated by the research influences the judgment of a conversation partner. In VR and AR worlds, special cloths and gloves, so-called wearable haptic interfaces [39, 40], translating virtual touches to haptic feedback have been proposed to enhance immersion.

Gestures relate to moving and shaping arms and hands to form the so-called emblems and illustrators. Emblems are signals that can substitute verbal messages, e.g., clapping hands. Illustrators rather support a verbal message, e.g., a clenched fist to highlight one's rage or feeling of success.

Moving the head, nodding, or initiating/stopping eye contact can indicate the desire to change the direction of verbal communication and tells the opposite to take the lead. Shirking from someone's look usually leaves the impression of nervousness or hiding something or even lying. However, it has been shown that someone avoiding direct eye contact does not necessarily mean that the person is lying.

Posture gives information about attitude, emotion, and status of a person. Persons with a higher status tend to take a more relaxed posture, e.g., crossed arms and hands in the pockets. In contrast persons with a lower status, rather take a stiff posture. There is only little research investigating the relation between personality and posture; however, it is popular to project negative personal attributes, e.g., to a slack posture or hands in the pockets of a person having a lower status.

Psychologists distinguish between unilateral and bilateral eye contact, while the latter form requires both communication partners looking into the eyes [36]. Depending on the context of the situation and the personality of the communication partner, eye contact can stand for confrontation, dominance, or devotion. The duration of eye contact needs to be chosen carefully since too long eye contact may turn the impression from friendly to threatening. Blinking eyes is a natural behavior. If eyes do not blink, the character seems to stare at the user, which leads to the uncomfortable feeling of being observed. Blinking eyes can also have a communicative function, e.g., to show non-verbally agreement, to manifest a social relationship, to greet, or to show stress or discomfort.

Facial expression is one of the most powerful communication mechanisms to transport emotions. A smiling opposite is judged more positively than a neutral mien. The French neurologist Guillaume-Benjamin-Amand Duchenne was the first scientist investigating the differences between an artificial, posing smile and a natural smile. The so-called *Duchenne Smile* describes in detail what makes a smile authentic, among, e.g., a parallel and symmetric contraction of the zygomatic muscle and the orbicularis oculi muscle. A facial expression is a composition of all parts of the face, i.e., mouth, eyes, and eyebrows, and different combinations of face motions result in different meanings.

Interpersonal distance is one of several subjects of the so-called proxemic. For a face-to-face interaction, Hall [41] differentiates four categories of personal spaces around the communication partner. The (1) intimate space (15–45 cm) is reserved for close social relationships such as family and life partners; (2) personal space (45–120 cm) is reserved for close relations to friends, e.g., informal conversations; (3) social space (120–370 cm) is used for professional and impersonal conversations; and (4) public space (370 cm and visual/hearing range) is intended for any other social relationship. These distances have been investigated with white North Americans [36]. Other cultures may use different ranges. It has also been shown that a communication needs a bigger range when face-to-face communication takes place seated (social or public distance) rather than standing (intimate or social).

Verbal Communication

Along with speaking, also listening belongs to verbal communication. An active listener showing interest and attention gives feedback using nonverbal messages such as nodding or smiling, keeping eye contact, preventing external interruption, as well as verbal feedback such as posing questions or culturally specified sounds such as “Ok” or “mhm.” Call centers using automatic data assessment with recorded voices such as “Please tell me your Zip code” usually check the correctness of the user input by repeating the understood content. In fact, psychologists use a very similar pattern for listening to their clients in order to get information but also to show interest, trustworthiness, and mindfulness to establish a relationship with the client. Their method of active listening consists of three components [36]:

1. Active following the spoken word, i.e., confirming sounds and eye contact,
2. Active understanding of the message, i.e., paraphrasing what was said.
3. Understanding and reflecting the emotional content of the message, i.e., repeating what was said.

During verbal communication, entirely missing eye contact, inappropriate facial expression, and hints indicating distractions are clear signs of inattention and should be avoided.

Questioning and answering are an important part of verbal communication. Questions initiate, maintain, and guide through a conversation. They are used to receive information “Which machine do you want to maintain?” or provide information “Did you think of pressing this button first?.” In addition, they can be used to attract the attention to a certain subject and provoke cognitive processes, e.g., “Why is it necessary to read the security instructions first?.” There are different types of questions, i.e., open questions, closed questions, and leading questions. Since verbal conversation with AI agents is still far away from being a truly intelligent communication, one would rather avoid open questions requiring deep analysis of what the user said to the AI. In addition, answering open questions requires time. This may not be helpful since most AR applications are designed for reaching a goal in a more efficient way. Even entertainment applications are designed to provide an interactive, diverting experience avoiding telling the story or necessary background information in a tedious way.

Closed questions can be designed to guide the user to the right direction of the story, the learning experience, etc. The formulation of a set of possible answers can influence the probability of giving an answer while still creating the impression of having a choice and controlling the interactive experience. Yes and No questions such as “Do you like to join me to explore … ?” can be easily answered with capabilities of today’s speech recognition APIs. Also, a selection of possible answers, such as “How experienced are you in AR worlds?” having answers such as (1) “‘I’m a beginner,” (2) “I have little experience,” or (3) “I’m an expert,” and variations spontaneously used by the user can be investigated by a decent piece of software. The same counts for questions to identify information such as “How old are you?”

Leading questions are a powerful tool to manipulate answering or achieve a higher probability for a certain answer being favorable for the overall AR experience, e.g., “Is there a useful tool around that we can use to open the machine cabinet?” or “Wouldn’t a clever maintenance worker read the safety instructions first?” Children are more responsive to leading questions than adults. Also, persons with lower IQ, bad memory, higher anxiety, or higher credulity rather follow suggestions hidden in leading questions. Parameters to categorize a user can be checked by open questions before an AR experience starts to design an individual and better AR experience for the user.

In many situations, answers are explanations, which can be categorized in (1) interpretative (What does this mean?), descriptive (How does it work?), and reasoning explanations (Why is this the case? Why does it work?).

When designing a dialog between an explaining AR character and the user, one can follow the P5 process [42] to design the communication:

1. **P**re-assessment of the receiver’s knowledge: AI poses closed questions to better understand the user’s foreknowledge.
2. **P**lanning the goals of the explanation: AI starts reasoning how to structure and compose the explanation.
3. **P**reparation of necessary material: The AI’s reasoner decides, which methods and materials shall be used, e.g., an explaining video, manuals, helping icons such as arrows pointing to the point of interest.
4. **P**resentation: AI starts explaining using the prepared material.
5. **P**ost-mortem: The AI checks, if the explanation has achieved its goals, e.g., by invoking a test.

AI characters shall be designed to facilitate a social relationship with it in order to exceed the simple acceptance of its existence. This relationship can be intensified by *We*-sentences in an explanation, e.g., “We can get this machine working again!” Humor is a powerful mechanism to influence one’s attitude towards a person or a product, which is frequently and successfully used in advertisement. It can be used to reduce stress, e.g., when a complex task needs to be managed or the user feels uncomfortable when using an AR device for the first time.

An AI character should be able to laugh in appropriate situations. The reason for laughing is not necessarily restricted to respond to a funny joke. It has been shown that in a 10 min conversation, humans laugh around 5,8 times [36]. Laughing is used to show commitment and attention to what was said, but also to highlight or moderate a statement.

Human-AI Communication

Behavioral and appearance parameters of virtual characters have been intensively studied in the past. Nahal et al. [14] review the "most influential user studies" published at the ACM intelligent virtual agents (IVA) conference between 2001 and 2015. Studies are categorized in "non-verbal and verbal behavior," "physical appearance and identities," and "applications." They found that only 1.1% of these studies use AR technology as the UI interface to AI-controlled agents and suggest that more research should be done in this field. In addition, they show that mostly behavioral aspects are investigated. These behavioral aspects can be differentiated between social and affective behavior. The most important aspects under investigation are "presence/absence of audio, pitch, prosody, backchannel (BC), turn taking, body posture/gesture (upper-torso, arms, hands, legs, etc.), facial expression, gaze ...." Social behavior appropriate to the social context strongly influences how the character is perceived by the human user/player. For this reason, the NPC needs a mechanism to perform the so-called social signal processing, i.e., sense, decide, and act according to the social context of a situation. The resulting backchannel (BC) will then positively influence the human user's attitude towards the character. The character-human relation is part of rapport studies, investigating parameters of BC signals such as "contingency and frequency, quantity, type, and timings." [14]. Characters can express emotions using "facial expressions, gaze, gestures, and behaviors" to influence the user's decision-making, the level of cooperation with agents, agent believability, and human-agent emotional contagion [14]. Researchers explored the emotional effect of "adding wrinkles, blushing, sweating and tears." The big majority of agents reference in [14] does not share an AR space with the application user; hence there is a hardware barrier such as a video screen between agent and human establishing a communicative barrier. This is different for AR interfaces. The user is no longer represented as an avatar or entirely invisible. The human user truly shares the same 3D space with the character in the same way as with another human interactor.

Another recent review study of Oh et al. [7] analyzed in 2018 overall 152 studies to identify predictors of social presence in VR environments. The term *presences* in the context of VR applications needs to be distinguished from the term immersion. *Immersion* depends on the quality of technology making the user feel to be in an alternative environment than its real surrounding, e.g., quality of display, sound, framerate, field of view, or realism of computer graphics representing the virtual human. Presence is a "subjective experience of actually being in the mediated virtual environment." Social presence is considered as the "subjective experience of being present with a real person and having access to his or her thoughts and emotions" [7] and requires a virtual character to interact with. According to Oh et al., factors of a virtual human enhancing its social presence are showing individual traits, maintaining physical proximity, allowing multimodal interactivity, providing identity cues, creating a communication context, allocating high quality technology to increase immersion, giving demographic information, as well as indicating "psychological traits associated with positive attitudes toward social interactions."

Although in both review studies [7, 14] AR literature is almost not present, in particular, research of VR [7, 43, 44–46] using HMD displays may inspire HMD-based AR applications, since both display UIs present their contents from the natural point of view.

It has been shown that a believable character does not necessarily need to have a realistic human-like appearance to be accepted by the user as a coexisting creature [3]. This is valid primarily for entertainment applications such as video games. As Kevin Dill mentions in his chapter "What is Game AI" in [4], players of video games are "willing participants" accepting also unnatural body shapes and motion or superhuman abilities. The *suspension of disbelieve* is a theory first described by Samuel Taylor Coleridge in 1817. It is used in any media form involving storytelling, e.g., theater, animation movies, or video games to describe that a user is not distracted by unnatural content and is ready to accept also the supernatural and fiction. However, there are many more AR applications beyond entertainment such as games, marketing, or interactive cultural experiences. The principle of *suspension of disbelieve* may work also in industrial applications if AR is considered as a novel media form. However, once it is established as efficient, cost-saving, serious technology, their user groups may have different expectations beyond getting entertained. For instance, users of military simulations or medical training application might even expect high realistic virtual humans; however, today's AR hardware capabilities do not allow for such high-quality renderings. Even, if the AR hardware existed, it is very likely to run into the uncanny valley effect, which is well-known when designing android robots or virtual characters. If one tries to realistically imitate a human face including its motion and this attempt is not entirely successful (this is today usually the case), the character leaves a rather spooky impression. Even though it has been shown that anthropomorphic characters of the current level or realism are better to gain commitment and cooperation, photorealistic NPCs are not necessary for cooperation. Higher realism increases the expectation of intelligence [21].

Conclusions for AR Characters

The appearance layer serves as the perceivable interface between agent and human user. With respect to the basic components of an AI structure introduced in Sect. 6.2.2, its mechanisms are options influenced by the motion layer and triggered by the reasoning layer. For instance, an AR character assistant needs to fetch an item at the other side of the room following a command of the human user. Depending on the voice volume and the success rate of preceding work orders, the agent needs to react rather relaxed and seasoned or agitated and nervously. Depending on the context of the situation, velocity of movement and related animations might differ.

Inspired by the literature review above, a list of design guidelines has been extracted to configure the appearance layer. With respect to the design of the character's body, i.e., the proximal components to create social perception, we recommend to create a clean, attractive, sympathetic look while fitting to the context of the AR application, e.g., a virtual human for serious applications and a comic-like character for gaming applications. The design including a shape, textures, and characteristic motion patterns using animations must meet the expectations of the user group. In addition, an appropriate voice and language needs to be chosen.

The character needs to be perceived as a living individual at any moment playing its role and guiding through the storyline or workflow:

- Create idle animations such as frequent body or head motion and blinking eyes, including standard gestures such as scratching the arm or nose, catching a deep breath, or looking randomly around.
- Adapt motion patterns and posture of animations to the personality of the character.
- Generate a set of harmonic facial expressions reflecting any emotional state of the character's role including animations to blend these facial expressions.
- Create sympathy with design and humor to establish a relationship and overcome stressful situations.
- If manageable, use wearable haptic interfaces to transmit touches for non-verbal communication.
- Use communication principles to AI to convince the user to act towards a certain direction, e.g., to return to the main story line while giving the impression of self-determination and control over what is happening.

The character needs to show awareness of the user's presence and personality:

- Create animations for coordinated looks to objects in AR space and the user, i.e., eyes move first. Then, if necessary, the neck follows and third the body rotates [12, 24].
- Send positive feedback to verbal input of the user and user actions to raise self-esteem.
- Create motion patterns and posture for the body and the face to show awareness of a speaking user. Body and facial expressions may to be controlled separately to create a big set of possible combinations.
- Respect and dynamically adapt the interpersonal distance between user and the agent depending on their intended social relation, e.g., companion versus coach.
- Limit sensor capabilities to observe and process social signals of the user, such as auditory ability limited with values for distance and volume, visual abilities limited to a field of view, and sense for a personal safety distance to maneuver out of risk zones.
- Make the character laugh in appropriate situations, e.g., to show commitment and attention.
- Adapt messages to the character's audience, e.g., avoiding technical terms when users are laymen or novices: clear instructions versus philosophical statements.
- Adapt of non-verbal and verbal output to the user's cultural background.

The character needs to communicate with the user:

- Manage the character's sequential audio output, i.e., avoid parallel audio plays and interruption of the user.
- Create options to be executed in parallel for verbal and non-verbal communication, e.g., for enhancing the spoken word with gestures and mimics. This may require parallel processing of multiple reasoners and option managers.
- Care about timing in sequential verbal communication, e.g., leave time for thinking and introduce golden silence.
- Adapt information transfer to the cognitive abilities and personality of the user group and the time constraints of the application scenario to trigger central and/or the peripheral knowledge transfer.
- Follow principles to enhance persuasion, i.e., rather give a contract, structured instruction than a general suggestion, sell information as free goods/presents, and emphasize an impression of a general social consent on the subject.
- Contextual alignment of verbal and non-verbal communication output to avoid annulling or even reversing information.
- Use well-established emblems and illustrators to enhance information of messages.
- Use establishing and releasing eye contact and head motions to control the flow of dialogs.
- Use eye contact and head motions to imitate emotions such as nervousness.
- Control gaze direction and duration to encode attitudes such as confrontation, dominance, or devotion but also to show awareness of a speaking user.
- Use eye blinking for non-verbal messages.

- Create mechanisms for active listening and give feedback according to the user's actions, speech, and motion.
- Prevent the character to pose open questions. Only use this type of questions to trigger cognitive processes, e.g., to think about a certain subject.
- Use closed questions to gain information from the user and leading questions to guide the user through the AR experience.
- Prepare the level of detail of information given through answers according to the cognitive abilities of the user and his/her objectives.
- Structure dialogs, e.g., using the P5 guideline or similar concepts.

Every property described above needs to be designed as options and higher-level behaviors usually being designed as a sequential or parallel executed set of options. One of the most noticeable option/gesture is the *coordinated look* of an agent to view either the user or another object of the AR scene. This control of the character's gaze may be induced by one of the following events.

- User gazes at NPC.
- User points at NPC.
- Users show interest in something by looking at it.
- User moves in space.
- User performs gestures to attract attention, e.g., waves.
- User makes noise to attract attention (scream, harrumph).
- User says commands.
- A real or virtual object makes noise.
- A real or virtual object moves.
- A real new virtual or real object enters the scene.
- A real new virtual character or real user enters the scene.
- An extraordinary event happens, e.g., an explosion or a collision.

All events triggering coordinated looks are measured by the sensor system gathering knowledge for reasoning and serving as parameters for movement algorithms.

6.2.4 Movement

Standard Game AI uses techniques to let characters autonomously move from one location to another. Movement must be differentiated from motion through animations. Animations are usually prerecorded motion patterns such as moving the legs once the state of the NPC changes from standing to walking. Animations typically handle combined movement of body parts such as head, legs, and arms and facial expressions. However, this section addresses the movement of NPCs in space. Such movements can be designed as options being a result of decision-making. There are two primary approaches how to move a character from A to B in space that can be combined or work as separate entities.

The first approach is known as *steering behaviors* introduced by Reynolds [47]. To achieve dynamic movement of a character, the Euler integration method is used that splits the continuous motion equation into two first-order differential equations.

1. $\vec{p}\,(t) = \vec{v_0}\,t + \vec{p_0}$
2. $\vec{v}\,(t) = \vec{a}\;t + \vec{v_0}$

These equations are usually called in every update frame of the game loop and are parametrized with delta time. The first equation combines the character's current position with its velocity vector which is a direction vector encoding the speed and direction of movement. The second equation manipulates the velocity using an acceleration force. This acceleration force is the output of different types of steering behaviors. Reynolds introduces a basic set of steering behaviors that move single NPCs but also realize coordinated movement of NPC groups. However, many more specialized steering behaviors have been reported since the work of Reynolds [47]. The most basic behaviors are as follows:

- FLEE: The NPC computes a direction vector from the player to the NPC serving as the acceleration force to move away from the player.
- SEEK: It is an opposite behavior of FLEE.
- ARRIVE: It works as SEEK but stops at the target location without oscillation effects.
- WANDER: This is an idle behavior, letting the NPC walk around randomly in its living space.
- FLOCKING: It is a coordinated movement of NPC groups resulting in a bird like swarm sharing velocity and keeping a minimum distance to each other without colliding.

More special steering behaviors can be implemented, e.g., to avoid collisions with other NPCs or geometries in AR space. In addition, they can be combined to create very complex and parametrized movement.

Figure 6.4 shows the result of the bachelor thesis of Michael Eisele at Hochschule Kempten [48]. The AR scene is generated with Microsoft's second iteration of the Kinect sensor. Characters are robot bees sensing the user's motion from the skeleton tracking features of the Kinect. Depending on the user's gestures such as position and velocity of hands, the bees react and perform actions, for example, attacking the user, randomly flying around, or seeking virtual food. Next to the position of the user's extremities, they are also aware of the user's body shape and avoid collisions by applying appropriate 3D steering behaviors. In addition, the body shape being determined by the depth camera of the Kinect

Fig. 6.4 AI bees move in AR space using 3D steering behaviors. They attack the user, react upon arm position and motion, and respect space occupied by the user including correct occlusion effects. (After [48] with permission of Michael Eisele)

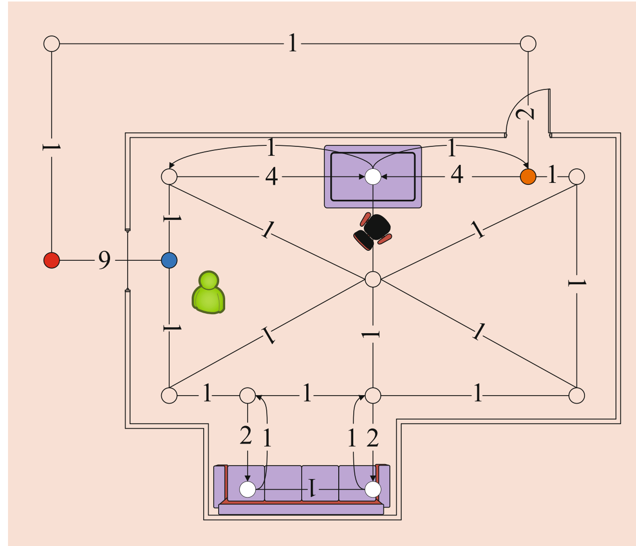


Fig. 6.5 Exemplary AR space with a navigation mesh. The blue node of the mesh is the closest one to the green NPC. Cost of edges in the graph represents arduousness of the route, e.g., jumping onto the desk is more difficult than walking on the ground. The red node is behind a portal to a virtual world behind the wall

allows for correct occlusion effects, i.e., bees behind the user's body will be occluded by reality.

In case movement shall be executed in a more predictable and controllable way to track and anticipate locations on the movement path, Game AI uses approaches known from graph theory and navigation techniques. This is favorable for movement in 2.5D space, i.e., the character moves on the ground while having the ability to jump over gaps or on platforms.

Basis of this approach is a so-called navigation mesh, which is a graph invisibly superimposed onto the NPC's living space. It hosts information about the moveable space, i.e., where can the NPC move and where movement is not possible. This includes, for example, steepness of a ramp, gaps between two objects, step heights, or narrowness of passages. Major game engines are shipped with this core functionality, e.g., Unity. Here a navigation map can be generated during the development process of the game taking all geometries of the NPC's living space into account to define a graph reflecting the movable area. More recently, such navigation maps can also be generated during runtime. This is essential for AR applications, since geometry of the AR space defining boundaries of the NPC's movement is not known beforehand. The space geometry may even change during the AR experience, when real objects are inserted into or removed from the scene. In an AR scenario, the basis of the navigation mesh generation is the spatial map, e.g., computed and regularly updated by the HoloLens in runtime. While this spatial map can serve as a navigation map as it is, the computation of the navigation map as a graph encoding rich information about movement space as described above provides many advantages.

Once the navigation map, the graph representing the moving space, has been created, characters are able to move from their current position (a 3D vector) to another position in space using *pathfinding algorithms*. The most popular algorithm to efficiently find the best (shortest and cheapest) path is the A* algorithm including its optimizations and variations [2]. A* evaluates costs of edges in the graph and heuristic information to find the shortest route from a starting node to an end node within the graph.

Figure 6.5 shows an exemplary navigation mesh for a room serving as an AR living space of an AR character (green character icon). The room has a magic window serving as a virtual portal to a virtual world behind the wall (red node). It further is equipped with a real working desk and a real couch. In case the character needs to move from his position (blue node) to the orange node next to the door, he has several options. Using Euclidian distance as costs for the edges of the graph, the path leading over the table would be the shortest route. However, costs of this route are higher, if we use edge weights as a measure of arduousness to take a connection between nodes. With the shown configuration of the navigation mesh, the AR character would rather take the route from the blue node to the node on the upper left corner of the room, to the center of the room, to the upper right corner, and then to the door node. Although the way from the blue node to the red node behind the virtual portal shows the shortest distance, it has a high weight. This may convince the character most of the time to use the open door to leave the room and appear later behind the portal. However, in some situations it would decide to take the direct connection through the virtual portal. The costs of the edges may also be a combination of area costs and Euclidian distance.

Usually, the character is not positioned exactly at one of the nodes of the navigation mesh. In addition, the destinations of the movement should not be restricted to nodes of the graph. To accomplish flexible movement, pathfinding is combined with steering behaviors:

1. NPC is positioned at location A_P and decides to move to a point B_P in space.
2. The navigation mesh is analyzed to find the closest node A_M on the map to A_P in AR space.
3. The navigation mesh is analyzed to find the closest node B_M on the map to B_P in AR space.
4. A* computes the PATH (an ordered set of 3D vectors/nodes) from A_M to B_M, i.e., A_M is the first node in PATH and B_M is the last node.
5. NPC uses the steering behavior SEEK to first move from A_P to A_M and then from one node to the next node in PATH.
6. Once the NPC has reached B_M, it uses the steering behavior ARRIVEs to reach B_P.

Figures 6.1 and 6.2 show pathfinding in AR space at work. The red lines visualize the path calculated with the A* algorithm and the navigation map being part of the Unity core functionality.

6.2.5 Reasoning

Reasoners let NPCs make decisions and induce a first level of intelligence. The type of reasoner determines the level of autonomy, i.e., how much the agent follows its own schedule while being responsive when appropriate. It is possible to rank reasoner algorithms with respect to their level of author control and autonomy as shown in Fig. 6.6. Algorithms with greater author control provide game designers with more options to script the agent's behavior. The agent will naturally have more specialized roles, e.g., providing standardized marketing information or telling a background story of a game. This kind of character has a limited living space and restricted responsiveness, and its behavior is rather predictable. On the other end of the ranking scale, algorithms show greater responsiveness to changing parameters of the character's living environment. Configuration of these algorithms is rather complex and is difficult to be managed by a game designer usually having less programming skills. Planning algorithms even combine decision-making and strategy. They create a plan, which increases the autonomy level and makes decision-making even more complex. The highest level of autonomy can be reached by the integration of learning algorithms. Most of these algorithms require a learning phase resulting in large parameter sets that are difficult to be interpreted. In other words, on the left side of the scale, the game designer decides, which knowledge serves as decision factors for a certain option. On the right part of the scale, learning algorithms take over the selection of knowledge and even generate secondary knowledge to make decisions.

Usually a character has multiple reasoners working in parallel and being responsible for different behavior levels. For instance, there can be a reasoner coordinating head and eye motion to allow for gazing at the user or a position in AR space. Another reasoner can be responsible for hand and arm gestures. A third reasoner may care about the movement in AR space. One may also consider a specialized high-level reasoner that processes social signals of the user to control designated facial expressions, gestures, etc.

All reasoners follow the so-called Sense-Think-Act cycle that is executed in every update frame as shown in Fig. 6.7.

When sensing, the character's AI architecture collects and updates required knowledge. The source of this knowledge can be from sensory input such as perceiving obstacles or objects of interest visually or acoustically within the NPC's perceptual range. It can consist of hardwired parameters such as internal states of the character or global states of the AR application. In addition, knowledge can be selected from so-called blackboards, a data structure that can share informa-

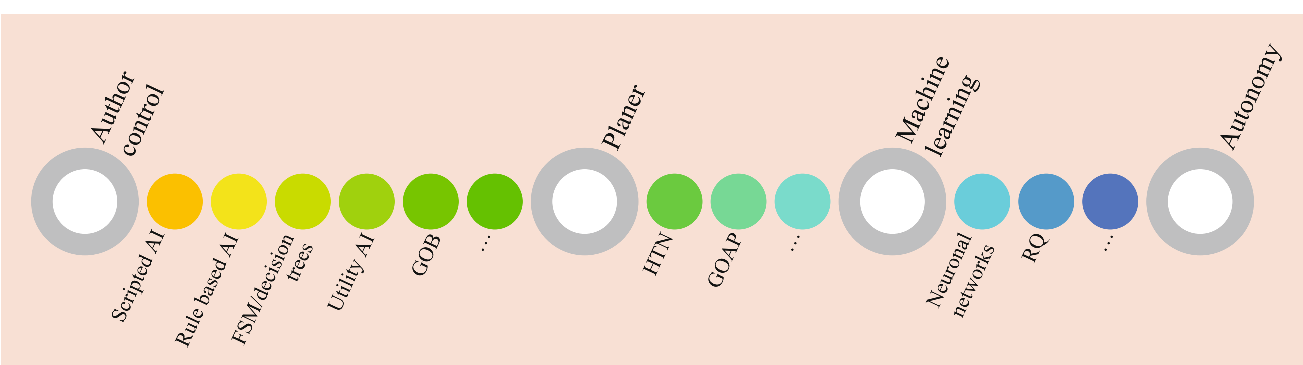


Fig. 6.6 Author control versus autonomy

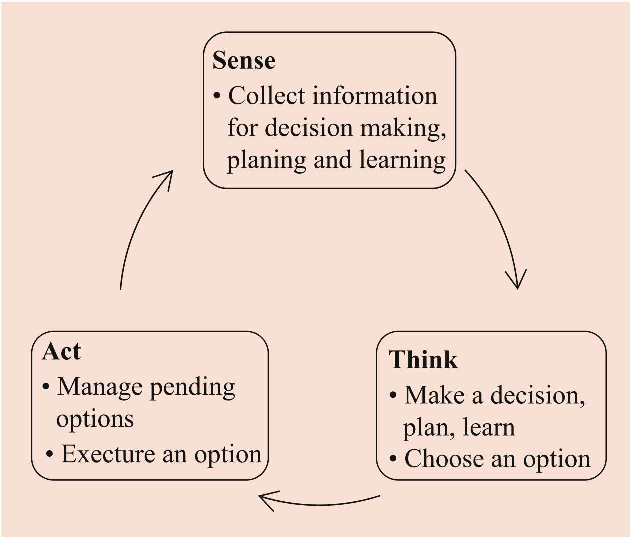


Fig. 6.7 Sense-Think-Act cycle

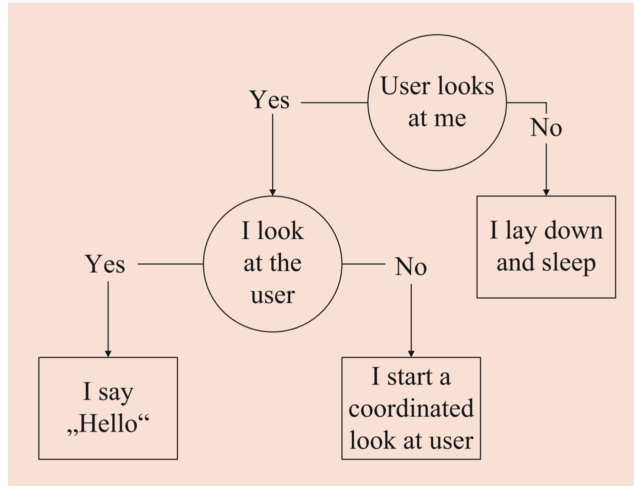


Fig. 6.8 A simple decision tree to trigger social reactions

tion among multiple AI characters or multiple reasoners of one agent.

Thinking evaluates knowledge to make decisions using a reasoner algorithm. It also includes making strategic plans to change the current state of the world to a desired goal state. In addition, learning algorithms may use the knowledge to adapt their parameter sets. When having accomplished the thinking process, reasoners will output one or a series of options that the character needs to execute.

The options selected by the reasoners need to be managed and processed to implement the Act process of the cycle. A simple architecture arranges incoming options in a to-do list and processes them sequentially. A more complex architecture uses an *option manager* to administrate pending options from various reasoners of the same NPC and differentiates varying types of options. For instance, a certain type of options may not accept postponements and need to stop and replace a current active option. Another type of options may cause a temporary interruption of the current running option and as for execution once it was selected by one of the reasoners. A good example for this option type is a hit reaction in battle scenarios for this type of options. Once the NPC senses a strike of the player or another NPC while fighting, it needs to execute an adequate animation, e.g., stumbling and/or yelling. After this animation has finished, the character continues fighting. Options having been selected by reasoners may not wait forever to be executed. They usually have a timestamp telling the option manager that the option is not relevant anymore since the world state of the character has changed meanwhile. In addition, options may be equipped with a cool down mechanism that helps avoid oscillation effects.

The following sections introduce a small selection of the most popular reasoners in Game AI. For a more complete overview of reasoners, we refer to the books [2–5].

Decision Trees

Decision trees are usually executed once every frame and realize fast and simple decisions. They usually show a simple binary tree structure that can be easily understood and edited by a designer having only limited programming skills.

The algorithm starts evaluating the decision tree in the root node. Every internal node including the root node of the tree is a consideration that evaluates a piece of knowledge and usually outputs a true or false value indicating which node/consideration needs to be evaluated next. Leaf nodes of decision trees are options. Once the tree traversal reaches a leaf, the reasoner terminates, and the decision tree reasoner returns the leaf option.

Figure 6.8 shows an example that can be implemented for a responsive AR assistant showing social awareness. The thinking process of the character starts with the root node hosting a consideration that evaluates the gaze direction of the player. Relevant knowledge can be *hardwired* in the code and consists a reference to the 3D vector in combination of a maximal distance value and an angle storing the player's field of view. A consideration may evaluate whether the NPC's position in space is within the player's field of view. In case the character is not within the player's perceptual range, the reasoners output an option to trigger an idle animation, and the thinking process is over. Otherwise, the character evaluates its own state. In case the character already looks at the user, it decides to say "Hello." However, if there is no eye contact between user and the player in the current situation, an animation performing a coordinated look at the player needs to be triggered.

Decision trees can be reused to compose more complex trees. As a design rule, the tree should be balanced, if possible, and more likely decisions should have shorter routes through the tree structure. Also cost-intensive considerations should be executed by nodes of lower levels in the tree structure . In order to avoid predictable behavior, special random nodes can be inserted that randomly choose, e.g., with certain probability one of the child nodes.

Finite State Machines

Outside the game engineering community, *finite state machines (FSMs)* are necessarily not considered as AI algorithms. However, in the gaming world FSMs have a long tradition, to control NPCs. Like decision trees, the structure of FSMs can be easily understood by game designers to design character behaviors without having a deeper technical background.

FSMs usually consist of a set of states and transitions between the states. At any time, an NPC being directed by a FSM has one current state. The current state can be left via predefined transition to enter another state, if certain conditions are fulfilled. These conditions are implemented as considerations, evaluating a piece of knowledge being relevant for the current state.

Figure 6.9 shows an exemplary FSM of a character behavior in a life simulation scenario. When the application starts, the character enters an initial state. The circle symbol connected with an arrow to this initial state *Working* indicates this initialization mechanism. States manage three types of actions. The first type is passed to the Act process, when the character enters the state. Following the FSM example, the character may trigger an option to say: "Good morning dear colleague!." The second type of option is delivered by the FSM reasoner, while the character stays in the state. For instance, the NPC may repetitively output an animation option mimicking its designated daily work. The third type of options attached to states is triggered once the NPC leaves the current state. In case the NPC changes to the state *Eating*, it plays an audio source: "I'm so hungry, I could eat a horse!." The FSM evaluates exclusively conditions of the current state that may result in a state change, a so-called transition. If the NPC is in the *Working* state, it would evaluate the global knowledge *time* with a consideration to check if it is time to eat. The second consideration of the state *Working* evaluates an internal value of the character managing its constantly increasing need to use the restroom. Once this condition is fulfilled, the NPC would change to the state *Rest Room*. Like states also transitions can manage multiple types of option types to design a fine-grained behavior for the event of the state change. For example, once it is time to eat, the transition activates a so-called trigger option to indicate that the condition for the transition has been met. This option animates the character, for example, to look at a clock or

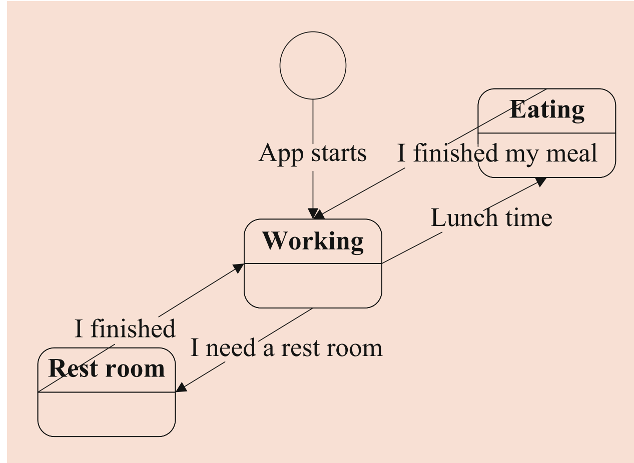


Fig. 6.9 A daily routine of an NPC realized with a FSM

the character turns off the machine he is working with. After the trigger option, the FSM reasoner returns a so-called fire option, visualizing that the character has started to change its state. For instance, the character begins to move from its working space to the food place.

The thinking process of decision trees may result in any option being assigned to this reasoner. In FSMs, only a subset of available options of the FSM reasoner can be selected depending on the current state. Like decision trees, also FSM designs can serve as a module to construct more complex FSM structures.

In *hierarchical FSMs*, an independent working FSM can be treated as a state of a higher-level FSM as shown in Fig. 6.10. The NPC has a standard behavior that serves as a lower-level state of a hierarchical state machine. If there are no extraordinary events, the NPC follows the rules of this standard behavior. However, special events may require exceptional behavior patterns. In this case, an alarm signal would lead to a transition to an *Escape* state. Regardless which state of the standard behavior has been the current state, the character would stop its actions and escape.

In order to handle the return to the standard behavior once the alarm signal has been stopped, one may follow two different strategies. The simplest approach would lead the NPC to the initial state of the standard behavior, here the *Working* state. A more plausible approach uses a state reference, remembering the last state once the NPC left the standard behavior. In Fig. 6.10, this so-called history state is shown as a dotted circle labeled with *H**.

Goal-Oriented Behavior

Goal-oriented behavior (GOB) is intensively used in simulation games, for example, to control daily routines of NPCs. Characters own individual sets of so-called needs or goals which are usually float values in the range of [0,1]. 0 means that the need is not pressing at all and will not influence the

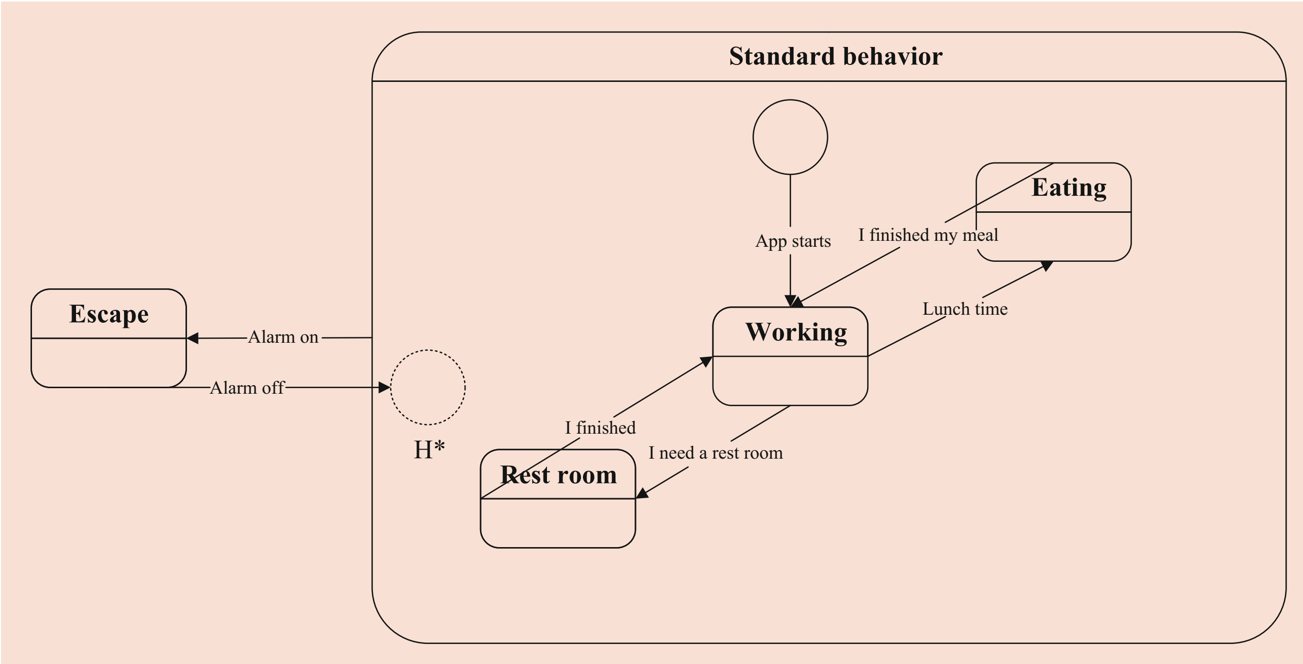


Fig. 6.10 With a hierarchical FSM, a daily routine can be interrupted with an extraordinary event requiring special behavior patterns

behavior of the character. On the other hand, a value of 1 means that the want is pressing and asks urgently for supply. Pressing needs are getting satisfied with available actions that effectively reduce the need value. For this reason, a GOB reasoner searches and returns actions that help to reduce painful needs.

Needs and their values can express emotions such as anger or happiness. They may also encode the social attitude towards the player/user. For instance, an AR assistant senses the time the user needs to finish a task or the number of failures. This information can be mapped to a special need encoding the NPC's willingness to offer help. Needs can also reflect physical properties of the character such as health or damage. One may even use psychological models such as the Maslow's hierarchy of needs to prioritize personal needs as shown in Fig. 6.11.

Values of these goals change over time. For example, the need for sleep will grow over daytime, and in the evening, it will turn into a pressing need. Alternatively, the value of needs is influenced by the NPC's actions and respectively its decisions what to do next. For instance, if the character performs an action to drink when the need *thirst* is pressing, this action will also influence its need *hunger*. Usually, a threshold is defined for each goal telling that the goal is pressing once its value excides the threshold.

Pressing needs look for healing options, i.e., options that offer satisfaction and reduce the need value. These options are usually provided by the game world or in this case the AR space. For example, a virtual fountain may offer an option to drink. In AR space, also real objects can provide options to supply wants, e.g., switching on a real lamp that has the function to communicate its state to the AR application can offer warmth when the character feels coldness.

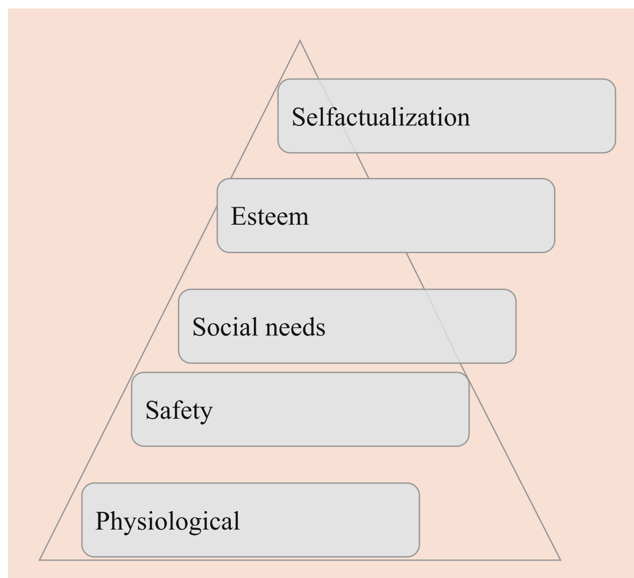


Fig. 6.11 Maslow's hierarchy of needs to design GOB needs

The simplest version of the GOB algorithm works as follows:

1. Update the float values of all needs that depend on a time function, e.g., hunger increases over time.

2. Select a pressing need, the *urgent_need,* from the list of needs. In case there are several pressing needs, choose one of them randomly.
3. Among all available options, select the option *best_option* that best satisfies *urgent_need*, i.e., results in the strongest decrease of the float value of *urgent_need.*
4. Return *best_option* to be handled by the option manager.
5. Once the *best_option* is executed and then terminated, all related need values are updated.

A more advanced version of the GOB algorithm takes the overall discontent into account. The overall discontent is the sum of the float values of all needs. The objective of this version of the algorithm is to select an option that has the highest impact on the overall discontent, i.e., reduces this sum to a minimum. Here, the side effects of options on various needs are considered. For instance, an option letting the NPC eat will first reduce the need *hunger*. In addition, also the needs thirst and sleep will be affected, since eating may make the character thirsty and tired. This variation of GOB works as follows:

1. Update the float values of all needs that depend on a time function, e.g., hunger increases over time
2. For all available options, compute their overall effect *discontent_per_option* on each need value. In other words, this piece of code checks what would happen to the overall discontent, if we chose a certain option.
3. Return the *best_option* with the lowest *discontent_per_option* to be handled by the option manager.
4. Once the *best_option* is executed and then terminated, all related need values are updated.

Figure 6.12 visualizes a very simple and idealized example of how the algorithm influences needs overtime. The character possesses three needs having a linear function over time, i.e., these needs automatically increase linearly with time. Alternatively, more complex functions such as a sigmoid function or a function with control points and interpolation between those control points may increase the realism of the simulation. In this exemplary setup, all needs use a threshold at 0.8, i.e., when the float value quantitatively representing the need exceeds the threshold, the need becomes urgent and looks for an option that promises satisfaction. The progress of their float values in the range [0;1] is visualized with lines and stacked area charts. This type of charts is recommended to check plausibility of evolving needs.

In addition, the NPC has three options that influence needs once they have been chosen to be executed. In this scenario, all options satisfy only one of the three needs. In this example, their effect reduces the float value to a minimum (=0). Usually, an option has a major effect on one of the needs and several positive and negative side effects on the other needs. In addition, the execution of available options in this idealized scenario does not require time. In a more realistic configuration of GOB, an option satisfying hunger has a varying and sometimes uncertain duration. First the character needs time to walk to the food place, prepare food via animation, and then eat the food. The uncertainty of the duration is primarily induced by the option manager, e.g., when other options are still pending to be executed and have a higher priority as the last option having been selected by the GOB reasoner.

Figure 6.13 shows a stacked area chart visualizing the evolution of three needs. The sum of the need values at each point in time shows the overall discontent in every situation of the NPC's live time.

Utility AI

Utility AI is modern and robust reasoner inspired by the utility theory that has been introduced and used in economics long before video games and augmented reality applications had been invented [5].

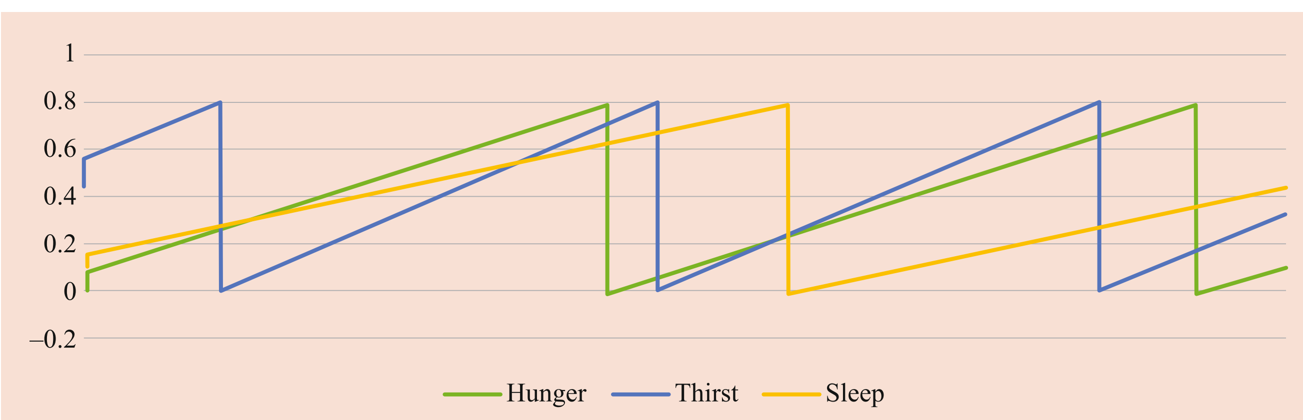


Fig. 6.12 Evolution of three needs over time. Once they become pressing needs, a designated option cares about satisfaction

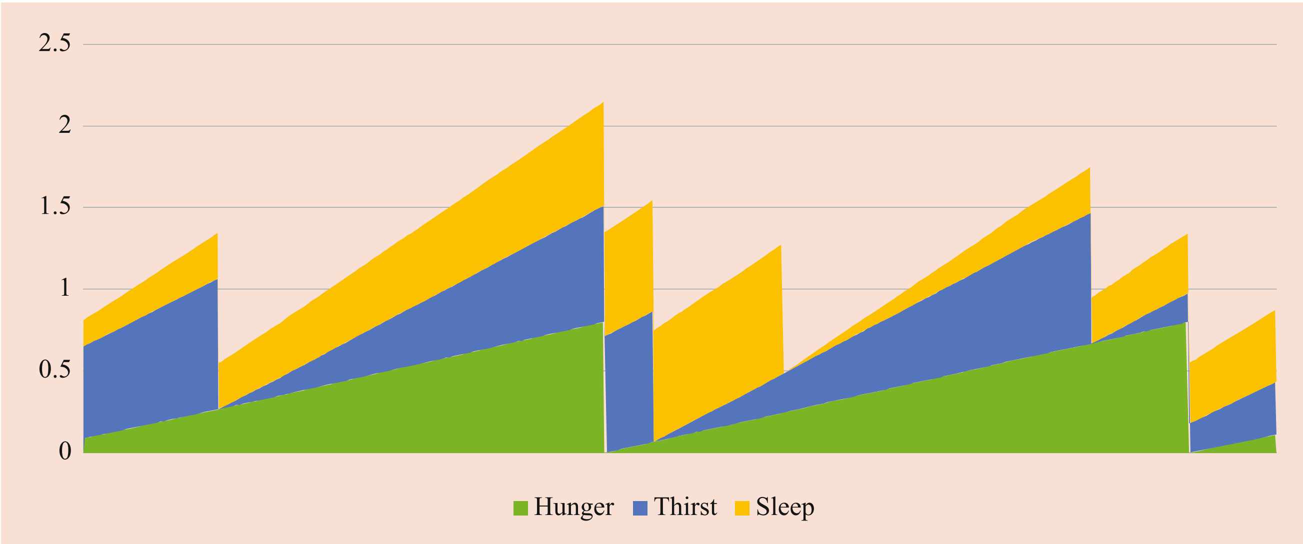


Fig. 6.13 Stacked area charts are helpful to visualize the overall discontent of the NPC

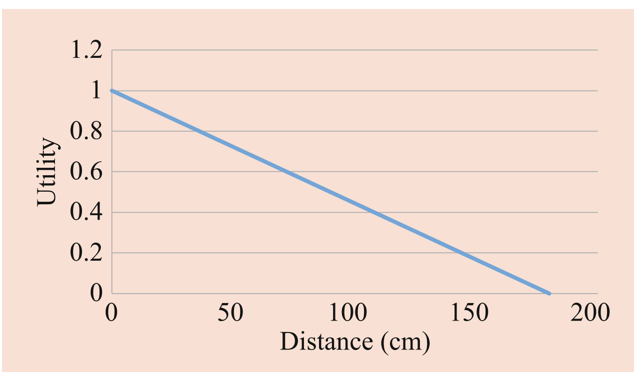


Fig. 6.14 This linear utility function converts a distance value to a utility value

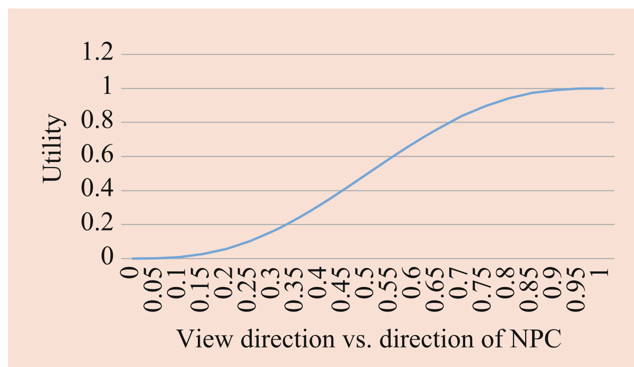


Fig. 6.15 A s-curve to map gaze information to utility values

Utility reasoners are designed to combine several decision factors being relevant for one specific option. All decision factors are designed as utility consideration functions $U(x)$ mapping knowledge and its evaluation to the continuous interval [0;1] resulting in expected utility values (EU), while 0 means that the utility of a decision factor is rather useless and 1 means that the utility of this factor is very helpful.

In an exemplary AR application, an assisting character needs to decide if it should execute an option offering help to accomplish a maintenance task. Decision factors for this option could be the relative distance to the user (DISTANCE; see Fig. 6.14), view direction of the user indicating that he may look for help (GAZE), and the number of unsuccessful attempts to finish a subtask (TRAILS).

Another option moves the NPC to a standard location in space and lets it wait for further instructions of the user. A relevant decision factor could be the number of helping offers that had been rejected by the user in the past indicating that this user rather prefers to manage problems on his own.

With respect to the first option and its decision factors, DISTANCE may use a linear utility consideration function equipped with a parameter m, which encodes the maximum acceptable distance that would evoke a reaction of the NPC.

$$U(x) = 1 - \frac{x}{m}$$

The decision factor GAZE in this example maps knowledge to utility space with a more complex function, an s-curve (see Fig. 6.15). The following function requires input of the interval [0,1]. A sophisticated approach would be the comparison of the direction vector from the user's position to the NPC's position and the user's gaze vector (view direction of the user). The absolute value of the angle $|\alpha|$ between the two vectors is then mapped to the input space.

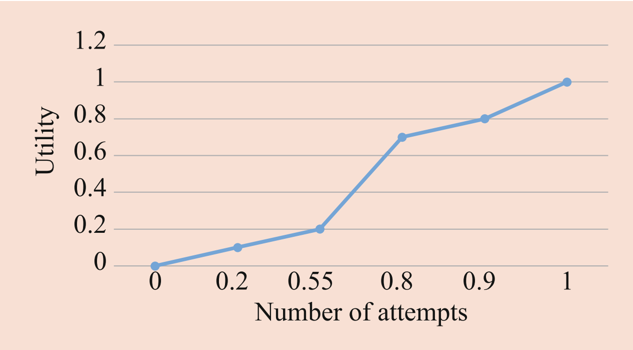


Fig. 6.16 A step function to convert number of attempts (% of maximum acceptable attempts) to utility values

The exponentiation $U(x)^k$ of this utility function with the parameter k (in this example $k = 1$) transforms its output to meet various configuration scenarios (Fig. 6.15).

$$U(x) = 10 * x^3 - 15 * x^4 + 6 * x^5$$
$$= x^3 (10 + x * (-15 + 6 * x))$$

In many situations, however, a rather simple step function using a set of control points and a linear interpolation between those control points may be more appropriate to the design of the decision behavior. In our example, TRAILS uses this step function as shown in Fig. 6.16.

Configuration of the utility functions for each decision factors as well as the definition of a set of relevant decision factors for a special option is usually part of the design process of a game or AR application. We recommend using and combining the following data structures to setup the utility reasoner:

- Each *option* is linked to a data structure *decision*.
- A *decision* is also linked to a set of decision factors.
- Each decision factor is a *utility consideration function* evaluating *knowledge*.

Decisions can share decision factors evaluating knowledge in the same manner. The same knowledge may also have different meanings for different decisions; hence it is evaluated with different utility consideration functions.

Each EU_{jl} is a result of an individual utility considerations function $U(x)$ encoding a decision factors l. A vector of these factors forms the basis of a decision j and must be combined to a final utility score (FU_j). Evaluation of the FU_j values leads to the selection of the best option in the current situation. The arithmetic mean value of all FU_j is a typical approach to compute this combination; however, there are further techniques described by the literature [4].

$$FU_j = \frac{\sum_{l=1}^{x} EU_{jl}}{x}$$

The utility reasoner returns in this case the option j linked to the decision j with the highest FU_j value, indicating that this option is most useful to manage the current situation.

6.3 Conclusion

This chapter addresses the integration of virtual characters in AR worlds. We start with a review of literature having investigated this subject. It is followed by a guideline on designing interactive AR agents and proposing three layers of an AI architecture consisting of an appearance layer, a movement layer, and a reasoning layer. We take a closer look at the appearance layer with the focus on human-human verbal and non-verbal communication to derive a list of design rules to create characters with high social presence in AR spaces. We further discuss a set of basic data structures and mechanisms that allow for sharing information among these layers to create a powerful AI system. We also introduce a selection of algorithms and techniques well known in the Game AI community to be used for AR purposes.

In a complex Game AI architecture, characters can be equipped with two additional layers, which would cover first strategy and second learning. The objective of strategy is to create more complex decisions including a plan of sequentially executed options to reach a certain goal. Learning allows the character to adapt to changing situations in the AR world, e.g., to explore its living environment to efficiently move from one place to another. It may also concern gaining knowledge about the user/player performance to adjust helping instructions, balancing the level of difficulty of a training experience, or even changing its social behavior and attitude according to the player's behavior. It may also be used to simply optimize certain parameters for its own behavior, which is important in AR living spaces that may change over runtime and cannot be planed and evaluated to the same extend as in entirely virtual worlds. However, in most scenarios, reasoning is powerful enough to create a believable artificial creature playing its designated role in an authentic manner.

References

1. Sutherland, I.E.: The ultimate display. In: Proceedings of the IFIP Congress, pp. 506–508 (1965)
2. Millington, I., Funge, J.: Artificial Intelligence for Games. CRC Press, Boca Raton (2009)
3. Rabin, S.: Game AI Pro 3: Collected Wisdom of Game AI Professionals. CRC Press, Boca Raton (2017)
4. Rabin, S.: Game AI Pro: Collected Wisdom of Game AI Professionals. CRC Press, Boca Raton (2014)
5. Rabin, S.: Game AI Pro 2: Collected Wisdom of Game AI Professionals. CRC Press, Boca Raton (2015)
6. Bates, J.: The role of emotion in believable agents. Commun. ACM. **37**(7), 122–125 (1994). https://doi.org/10.1145/176789.176803

6

7. Oh, C.S., Bailenson, J.N., Welch, G.F.: A systematic review of social presence: definition, antecedents, and implications. Front. Robot. AI. **5**, 483 (2018). https://doi.org/10.3389/frobt.2018.00114
8. Kim, K., Bruder, G., Welch, G.: Blowing in the wind: increasing copresence with a virtual human via airflow influence in augmented reality. In: ICAT-EGVE 2018 – International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments (2018)
9. Kim, K., Schubert, R., Hochreiter, J., Bruder, G., Welch, G.: Blowing in the wind: increasing social presence with a virtual human via environmental airflow interaction in mixed reality. Comput. Graph. **83**, 23–32 (2019). https://doi.org/10.1016/j.cag.2019.06.006
10. Hartholt, A., Mozgai, S., Fast, E., Liewer, M., Reilly, A., Whitcup, W., Rizzo, A.: Virtual humans in augmented reality. In: Proceedings of the 7th International Conference on Human-Agent Interaction – HAI '19, pp. 205–207. Kyoto (2019)
11. Anabuki, M., Kakuta, H., Yamamoto, H., Tamura, H.: Welbo. In: CHI '00 Extended Abstracts on Human Factors in Computing Systems – CHI '00, p. 10. The Hague (2000)
12. Barakonyi, I., Schmalstieg, D.: Augmented reality agents in the development pipeline of computer entertainment. In: Kishino, F., Kitamura, Y., Kato, H., Nagata, N. (eds.) Lecture Notes in Computer Science, Entertainment Computing – ICEC 2005, pp. 345–356. Springer Berlin Heidelberg, Berlin/Heidelberg (2005)
13. Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., Welch, G.F.: Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In: 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 105–114. Munich, Oct 2018 – Oct 2018
14. Norouzi, N., Kim, K., Hochreiter, J., Lee, M., Daher, S., Bruder, G., Welch, G.: A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents – IVA '18, pp. 17–22. Sydney (2018)
15. Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The ALIVE system: full-body interaction with autonomous agents. In: Proceedings Computer Animation '95, pp. 11–18. Geneva, Apr 1995
16. Cassell, J., Ananny, M., Basu, A., Bickmore, T., Chong, P., Mellis, D., Ryokai, K., Smith, J., Vilhjalmsson, H., Yan, H.: Shared reality: physical collaboration with a virtual peer. In: CHI '00 Extended Abstracts on Human Factors in Computing Systems – CHI '00, pp. 259–260. The Hague (2000)
17. Cheok, A.D., Weihua, W., Yang, X., Prince, S., Wan, F.S., Billinghurst, M., Kato, H.: Interactive theatre experience in embodied + wearable mixed reality space. In: Proceedings of the International Symposium on Mixed and Augmented Reality, pp. 59–317. Darmstadt, Sept 2002
18. Balcisoy, S., Kallmann, M., Torre, R., Fua, P., Thalmann, D.: Interaction techniques with virtual humans in mixed environments. In: 5th IEEE EMBS International Summer School on Biomedical Imaging, p. 6. Berder Island, June 2002
19. Cavazza, M., Charles, F., Mead, S.J., Martin, O., Marichal, X., Nandi, A.: Multimodal acting in mixed reality interactive storytelling. IEEE Multimedia. **11**(3), 30–39 (2004). https://doi.org/10.1109/MMUL.2004.11
20. Magnenat-Thalmann, N., Papagiannakis, G.: Virtual worlds and augmented reality in cultural heritage applications. In: Recording, Modeling and Visualization of Cultural Heritage, pp. 419–430 (2005)
21. Wagner, D., Billinghurst, M., Schmalstieg, D.: How real should virtual characters be?. In: Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology – ACE '06, p. 57. Hollywood (2006)
22. Miller, T., Weising, G.: Method for an Augmented Reality Character to Maintain and Exhibit Awareness of an Observer. US2018185757 (A1), US US201815906448 20180227, 5 July 2018
23. Tomi, A.B., Rambli, D.R.A.: An interactive mobile augmented reality magical playbook: learning number with the thirsty crow. Procedia Comp. Sci. **25**, 123–130 (2013). https://doi.org/10.1016/j.procs.2013.11.015
24. Arroyo-Palacios, J., Marks, R.: Believable virtual characters for mixed reality. In: 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 121–123. Nantes, Oct 2017 – Oct 2017
25. Ishii, H., Lee, N., Natkin, S., Tsushima, K. (eds.): Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology – ACE '06. ACM Press, New York (2006)
26. Kim, K., Maloney, D., Bruder, G., Bailenson, J.N., Welch, G.F.: The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR. Comp. Anim. Virtual Worlds. **28**(3–4), e1771 (2017). https://doi.org/10.1002/cav.1771
27. Miller, M.R., Jun, H., Herrera, F., Yu Villa, J., Welch, G., Bailenson, J.N.: Social interaction in augmented reality. PLoS One. **14**(5), e0216290 (2019). https://doi.org/10.1371/journal.pone.0216290
28. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE TRANSACTIONS on Information and Systems **77**(12), 1321–1329 (1994)
29. Stapleton, C., Hughes, C., Moshell, M., Micikevicius, P., Altman, M.: Applying mixed reality to entertainment. Computer. **35**(12), 122–124 (2002). https://doi.org/10.1109/MC.2002.1106186
30. Stapleton, C.B., Hughes, C.E.: Making memories of a lifetime. In: Emerging Technologies of Augmented Reality: Interfaces and Design. IGI Global, S. 329–351 (2007). https://books.google.de/books/about/Emerging_Technologies_of_Augmented_Reali.html?id=ItLhfvNGTssC&redir_esc=y
31. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. **6**(4), 355–385 (1997). https://doi.org/10.1162/pres.1997.6.4.355
32. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Grap. Appl. **21**(6), 34–47 (2001). https://doi.org/10.1109/38.963459
33. Bichlmeier, C.: Immersive, interactive and contextual in-situ visualization for medical applications. Dissertation, Chair for Computer Aided Medical Procedures & Augmented Reality, Technische Universität München, München (2010). Available: http://mediatum.ub.tum.de/node?id=977862
34. Agusanto, K., Li, L., Chuangui, Z., Sing, N.W.: Photorealistic rendering for augmented reality using environment illumination. In: The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings, pp. 208–216. Tokyo, Oct 2003
35. Scherer, L.: Lebenssimulation und Aufbauspiel in einer Augmented Reality Umgebung. Bachelor's thesis, Faculty of Computer Science, Kempten University of Applied Sciences, Germany, Kempten (2019). Available: http://games.hs-kempten.de/lebens-simulation-und-aufbauspiel-in-einer-augmented-reality-umgebung/
36. Röhner, J., Schütz, A.: Psychologie der Kommunikation. Springer Fachmedien Wiesbaden, Wiesbaden (2016)
37. Cialdini, R.B.: Influence: Science and Practice, 5th edn. Pearson Education, Boston (2009)
38. Petty, R.E., Cacioppo, J.T. (eds.): Communication and Persuasion. Springer New York, New York (1986)
39. Pacchierotti, C., Sinclair, S., Solazzi, M., Frisoli, A., Hayward, V., Prattichizzo, D.: Wearable haptic systems for the fingertip and the hand: taxonomy, review, and perspectives. IEEE Trans. Haptics. **10**(4), 580–600 (2017). https://doi.org/10.1109/TOH.2017.2689006
40. Furukawa, T., Niwa, D., Hirano, Y., Hanamitsu, N., Kamiyama, Y., Nii, H., Krekoukiotis, C., Minamizawa, K., Noda, A., Yamada,

J., Kitamura, K.: Designing a full-body customizable haptic interface using two-dimensional signal transmission. In: ACM SIGGRAPH 2019 Posters on – SIGGRAPH '19, pp. 1–2. Los Angeles (2019)
41. Hall, E.T.: The Hidden Dimension. Anchor Books, New York (1990). Available: http://www.loc.gov/catdir/description/random043/90034870.html
42. Hargie, O., Dickson, D.: Skilled Interpersonal Communication: Research, Theory, and Practice, 4th edn. Routledge, London (2004)
43. Blascovich, J., Bailenson, J.: Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution, 1st edn. William Morrow, New York (2011)
44. Slater, M., Sadagic, A., Usoh, M., Schroeder, R.: Small-group behavior in a virtual and real environment: a comparative study. Presence Teleop. Virt. **9**(1), 37–51 (2000). https://doi.org/10.1162/105474600566600
45. Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M.: Equilibrium theory revisited: mutual gaze and personal space in virtual environments. Presence Teleop. Virt. **10**(6), 583–598 (2001). https://doi.org/10.1162/105474601753272844
46. Bonsch, A., Radke, S., Overath, H., Asche, L.M., Wendt, J., Vierjahn, T., Habel, U., Kuhlen, T.W.: Social VR: how personal space is affected by virtual agents' emotions. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 199–206. Reutlingen, Mar 2018 – Mar 2018
47. Reynolds, C.W.: Steering behaviors for autonomous characters. In Game developers conference (Vol. 1999, pp. 763–782) (1999)
48. Eisele, M.: Techniken zur Interaktion mit autonom gesteuerten Agenten NPCs in Augmented Reality Spielen. Bachelor's thesis, Faculty of Computer Science, Kempten University of Applied Sciences, Kempten, 16 Mar 2017. Available: http://games.hs-kempten.de/techniken-zur-interaktion-mit-autonom-gesteuerten-agenten-npcs-in-augmented-reality-spielen/

Dr. Christoph Bichlmeier is professor at the faculty of informatics, Hochschule Kempten, Germany. He teaches Augmented Reality and AI for Games in the bachelor study program "Game Engineering" and the master study program "Game Engineering and Visual Computing." In 2010, he completed his PhD at TU Munich, Germany with the subject "Immersive, Interactive and Contextual In-Situ Visualization for Medical Applications."

Privacy and Security Issues and Solutions for *Mixed Reality* Applications

7

Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne

Contents

7.1 The Mixed Reality Present 157
7.1.1 Overview on Mixed Reality Processing 158
7.1.2 Towards MR Mobility 159

7.2 Security and Privacy Risks with Mixed Reality 159
7.2.1 Risks with MR Data Processing 159
7.2.2 Mobility and Privacy 160

7.3 Protection Approaches for Mixed Reality 161
7.3.1 Input Protection 162
7.3.2 Output Protection 167
7.3.3 Protecting User Interactions 168
7.3.4 Device Protection 170
7.3.5 Open Research Challenges 171
7.3.6 Future Directions 173

7.4 Towards Everyday MR Services 175

References 179

Abstract

Mixed-reality (MR) technology development is now gaining momentum due to advances in computer vision, sensor fusion, and realistic display technologies. Despite this, most of the research and development has been focused on delivering the promise of MR; concerns on potential *security* and *privacy* risks are continuously being pointed out, and only a few are working on the privacy and security implications of the technology. We put into light these risks and look into the latest security and privacy work on MR. In this chapter, we present an exposition and categorization of the latest security and privacy work on MR.

J. A. De Guzman (✉)
University of the Philippines Diliman, College of Engineering Electrical and Electronics Engineering Institute, Quezon City, Philippines
e-mail: jaybie.de.guzman@eee.upd.edu.ph

K. Thilakarathna
The University of Sydney, School of Computer Science Faculty of Engineering, Sydney, NSW, Australia
e-mail: kanchana.thilakarathna@sydney.edu.au

A. Seneviratne
University of New South Wales, School of Electrical Engineering and Telecommunications, Sydney, NSW, Australia
e-mail: a.seneviratne@unsw.edu.au

Keywords

Augmented reality · Mixed reality · Virtual reality · Security · Privacy · Usability in security and privacy · Safety · Survey · Spatial data · 3D data · Computer vision · Collaborative interactions

7.1 The Mixed Reality Present

The future with *mixed reality* (MR) is now. Starting 2015 or even earlier, we have seen an increase in *augmented* (AR) and *mixed reality* (MR) devices and applications either using head-worn or hand-held form factors. With Apple's release of the ARKit SDK for the iOS during the 2018 Apple Worldwide Developers Conference, and, soon after, Google followed with ARCore for Android, the year 2019 saw a boom in hardware and software platforms that aim to bring MR to the mainstream. In this chapter, we will refer to augmented, virtual, and mixed reality, collectively, as mixed reality or MR.

The Current Focus with MR Majority of the research and development efforts over MR have primarily been on delivering the technology: specifically on improving the visual experience and the mobility of these services. Early surveys on MR have focused on categorizing or determining these necessary technologies. In 1994, a taxonomy for classifying MR displays and platforms based on the user interface and plotting these devices along a reality-virtuality continuum was presented [1]. Subsequent classifications were also presented based on the concepts of transportation (the extent to which the users are transported from their physical locality to a virtual or remote space), artificiality (the extent to which the

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_7

user space has been synthesized from a real physical space), and spatiality (the extent to which properties of natural space and movement are supported by the shared space) [2]. Consequently, a survey has identified and discussed the various early challenges in MR: specifically focusing on physical-virtual alignment, optical distortion, and object tracking [3]. It was complemented with a follow-up survey that focuses on the enabling technologies, interfacing, and visualization [4]. Another recent survey updated the challenges in MR systems to include performance, interactions, and mobility [5]. Again, While it is important to highlight these classifications and challenges, it is also equally important to highlight the security and privacy issues and solutions in MR.

Nonetheless, a few recent works have pointed out the ethical considerations [6] as well as value-sensitive design approaches that consider data ownership, privacy, secrecy, and integrity [7]. Another recent study has highlighted the potential perceptual and sensory threats that can arise from MR outputs, such as photosensitive epilepsy and motion-induced blindness [8]. Moreover, a relevant recent work has emphasized the *input*, *data access*, and *output* aspects for protection in MR [9]. In a recent survey on MR security and privacy approaches, we have extended these to include *interaction* and *device* protection as equally important aspects [10]. Now, in this chapter, we will present a re-categorization of these aspects and update the collection with the latest security and privacy work over MR.

7.1.1 Overview on Mixed Reality Processing

Figure 7.1 shows a generic pipeline for MR processing. MR platforms collect information through sensors that may capture various signals from the environment. MR has particularly put emphasis on visual information as the primary mode of information and experience. The captured visual or spatial information are processed to construct machine-understandable representations of these information. Common examples include a spatial mapping (or digital representation) of the environment, or a skeletal abstraction of user anatomy for gesture detection. This allows the MR platform to have an environmental understanding and detect the information-of-interest, which can be a structural feature

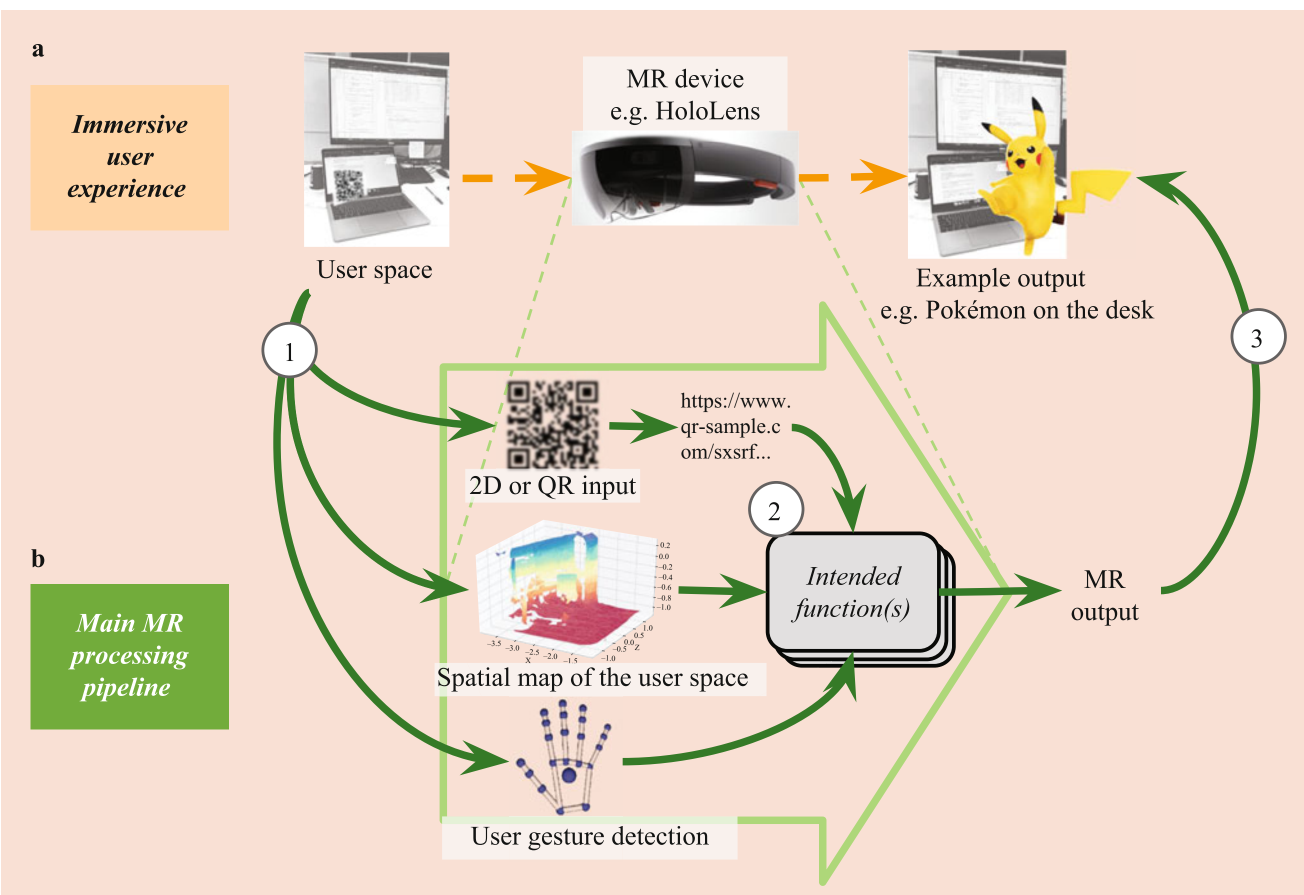


Fig. 7.1 Mixed Reality pipeline: (**a**) the immersive experience as viewed by the user; (**b**) the main processing pipeline of (1) detection or sensing, (2) transformation, and (3) rendering or actuation

(e.g., wall and floor), visual target (e.g. 2D QR code, or an actual object), or user gesture (e.g., hand or eye movement). Then, the MR application extracts the necessary information (e.g., object orientation in space), which are transformed to usable information (e.g., virtual features that can be anchored by virtual objects). Finally, the MR output is rendered on to the scene to make it seem like it inhabits the real world as it is viewed through a display. Spatial audio may also be added for an immersive MR experience.

7.1.2 Towards MR Mobility

Given that our mobile hand-held devices are primarily used for non-MR applications, it is only apt that new and dynamic ways of processing need to be designed to handle the processing demands of MR. More so that MR is one of these upcoming services that expected to further push the limits of hand-held devices and cloud computing. For example, Microsoft now offers remote rendering for the HoloLens through their Azure cloud service for rendering more complex 3D models, especially those used in engineering and architecture.

Various recent works on cloud- and edge-assisted MR processing include the reverse offloading processing back to the user devices as edge computing devices to alleviate cloud infrastructure cost and reduce latency [11, 12], utilizing GPU acceleration with cloud offloading [13], or using convolutional neural networks (or CNN) for object detection in the cloud [14]. While most of these works were mostly on 2D (i.e., image-based) data, it is also necessary to investigate actual 3D detection, such as those specifically used by current MR platforms as listed in Table 7.1, particularly looking at how auxiliary servers can be utilized for offloading 3D processing tasks. Perhaps, we can look towards techniques proposed in the area of vehicular AR for improving vehicular visibility collaboratively [15]. Likewise, we may also look towards efforts in wireless virtual reality (VR) for remote 3D rendering [16].

Ultimately, all these developments in mobile MR processing needs to be scrutinized in terms of the privacy and security risks that these developments may pose to the users. As we proceed in this chapter, we will present an overview of the various security and privacy risks that MR poses. Afterwards, we will present a review of the various approaches as well as a categorization of these approaches. Then, we present the remaining challenges that need to be addressed before finally concluding the chapter.

7.2 Security and Privacy Risks with Mixed Reality

Given these capabilities and the continuous development in machine vision and sensor fusion technologies, MR users face even greater risks as richer information can be gathered using a wide variety of methods. For example, the Microsoft HoloLens has a multitude of visual sensors: four (4) cameras for environment understanding, a separate depth sensor, two (2) infrared cameras for eye tracking, and another camera for view capture. Figure 7.2b shows an example of spatial map captured using HoloLens. These spatial maps are more memory-light than video despite containing near-accurate 3D digital representations of user spaces. Once these information are made available to applications and services, users may no longer have control over how these data are further utilized.

7.2.1 Risks with MR Data Processing

Despite these capabilities being seemingly necessary in delivering the promise of MR, not all MR functionalities or services require extremely rich information. Privacy concerns are further exacerbated by recent advances, such as near-real-time visual object detection using machine learning, which enables inference beyond the intended functionality [17]. For example, visual sensors in the MR device can subtly capture images and video without the knowledge of those around the user. This violates bystander privacy or the unauthorized capture of information about the other users, or so-called bystanders. Moreover, it has been demonstrated how easy it is to use a simple facial recognition algorithm to match live-captured photos with publicly available photos online (from online social networks such as Facebook) and extract personal information such as names and social security numbers [18]. Various endeavors have highlighted these risks over captured visual data, and likewise, various protection mechanisms have been posed.

Risks with 3D Spatial Data However, it is not only visual data that poses risks but also the spatial maps that provide the necessary environment understanding to MR platforms. This

Table 7.1 API class or command handling 2D and 3D detection

Software platform	2D	3D	Shareable?
Google's ARCore	*AugmentedImage*	*Trackables*	Yes, via *CloudAnchors*
Apple's ARKit	*ARReferenceImage*	*ARWorldTracking*	Yes, via *ARAnchors*
Window's MR API 2	No native support from API but can use any third party library or API that can run on Windows	*SpatialMapping*	Yes, via *SpatialAnchors*

Fig. 7.2 Comparison of different representations of the (**a**) physical environment: (**b**) the captured spatial map, which is stored as an unordered list of 3D points and usually accompanied by triangle mesh information to represent surfaces; (**c**) the rendered digital map with a color texture extracted from image captures to create an almost accurate copy of the physical space. (**a**) Real representation. (**b**) 3D digital representation. (**c**) Rendered reconstructed representation

capability further poses unforeseen privacy risks to users. Once these captured 3D maps have been revealed to untrusted parties, potentially sensitive spatial information about the user's spaces is disclosed. Adversaries can vary from a benign background service that delivers unsolicited advertisements based on the objects detected from the user's surroundings to malevolent burglars who are able to map the user's house and, perhaps, the locations and geometry of specific objects in their house based on the released 3D data. Furthermore, turning off GPS tracking for location privacy may no longer be sufficient once the user starts using MR applications that can expose their locations through the 3D and visual data that are exposed. For example, Google has unveiled a *Visual Positioning Service* (or VPS) using visual and 3D data to locate users—an offshoot of Project Tango—during their 2018 I/O keynote event.

User Perception of 3D Spatial Data With images and video, what the "machine sees" is practically what the "user sees," and a great deal of privacy work have been done on these data forms. Contrariwise, in most MR platforms, the experience is exported as visual data (e.g., objects augmented on the user's view), while the 3D nature of the captured spatial data is not exposed to the user: what the machine sees is different—arguably, even more—from what the user sees. That is, the digital representation of the physical world that the machine sees and understands is not exposed to the user. This inherent perceptual difference creates a latency from user perception and, perhaps, affects the lack of user-perceived sensitivity over the captured spatial information.

Moreover, no privacy preservation is currently applied before providing these spatial data to third-party applications. Aside from the spatial structural information, the mapping can also include 3D maps of objects within the space. Figure 7.2b shows that the digital representation not only accurately models the real representation but also includes information regarding the orientation of objects in the user environment at the time of interaction. Even the slightest change in the orientation will result in a different 3D point data, allowing the MR machine to detect such changes that otherwise would have been unnoticeable to a human eye—hence 'seeing' more than human perception.

Since users cannot readily understand how the machine sees in 3D, we speculate that users might inherently not bother with the sensitive information that 3D MR data captures. A recent study explored users' concerns in a multi-user MR environment where they highlighted the lack of concern from participants about how the technology can be maliciously used by other users and applications [19]. However, further study on the differences of user perception between 2D and 3D data representations as well as comparison of privacy concerns between the two need to be conducted to further corroborate our previous speculation.

Furthermore, despite the varying underlying vision processing algorithms employed, most of these MR platforms directly operate on these 3D spatial maps and offer cross-platform interoperability. Table 7.1 lists three APIs or SDKs for three popular MR platforms. It shows how each MR software platform handles, or which command or class handles, 2D and 3D detection. Due to the inherent differences in the algorithms used for 2D and 3D detections, separate commands or classes are used; however, they have cross-platform interoperability for visual anchors as presented in the last column. These visual anchors can be shared online between users of different MR platforms. However, the interoperability may pose new security and privacy risks.

7.2.2 Mobility and Privacy

In Sect. 7.1.2, various strategies in mobile MR processing were discussed to primarily distribute MR processing to reduce latency and provide a better mobile MR experience.

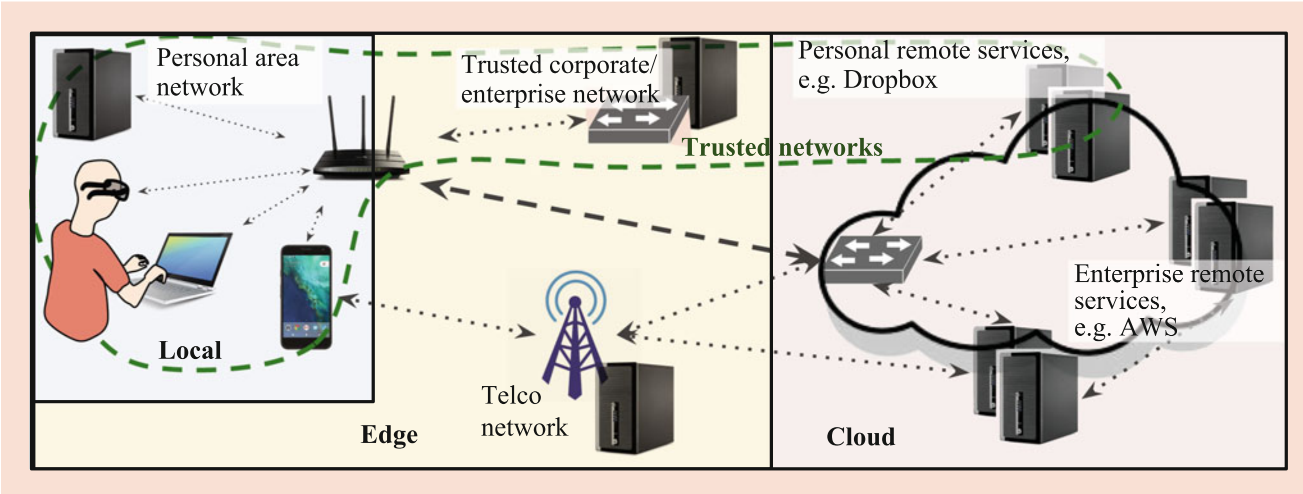


Fig. 7.3 Network of local devices, edge, and cloud auxiliary servers. A trusted subnet of this network is also shown, which is encircled by the green broken line

However, offloading the user data to remote edge and cloud servers may introduce risks. Enterprise networks or cloud storage servers may be considerably safe, but the trustworthiness of edge devices cannot be ensured. In general, as seen in Fig. 7.3, as we move away from the user device, the processing capacity increases, but the trustworthiness may not necessarily be ensured.

In order to investigate how much data were being offloaded by current MR apps, we performed packet transaction type analysis over a number of ARCore, ARKit, and Hololens applications. We used the following transaction-type categories: Application (`App`), Utilities (`Utilities`), and Advertising and Analytics (`Ads&Analytics`). We also separated transactions into download (`DL`) or upload (`UP`) to obtain the ratio of data packets being offloaded. Figure 7.4 shows that, while packets related to advertising and analytics were seen in each of the investigated ARCore apps, packets under this category are mostly not present in the current MR apps. Not only are the observed packet transaction lengths predominantly related to MR processing-and-services but also seem to be mainly used for offloading packets to auxiliary servers in some applications, such as HoloLens' holotour. These findings therefore highlight the potential privacy and security threats users are exposed to while using MR apps and affirm the need for countermeasures to ensure user privacy and security.

7.3 Protection Approaches for Mixed Reality

As vision processing and mobility are both being addressed and discussed well, privacy remains as a big challenge for MR. Given the sensing power and capabilities of MR, nascent as well as unknown privacy and safety risks arise. Now, we will present the various security and privacy works in MR according to the subsequent categorization discussed below.

Categorizing the Approaches Figure 7.6 shows an example MR environment with sample target elements for protection. Physical entities (e.g., desk, cup, or keyboard) from the environment are captured or detected. After detection, the resulting entities will be transformed or processed to deliver services accordingly. Depending on the service or application, different transformations are used. Finally, the results of the transformation are delivered to the user by rendering them (such as the virtual pet bird or the cup-contents indicator) through the device's output interfaces.

Figure 7.5 shows the four categories (and their subcategories) to which we will distribute the various related works. The first two categories of *input* and *output* protection are directly mapped to the associated risks with the *detection* and *rendering* stages of the MR processing pipeline, respectively. While *interaction* and *device* protection cannot be directly mapped along the pipeline but can be visualized within an MR environment as labeled in Fig. 7.6.

Table 7.2 lists the security and privacy properties, associated threats, and definitions. We follow the same combined list of properties that were presented in our original survey paper [10]. The top six properties are security-oriented, while the lower six are considered as privacy-oriented. The confidentiality property is considered both a security and privacy property. Consequently, some security properties are conversely considered as privacy threats and vice versa: for example, non-repudiation is the "threat" to plausible deniability. Nonetheless, these properties are not necessarily

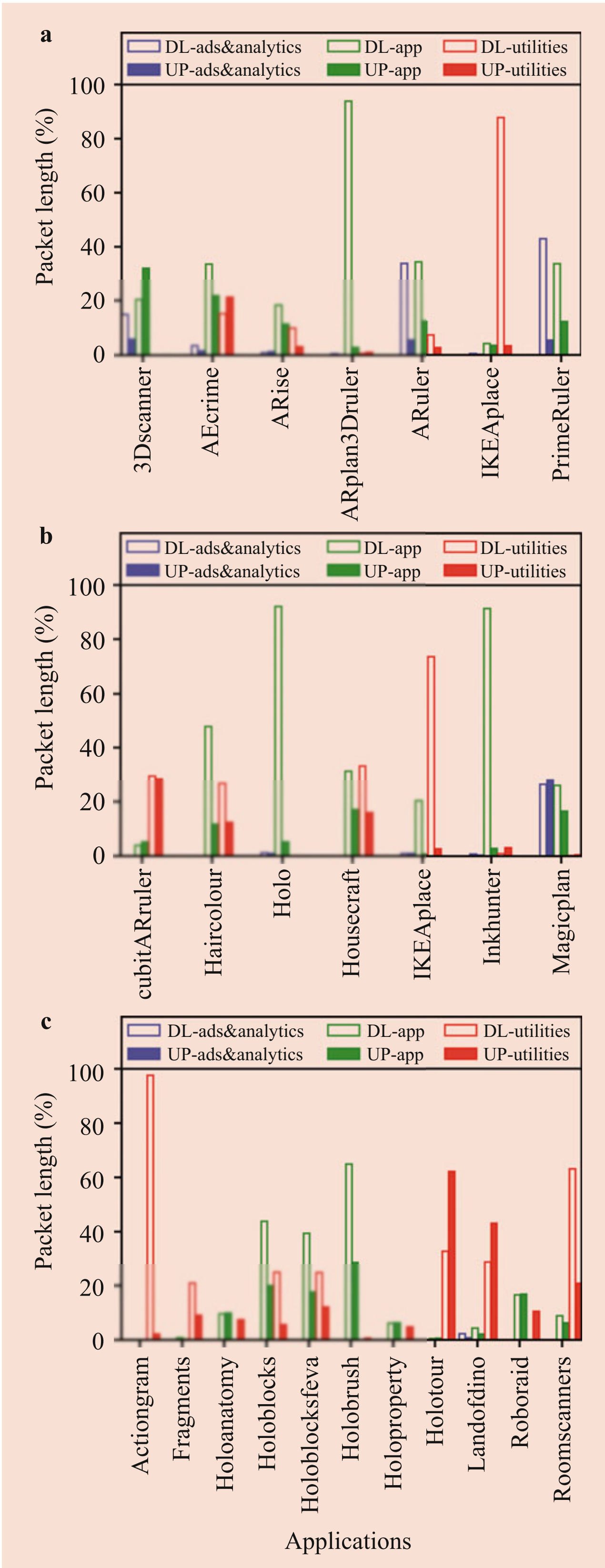


Fig. 7.4 Upload(UP) and download(DL) packet length distribution of MR apps to generalized categories in the first 5 min of app usage. NOte: This experiment was conducted in May 2019 (**a**) ARCore. (**b**) ARKit. (**c**) Hololens

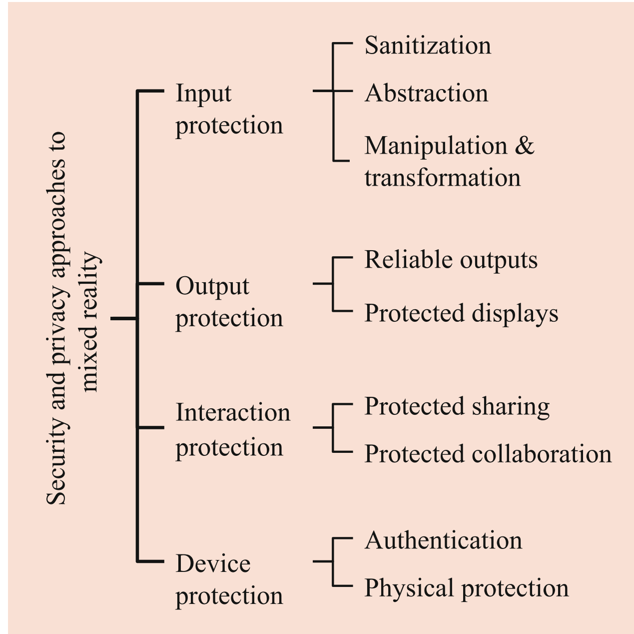


Fig. 7.5 A data-centric categorization of the various security and privacy work or approaches on mixed reality and related technologies

mutually exclusive and may be desired at the same time and, thus, be applied to different elements along the processing pipeline.

7.3.1 Input Protection

Several MR and related PETs have been proposed in the literature, and most of these early privacy works focused on 2D visual data protection (Fig. 7.7).

Visual Information Sanitization

Early approaches primarily involved applying sanitization techniques on visual media (i.e., image and video), e.g., selective blurring, replacement, or outright removal of sensitive portions in images or video frames. Sanitization removes latent and sensitive information from the input data stream. Various approaches use different methods to sanitize sensitive visual information, e.g., removing RGB and only showing contours [20]; detecting markers that signify sensitive content to be sanitized [21, 22], based on context [23, 24] or through user gestures [25]; or mechanically blocking the visual feed [26]. However, these methods only employ intrinsic policy enforcement or are "self-policing," which can potentially have a myopic view of privacy preferences of external objects. Thus, other works have focused on extrinsic policy enforcement to allow objects (including other users who are so-called bystanders) to communicate their privacy preferences. An early implementation [27] involved outright capture interference to prevent sensitive objects from being

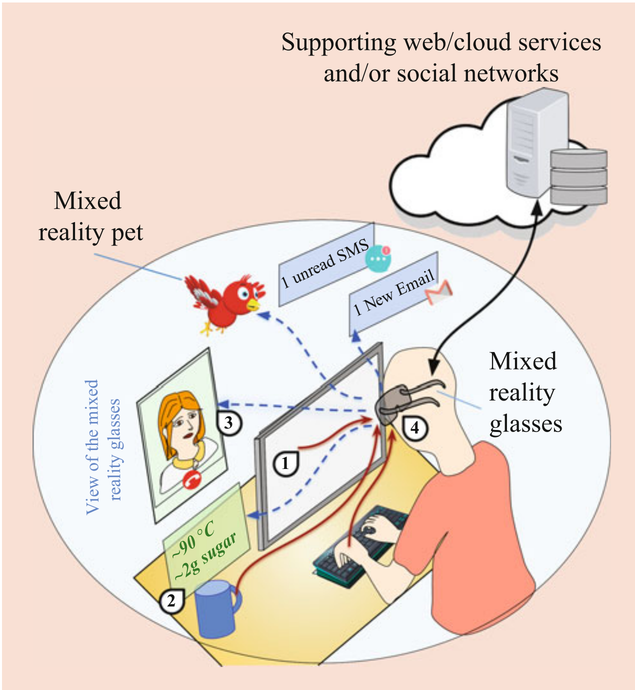


Fig. 7.6 A *mixed reality environment* with example points of protection as labeled: (1) *inputs* such as contents of the display monitor, (2) *rendered outputs* such as virtual display of a smart mug, (3) interactions such as collaborating with other users, and (4) *device* access to the mixed reality eyewear

captured by unauthorized visual capturing devices. It uses a camera that detects other cameras and flashes a strong light source to essentially blind these other cameras. Other methods allow for external policies to be communicated through shared databases [28] or through personal area network (PAN) technologies such as Bluetooth [29–31]. However, whether intrinsic or extrinsic, these sanitization approaches are still primarily focused on post-captured information, which can still pose security and privacy risks, and they have only been applied on the wide use case of visual capturing devices and not on actual MR devices or platforms.

Visual Information Abstraction

Abstraction addresses the earlier issues by reducing or eliminating the necessity of accessing the raw visual feed directly. In the `Recognizers` work, a hierarchical recognizer is inserted as an intermediary input protection layer, which also provides *intrinsic* input access control [32]. It follows the concept of least privilege to application input access—applications are only given the least amount of information necessary to run [33]. The least privilege access control has also been applied to secure user gesture detection. The PREPOSE gesture core, which is also inserted as an intermediary layer, only sends gesture events to the applications instead of the raw visual feed [34].

The same approach has also been used for providing spatial information while maintaining visual privacy [35] in a room-scale MR environment. Again, these applications do not need to know what the contents on the wall are; it only has to know that there is a surface that can be used for projecting outputs. Another recent work demonstrated how the visual processing and network access of a mobile AR/MR application can be siloed to protect visual information from malicious AR/MR applications [36]. However, most of these works employing abstraction are functionality or data specific and have neither presented or exposed actual risks with 3D MR data nor provided demonstration and evaluation against actual attacks.

`SafeMR`: Visual Information Access Control Thus, we designed and developed a proof-of-concept object-level abstraction we call `SafeMR` that can form the basis for providing the necessary visual protection for emerging MR applications [37,38]. It (1) provides visual information protection while (2) reducing visual processing latency by taking advantage of concurrent processing through reusable object-level abstractions. Our system follows the least privilege paradigm where applications are only provided with the minimum amount of information necessary for their intended functionality and/or permitted by the user as illustrated in Fig. 7.8. It shows a cascading and diminishing visual information representation which presents the different levels of privilege and, in turn, the amount of information an application has access to.

`SafeMR` is inserted as an intermediary layer between the trusted device APIs (e.g., ARCore) and the third-party applications to provide information access to visual data via object-level abstractions by detecting the objects and exposing them as abstractions to the applications. Applications' access to the object abstractions and the privilege level is specified by the user privacy preferences. Sample screenshots from our demo application are shown in Fig. 7.9. As shown in Fig. 7.9b, object sensitivity can be specified by toggling the detected objects from a list populated by an initial detection process (seen as the top button on the menu with label "MR DEMO INITIALIZE"). This shows object-level permissions and access control. Likewise, we implemented a "privacy slider" to allow the users to change the privilege levels as defined in Fig. 7.8. Figure 7.9c, d illustrates the use of the slider to set two different privilege levels, i.e., level B and D, respectively.

Data Manipulation and Transformations

Sanitization and abstraction methods primarily focus on providing protection by controlling or blocking the data flow. These methods can potentially be limiting in providing data utility. Others have proposed employing data manipulation or transformation to protect sensitive information.

Table 7.2 Security and privacy properties and their corresponding threats as defined in [10]

	Property	Threat	Definition
↑ *Security-oriented*	Integrity	Tampering	Storage, flow, or process of data in MR is not and cannot be *tampered* or *modified*.
	Non-repudiation	Repudiation	Modification or generation of data, flow, or process *cannot be denied*.
	Availability	Denial of Service	Necessary data, flow, or process for an MR system should be available.
	Authorization	Elevation of Privilege	Actions or processes should be originated from authorized and verifiable entities.
	Authentication	Spoofing	Only the legitimate entities should be allowed to access the MR device or service.
	Identification	Anonymity	All actions should be identified to the corresponding entity.
	Confidentiality	Disclosure of Information	All actions involving sensitive or personal data, flow, or process should remain undisclosed.
Privacy-oriented ↓	Anonymity and Pseudonymity	Identifiability	Entities should be able to remove the identifiable association or relationship to the data stored, flow, or process.
	Unlinkability	Linkability	Any link or relationship of the entity, i.e., user or party, to the data stored, flow, or process as well as with other entities (e.g., data to data, data to flow, and so on) cannot be identified or distinguished.
	Unobservability and Undetectability	Detectability	An entities' existence cannot be ensured or distinguished by an attacker, or an entity can be deemed unobservable or undetectable by an adversary, or the entity cannot be distinguished from randomly generated entities.
	Plausible Deniability	Non-repudiation	An entity should be able to deny that they are the originator of a process, data flow, or data storage.
	Content Awareness	Unawareness	An entity (usually the user) should be aware of all data flows or processes divulged, especially those that are personally identifiable or sensitive.
	Policy and Consent Compliance	Non-compliance	An MR system should follow and provide guarantees that it follows the policies that aim to protect the user's privacy or security.

Encryption-Based Techniques Among these methods are encryption-based techniques, which allow for data queries or computations over encrypted data. Various flavors of encryption have been employed in privacy-preserving image and video processing. For example, HE-SIFT [39] performs bit-reversing and local encryption to the raw image before feature description using SIFT [40] to make dominant and sensitive features recessive. Image feature extraction, description, and matching are all performed in the encrypted domain using near full homomorphism and, thus, have a very slow computation time. Improvements such as Leveled-HE reduces the computation time of HE-SIFT [41]. SECSIFT [42,43] also improved on the computation time of HE-SIFT by instead using an order-preserving encryption. Other improvements utilized big data computation techniques to expedite secure image processing such as the use of a combination of MapReduce and ciphertext-policy attribute-based encryption [44], or the use of Google's Encrypted BigQuery Client for Paillier HE computations [45].

Secure Multi-Party Visual Processing Another cryptographic approach used in image and video processing is secure multi-party computation (SMC) or secret sharing, which allows computation of data from two or more sources or parties without necessarily knowing about the actual data each party has. Figure 7.10 shows a possible SMC setup. A privacy-preserving photo-sharing service has been designed using a two-party secret sharing by "by splitting a photo into a public part, which contains most of the volume (in bytes) of the original, and a secret part which contains most of the original's information" [46]. A virtual cloth try-on service also used secret sharing and two-party SMC [47]: the body measurements of the user is split between the user's mobile device and the server and are both encrypted. The server, which has the clothing information, can compute a 3D model of the user wearing the piece of clothing by combining the body measurement information and the clothing information to generate an encrypted output, which is sent to the user device. The user device decrypts the result and combines it with the local secret to reveal the 3D model of the user "wearing" the piece of clothing. However, like most cryptography-based methods, they are algorithm-specific; thus, every algorithm has to be re-engineered to apply cryptographic protocols on their computations.

Facial de-identification Other techniques have focused on facial de-identification using non-cryptographic image manipulation to achieve k-anonymity for providing identity privacy [48–50]. The succeeding face de-identification work has focused on balancing utility and privacy [51]. Contrarily,

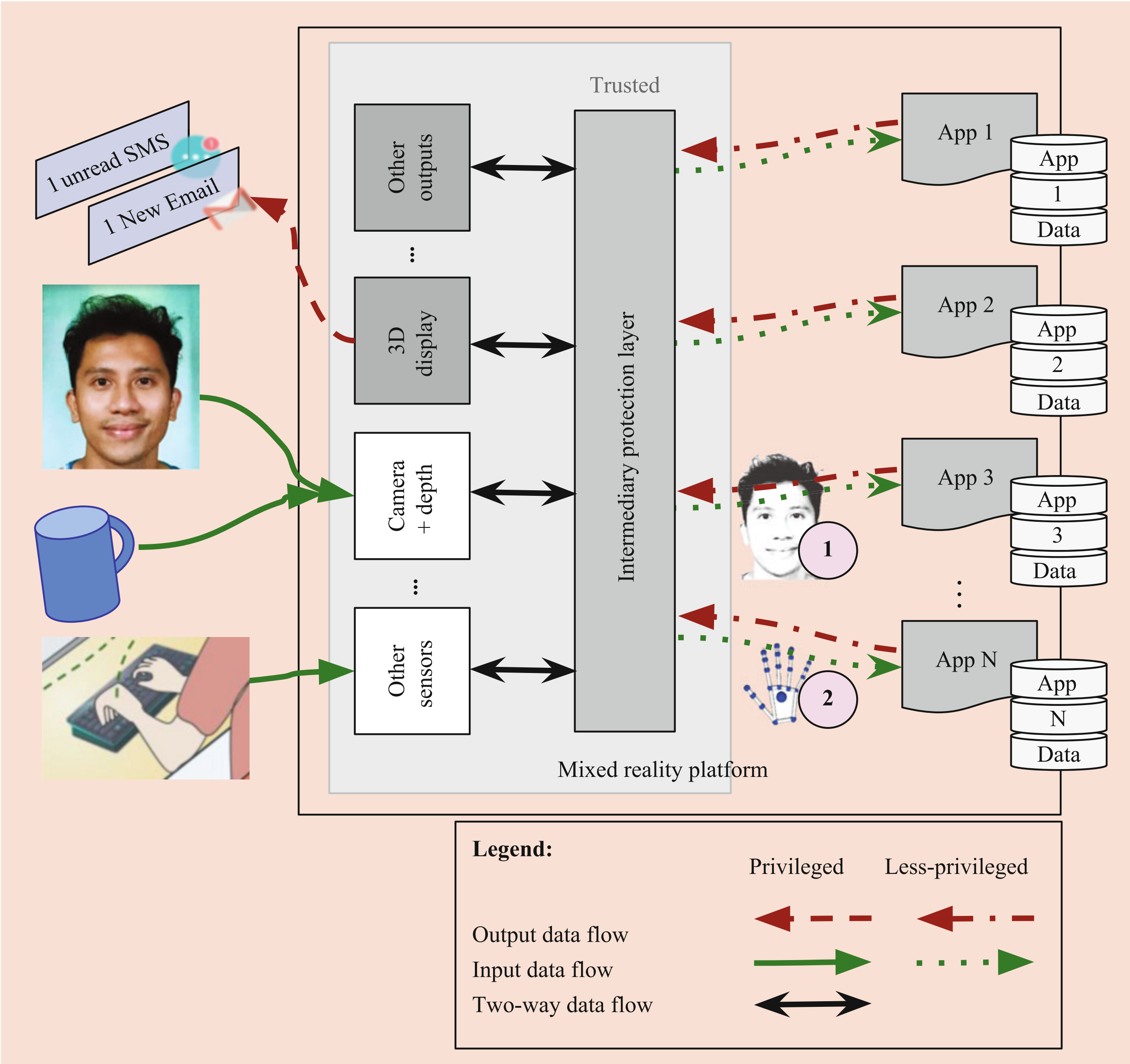


Fig. 7.7 A generic block diagram that shows an inserted ***intermediary protection*** layer between the applications and device resources. Example strategies for input protection are also shown: (1) *information reduction* or *partial sanitization*, e.g., from RGB facial information to facial outline only, or (2) skeletal information instead of raw hand video capture

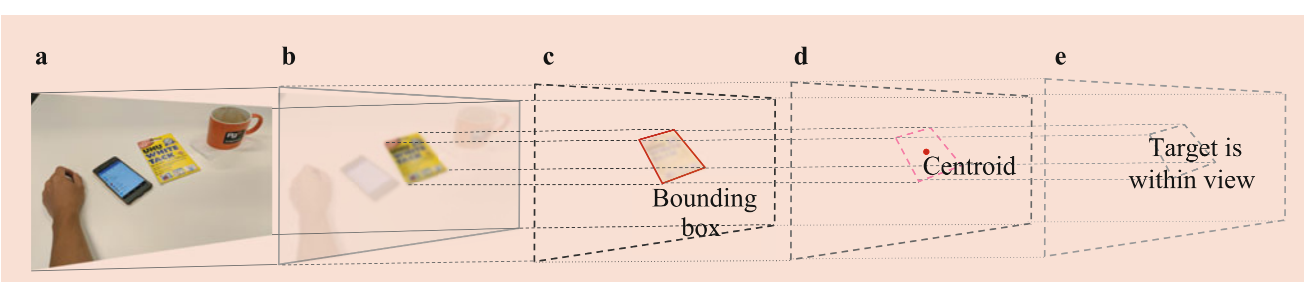


Fig. 7.8 Diminishing information: (**a**) the raw visual capture; (**b**) the target is cropped out but still with complete visual information of the target; (**c**) only the bounding box of the target is exposed; (**d**) only the centroid of the target is exposed; and (**e**) only the binary presence, whether the target is within view or not, is exposed

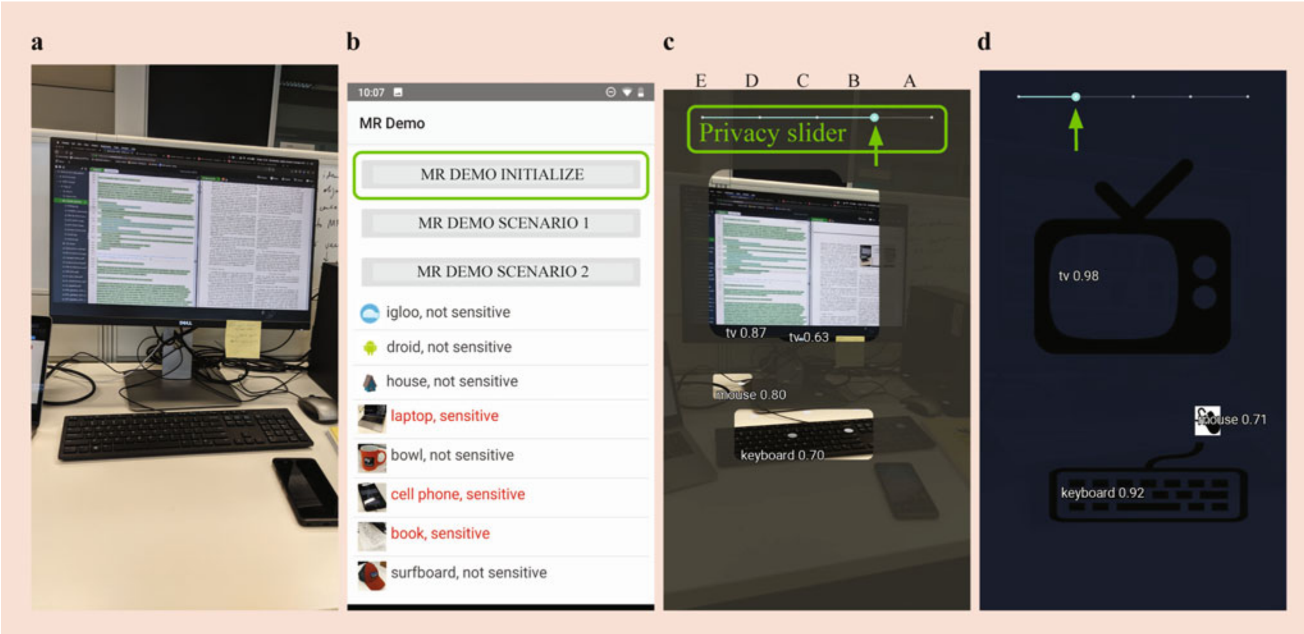


Fig. 7.9 `SafeMR` demo showing different privilege levels. (**a**) Scene. (**b**) Defining object sensitivity. (**c**) Privilege Level B. (**d**) Privilege Level D

much recent works have leveraged generative adversarial networks for deceiving a potentially adversarial data collector to de-identify faces but ensuring high demographic utility (say, only revealing gender) of the resulting de-identified face [52, 53].

3D Data Attacks and Protection The same manipulation can be extended over 3D spatial data that is utilized in MR systems. Instead of providing complete 3D spatial data, a sanitized or "salted" virtual reconstruction of the physical space can be provided to third-party applications. For example, instead of showing the 3D capture of a table in the scene with all 3D data of the objects on the table, a generalized horizontal platform or surface can be provided. The potentially sensitive objects on the table are thus kept confidential. A tunable parameter provides the balance between sanitization and utility. Using this tunability, similar notions of privacy guarantee to differential privacy and k-anonymity can be provided. However, this approach is yet to be realized, but virtual reconstruction has been used to address delayed alignment issues in AR [54]. This approach can work well with other detection and rendering strategies of sanitization and abstraction as well as in private collaborative interactions. It also opens the possibility to have an active defense strategy where "salted" reconstructions are offered as a honeypot to adversaries. A recent work demonstrated how original scenes from 3D point cloud data can be revealed using machine learning [55]. As a countermeasure, a concurrent work designed privacy-preserving method of pose estimation to counter the scene revelation [56]: 3D "line" clouds are used instead of 3D point clouds during pose estimation to obfuscate 3D structural information; however, this approach only addresses the pose estimation functionality and does not present the viability for surface or object detection, which is necessary for a virtual object to be rendered or "anchored" onto. Thus, it is still necessary for 3D point cloud data to be exposed but with privacy-preserving transformations to hide sensitive content and prevent spatial recognition.

3D Spatial Data Transformations We have investigated the viability of surface-to-plane generalizations as spatial privacy preservation for spatial data captured by the Microsoft HoloLens. Our preliminary work showed evidence on how an adversary can easily infer spaces from captured 3D point cloud data from Microsoft HoloLens and how, even with spatial generalization (i.e., the 3D space is generalized into a set of planes), spatial inference is still possible at a significant success rate [57]. Furthermore, we are also working on deriving a heuristic measure for spatial privacy risk. We refer to spatial privacy as an extension of location privacy; specifically, we pose it as the likelihood of an adversary to identify the space a user is in from the revealed spatial data that is captured by MR devices and platforms like the HoloLens or from ARCore.

Our current work focuses on augmenting the currently inadequate surface-to-plane generalizations with conservative plane releasing (as shown in Fig. 7.11b) as a stronger countermeasure against spatial inference attacks [58]. Our experiments over accumulated data from continuously and

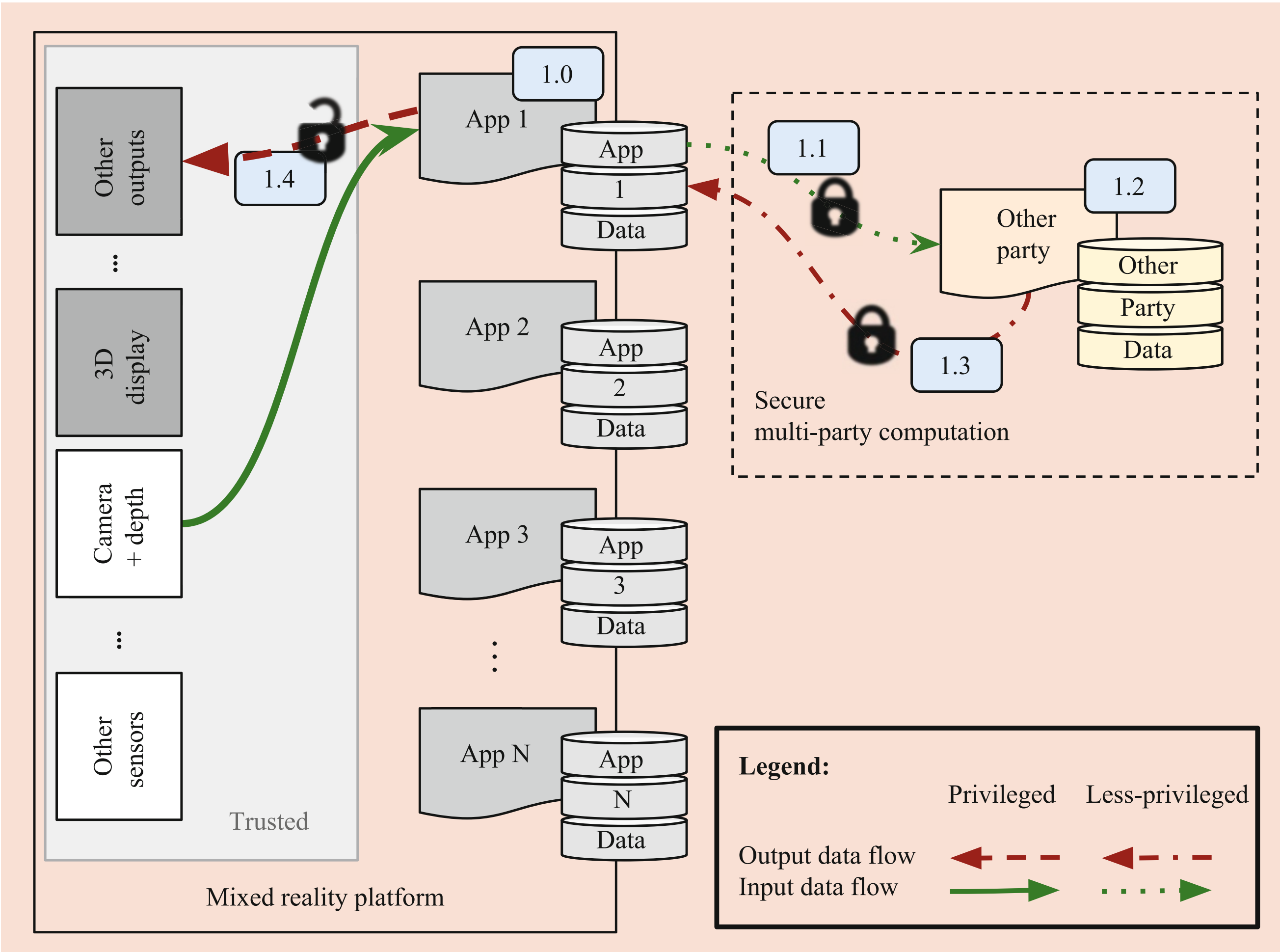


Fig. 7.10 Generic block diagram of a cryptographic technique using *secure multi-party* computation where two or more parties exchange secrets (1.1 and 1.3) to extract combined knowledge (1.2 and 1.4) without the need to divulge or decrypt each other's data share

7

successively released spatial data showed that we can reveal up to 11 planes and avoid spatial recognition for at least half of the time for sufficiently large revealed spaces, i.e., radius $\leq$ 1.0 meters. And, for the occasions that the adversary correctly recognizes the space, an adversary's guess spatial location can be off by at least 3 meters. Given that plane generalization is already (or can potentially easily be) implemented in most existing MR platforms, thus, perhaps, what remains is the implementation of conservative plane releasing on actual MR platforms as a viable countermeasure against spatial inference attacks.

7.3.2 Output Protection

After capturing and processing data, the resulting outputs of MR applications are sent out to the displays for consumption of the user. Similar to input protection, it is also desirable for applications to only have access and control over their outputs and should not interfere with other outputs or objects.

For example, in the smart information hovering over the cup in Fig. 7.6, malicious applications should not be able to modify the sugar-level information. Other adversarial output attacks include clickjacking, i.e., deceives users to "clicking' on sensitive elements through transparent interfaces [9], as well as physiological attacks such as inducing epileptic seizures through a visual trigger [8]. A recent study using immersive VR has demonstrated how to "disorient users, turn [their HMD] camera on without their knowledge, overlay images in their field of vision, and [...] control immersed users and move them to a location in physical space without their knowledge." [59].

Output Reliability and User Safety
Despite these output-targeted attacks being brought up as early as 2014 [9], current MR systems still have loose output access control. As a result, adversaries can still potentially *tamper* or *spoof* outputs that can compromise user safety. Despite these output-targeted attacks being brought up as early as 2014 [9], current MR systems still have loose output

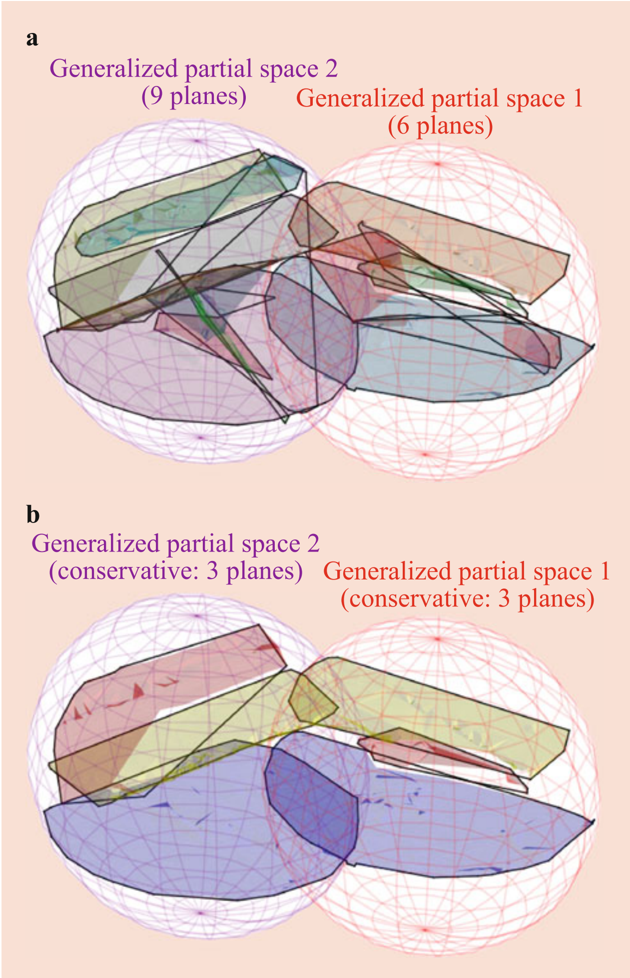


Fig. 7.11 Example of conservative plane releasing. (**a**) Reference generalized partial spaces. (**b**) Conservatively released planes

access control. As a result, adversaries can still potentially tamper or spoof outputs that can compromise user safety. In 2016, an output access control framework with an object-level granularity has been proposed to make output-handling enforcement easier [60]. It can be implemented as an intermediary layer, as in Fig. 7.7, and follows a set of output policies to manage output rendering priority in terms of synthetic object transparency, arrangement, occlusion, and other possible spatial attributes to combat attacks such as clickjacking. In 2017, they followed it up with a design framework [61] called `Arya` for output policy specification and enforcement to ensure policy compliance, integrity, non-repudiation, availability, and authorization. This ensures that correct outputs are always available, an output's originator cannot be denied, and only authorized applications can produce such outputs. A follow-up work focused on mediating rendering conflicts from multiple MR applications [62]. Succeeding work builds up on Arya's weakness when it comes to dynamic and complex environments requiring heterogeneous sources of policies using reinforcement learning [63].

Protecting External Displays
When input and output interfaces are on the same medium or are integrated together such as on touch screens, these interfaces are vulnerable to physical inference threats or visual channel exploits such as shoulder-surfing attacks. MR can be leveraged to provide secrecy and privacy on certain sensitive contexts requiring output confidentiality. Various MR-leveraged strategies include content hiding where a near-eye HMD can deliver secret and private information on public [64]. Most MR eyewear such as the Google Glass can be utilized for this purpose. Other approaches involve the actual hiding of content using optical strategies such as rolling shutter or variable frame rate [65–67]. This technique hides content from human attackers, i.e., shoulder-surfers, but is still vulnerable to machine-aided inference or capture, i.e., setting the capturing device's frame rate to that of the rolling shutter. Other secret display approaches have also used visual cryptographic techniques such as visual secret sharing (VSS) schemes, which allows optical decryption of secrets by overlaying the visual cipher with the visual key. However, these were primarily aimed at printed content [68] and require strict alignment, which is difficult in MR displays. The VSS technique can be relaxed to use standard secret sharing using codes, i.e., barcodes and QR codes. An MR device which has the secret code key can be used to read the publicly viewable code cipher and augment the decrypted content over the cipher. This type of visual cryptography has been applied to both print [69] and electronic displays [70, 71]. These techniques can further be utilized for protecting sensitive content on displays during input capture. Instead of providing privacy protection through, say, post-capture sanitization, the captured ciphers will remain secure as long as the secret shares or keys are kept secure. Thus, even if the ciphers are captured during input sensing, the content stays secure.

7.3.3 Protecting User Interactions

In contrast to current widely adapted technologies like computers and smart phones, MR can enable entirely new and different ways of interacting with the world, with machines, and with other users. Figure 7.12 shows a screenshot from a demonstration video from Microsoft Research on their Holoportation project, which allows virtual teleportation in real time. Consequently, one of the key (yet latent) expectations with these kinds of services and functionalities is how users can have shared space experiences with assurances of security and privacy.

Imbalance of Power from MR Boundaries As early as 1998, concerns on the boundaries between physical and virtual spaces (Fig. 7.6) in MR and on the directionality of these boundaries have been raised [72]. The directionality can influence the balance of power, mutuality, and privacy between users in shared spaces. For example, the boundary (labelled 1) in Fig. 7.13b allows User 2 to receive full information (solid arrow labeled 2) from User 1, while User 1 receives partial information (broken arrow labeled 3) from User 2. The boundary enables an "imbalance of power," which can have potential privacy and ethical effects on the users. For example, early observation work shows territoriality in collaborative tabletop workspaces [73], while a much recent work on multi-user interactions in MR showed the conflicts that can arise, i.e., misuse by multiple actors, apart from the varying degrees of concerns that the user has [19]. Furthermore, this imbalance is not only confined to collaborative but also on non-collaborative shared spaces where, for example, only a subset of the users are using an HMD, such as the Google Glass, while others are not. A great deal of criticism has already been received by Google Glass on its potential violations of privacy.

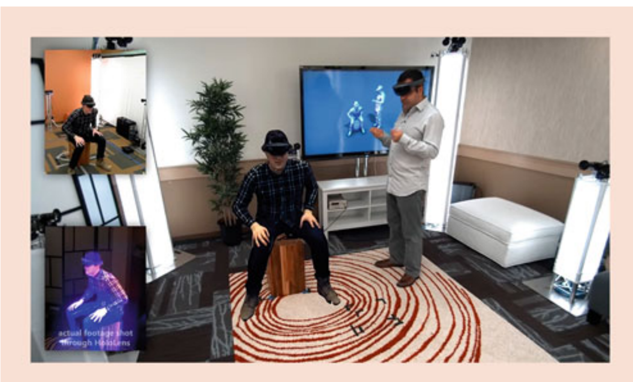


Fig. 7.12 `Holoportation` by Microsoft Research: an example of shared-space service. The person sitting (left) is "holoported" to the room with the person standing (right) using MR technology. Screenshot from https://youtu.be/7d59O6cfaM0. Used with permission from Microsoft

Protecting Collaborative Interactions
Unsurprisingly, the same extrinsic protection is also desired in collaborative interactions to enable user-originated policies and privacy preferences. Early implementations demonstrated how users can change the privacy of virtual objects in a shared virtual environment using privacy lamps and vampire mirrors [74, 75]. (Privacy lamps "emit" a virtual light cone in which users can put objects within the light cone to mark them as private, while *vampire mirrors* are used to determine privacy of objects by showing full reflections of public objects while private objects are either invisible or transparent.) Succeeding work demonstrated the use of hand gestures to signify user privacy preferences [76]. Other developments worked on mediating conflicts in digital workspaces explored the use of multi-user coordination policies [77]. For example, to increase group awareness, they employed cooperative gestures, which requires gesture contributions from more than one user to enforce a single command, such as clearing the entire screen when users do the erase gesture together [78].

Feed-Through Signaling Other developments have also focused on providing feed-through information to deliver signals that would have been available in a shared physical space but is not readily cross-conveyed between remote physical spaces [79]. Feed-through mechanisms were also used to cross-convey user privacy preferences [80]. For example, Fig. 7.13b shows a situation in which User *n* enters the shared space (labelled 4) on the same physical space as User 2, which triggers an alarm (labelled 5) or notification for

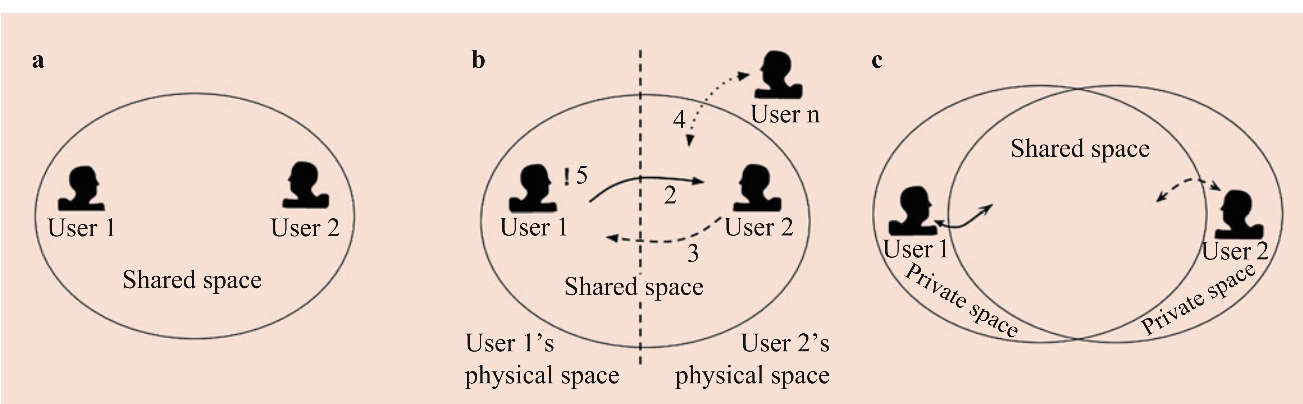


Fig. 7.13 Shared spaces. (**a**) A simplified virtual *shared space* diagram. (**b**) A possible separation in the underlying physical space that creates boundaries between users and devices. (**c**) A collaborative space with a shared space and *private spaces*

User 1. The notification serves as a feed-through signal that crosses over the MR boundary. By informing participants of such information, an imbalance of power can be rebalanced through negotiations. Non-AR feed-through signaling has also been used in a non-shared space context such as the use of wearable bands that lights up in different colors depending on the smartphone activity of the user [81]. However, the pervasive nature of these feed-through mechanisms can still pose security and privacy risks; thus, these mechanisms should be regulated and properly managed. A careful balance between the users' privacy in a shared space and the utility of the space as a communication medium is ought to be sought.

Various strategies also arose from competitive gaming, which demands secrecy and privacy in order to make strategies while performing other tasks in a shared environment. Some of these strategies include Private Interaction Panels (or PIPs) that provides a virtual private region on your gaming console [82] and variable view from see-through AR to full VR [83]. The privacy lamps and mirrors also act as private spaces similar to a PIP. Other PIP-like private regions have been demonstrated on handheld gaming devices [84, 85, 118]. Overall, what rises from these strategies is the utilization of different portions of space as public and private regions where users can actively move objects across the regions to change their privacy.

Protecting Sharing Initialization

Similar to protected external displays, MR can also be leveraged to provide a protected method for securely initializing a collaborative channel. Out-of-band techniques using MR platforms and, thus, leveraging MR capabilities can be employed in secure channel initialization. Such techniques have been demonstrated, including the use of wireless localization paired with facial recognition information for cross-authentication [86], or the use of visual cues for confirming a shared secret using HoloLens [87, 88].

7.3.4 Device Protection

This last category focuses on the actual *physical* MR device and its input and output *interfaces*. This implicitly protects data that is used in the earlier three aspects by ensuring device-level protection. Authentication, authorization, and identifiability are among the most important properties for device protection.

Protecting Device Access

The primary threats to device access are identity spoofing and unauthorized access; thus, all subsequent approaches aim to provide protection against such threats. Currently, password still remains as the most utilized method for authentication [89]. To enhance protection, multi-factor authentication (MFA) is now being adopted, which employs two or more independent methods for authentication. It usually involves the use of the traditional password method coupled with, say, a dynamic key (e.g., one-time password) that can be sent to the user via SMS, email, or voice call. The two-factor variant has been recommended as a security enhancement, particularly for online services like E-mail, cloud storage, e-commerce, banking, and social networks.

To allow users to conveniently wear the MR device (i.e., as a pair of smart glasses), most of the MR device design goal is to achieve the eye-wear form factor. With this, the device is mostly in contact with the user, which allows for various novel authentication methods leveraging physiological signals and other user gestures. Some of these novel strategies include the following: finger and hand gestures using a 3D sensor [90]; head and blinking gestures triggered by visual cues [91]; head movements triggered by an auditory cue [92]; and active physiological signals such as breathing [93], PPG signals [94], and even bone (sound) conduction [95]. Other methods combine two or more modes in a singular method, such as a combination of facial, iris, and periocular information for user authentication [96], and gaze gestures and touch keys [97]. With most HMD MR devices having gaze-tracking and other near-eye sensors, these authentication methods can be easily applied on MR platforms.

Protecting Physical Interfaces

As discussed in the section on external displays (in section "Protecting External Displays"), MR interfaces are vulnerable from malicious inference, which can lead *disclosure* of *input* activity, and/or *output display information*. Currently available MR HMDs project content through see-through lenses. The displayed content on the see-through lenses can leak private information and be observed externally. Visual capturing devices can be used to capture and extract information from the display leakage. Moreover, external input interfaces suffer from the same inference and side-channel attacks.

Various *strategies* have already been discussed in the section on protecting external displays. The same visual cryptographic techniques can be used to protect input/output interfaces [98], or using key scrambling to hide input activity from external inference [99]. Further optical strategies include the use of *polarization* (Fig. 7.14) combined with *narrowband illumination* to maximize display transmission while minimizing leakage [100]. Other strategies use optical reflective properties to only show content at a certain viewing angle [101]. Active camouflaging techniques have also been demonstrated to allow the screens to blend with its surrounding [102]. Both TaPS widgets and the chameleon-inspired camouflaging are physically hiding sensitive objects or information from visual capture. The content-hiding

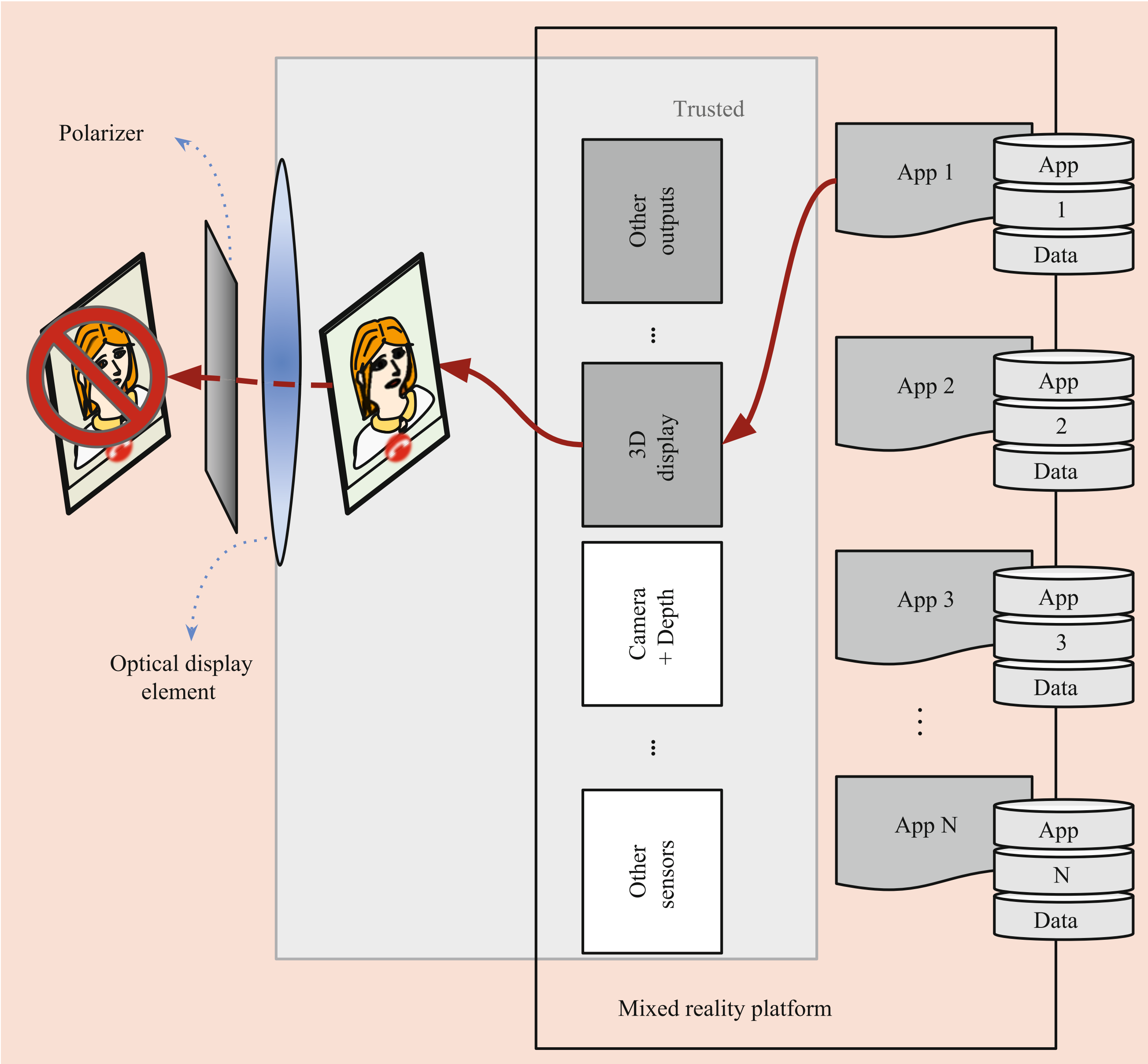


Fig. 7.14 Sample interface protection strategies: inserting a *polarizer* to prevent or block display leakage. All elements to the left of the *optical display element* are considered *vulnerable* to external inference or capture

methods discussed in section "Protecting External Displays" to hide outputs are also optical strategies.

7.3.5 Open Research Challenges

As we have presented a wide understanding of security and privacy in mixed reality, various challenges and opportunities still remain for every aspect. While uptake in mixed reality continues, devices and services will also evolve, leading to new research directions. In this section, we discuss the remaining challenges, potential new opportunities, and future directions.

Remaining Challenges to Input Protection Despite the variety of proposed protection measures, it is still a challenge to determine which objects have to be protected. Specifically, the machine must know what type of inputs or objects the different applications will require and, hence, which sensitive information blended with these inputs need protection. This then requires the necessity for an ontology and taxonomy of objects (including user gestures and inputs) and application requirements, so that specific protection can be applied based on the user-specified sensitivity of the object and as required by the application as well. For example, an MR-painting application may require the detection of different types of brushes, but the machine does not know how to "see" or detect the brushes. We can use object detection, but we also want to prevent the machine from seeing other sensitive objects. A recent approach has proposed a "a collaborative authorization framework [that integrates] organizations, developers, device manufacturers and users, [...] to handle different cases and needs by leveraging attribute-based policies" [103].

Moreover, given the rather specific and heterogeneous challenge, a heterogeneous sensing management framework is also required to allow users to specify and control fine-grained permissions to applications accessing data not just at the sensor level but also at an object level. With this framework, applications must only have access to objects and information they require and must not be able to access the information stream of other applications. Likewise, users should be able to enforce heterogeneous permissions to the different applications. A sensor-level management framework has already been demonstrated in Android [104]. A similar but more fine-grained framework as we have described is necessary for a privacy-aware *near*-future with MR. In conjunction, aside from allowing users to specify the fine-grained permissions, we can also leverage machine learning and artificial intelligence to *dynamically* apply the necessary protection—whether via sanitization, abstraction, or transformation—for both pre-determined and *previously unseen potentially* sensitive objects.

Differently, as MR is now being leveraged in increasing awareness for *dangerous situations* or tasks, say, on manufacturing, it is counter-intuitive to limit the data access of safety and danger warning applications. Thus, it is imperative to employ *heterogeneous access* mechanisms to provide mission-critical applications with full access to environment data. MR device manufacturers (and their developers) can integrate such danger detection and warning systems to the core APIs of their MR device, while other non-critical third-party applications may remain to have minimal access to the environment information.

Remaining Challenges to Output Protection Similar to input protection, an ontological challenge to output protection also arises. This comes from output access control policies being applied also as an intermediary access control layer (see Fig. 7.7) between applications and output interfaces; thus, policies require a heterogeneous mixture of outputs, output channels, and applications. And, to enforce these output control policies, a reference policy framework has to exist through which the protection is applied. A further challenge is the *specification* and *enforcement* of these policies, particularly on who will specify them and how they will be effectively enforced. Furthermore, in the output side, risks and dangers are more imminent because once adversaries have access to the output interfaces, they can actuate malicious response or output, thus the more these output protection strategies are necessitated.

Remaining Challenges to User Interactions The additional dimension of varying user preferences are further compounded by the varying use cases of user interactions. Depending on the context or situation, privacy and security concerns, as well as the degree of concern, can vary. Thus, there is a great deal of subjectivity to determine what is the most effective protection mechanism during sharing or interactions. A recent work has posed a set of security and functionality goals for multi-user MR and demonstrated it using HoloLens [105]. They proposed a design for a sharing control module that can be implemented either as an application- or OS-level interface.

However, a similar ontological challenge also arises as we design and develop these sharing mechanisms, especially as we provide cross-platform-sharing capabilities. Thus, we reiterate that, before everything else, we should probably ask first: "Who or what are we protecting?" Ultimately, the MR platform should be ready to accommodate the various answers to this question, that is, the users (as well as bystanders) should be empowered and be allowed to specify their privacy preferences on these shared MR environments.

Remaining Challenges to Device Protection As MR device and hardware technology research and development continue, various device-level vulnerabilities may arise, which we may have previously been unknown. Thus, we also have to continue to scrutinize the various MR devices that are being released to assess and determine these vulnerabilities with the intention of reducing and mitigating user risks. Nonetheless, some developments on MR (or, generally, wearable) user authentication for device access may include the translation of existing methods from mobile technology, such as fingerprint authentication.

As presented in Sect. 7.3.4, external display protection approaches are particularly helpful in providing security and privacy during certain activities in shared or public spaces due to the secrecy provided, for example, by the near-eye displays, which can perform the decryption and visual augmentation. However, they are mostly applicable to pre-determined information or activities that are known to be sensitive, such as password input or ATM PIN input. Moreover, the visual cryptography or similar approaches are limited by the alignment requirement. Nonetheless, these techniques are still helpful in the contexts they are designed for.

Furthermore, issues on display leakage have also been raised. Given that various display technologies are used for MR devices, various protection methods are also necessary. For example, polarization may not work for certain displays due to the refractive characteristics of the material used in the MR display.

7.3.6 Future Directions

Aside from the previous remaining challenges on input, output, interaction, and device aspects, we pose the following further challenges and opportunities for future academic endeavors.

Exploring and Extending MR Data Attack Scenarios In Sect. 7.3.1, 3D data protection strategies were presented, which primarily utilize 3D data manipulation to confuse adversaries trying to extract information about the 3D data (e.g., identify the location of the space represented by the 3D data). However, these countermeasures are only as good as the best attack it can defend against. Thus, it is a worthwhile effort to explore improvements on attack scenarios as future work. The current attack scenarios used were based on classifiers using the nearest-neighbor search and deep neural network. Further investigation is required over attack performance of adversaries with strong background knowledge (for example, having access to other user information that can reveal their location) and how these attackers will perform in spatial inference attacks.

Furthermore, there are other platforms and use cases in which 3D data is also used: for example, 3D lidar data used in geo-spatial work, captured by self-driving cars, and, now, by many other applications on recent smartphones. Thus, adversaries from across these various 3D data sources can potentially collude to extract sensitive user information despite current 3D data protection measures.

Risk Analysis of Existing/Future Devices and Platforms As devices and their capabilities continue to evolve, a systematic security and privacy analysis of MR applications, devices, and platforms should still be performed in order to identify other potential and latent risks, particularly on their input capabilities. For example, the scanning capability of current devices, such as the HoloLens, should be further investigated whether it can potentially be used to detect heartbeats or other physiological signals of bystanders or, aside from what we have revealed [57], how these spatial data can be further abused by adversaries.

An integral aspect of device risk analysis is also the risk analysis of the data collected by these devices. In one of our recent work, we presented a measure of risk (in terms of spatial identifiability) using spatial complexity based on stochastic and geometric measurements over the spatial data [106]. Using these measures, we compared the risk of spatial data captured by the Microsoft HoloLens and of Google ARCore (using Google Pixel 2) and showed that Microsoft HoloLens poses more identifiability risks. Similar assessments can be performed on other platforms, such as Apple's ARKit and Oculus VR devices.

Utilization of Other Wearables Majority of past work in security and privacy of MR focused on developing solutions on the primary head-mounted device and its associated peripherals. All indicators are pointing toward the fact that users will wear more than one wearable at all times. For example, smartwatches and smartbands have recently become quite popular among consumers. In the case of security, these devices can be leveraged for continuous or multi-factor authenticators, an alternative channel of communication, environmental or body sensing, etc. In the case of privacy, wearables can be utilized for local processing, alternate non-sensitive form of identifiers, input detection and identification, etc. However, effective integration of such general-purpose wearables into MR sub-system in and around the body will be a challenge for developers when these devices are potentially developed by different vendors. Therefore, extensions of local standards such as UPnP and NSD (Network Service Discovery) would be useful contributions for MR systems.

Protection of Bystander Privacy This is one of the most challenging aspect in the privacy of MR, while it is one of the most impact-full aspect in terms of public acceptance of MR technology. The notion that a person (say, their physical appearance through images) can be inadvertently become part of someone else data collection is a complicated matter to resolve. When Google Glass was first released, there were intense debate about bystander privacy [107]. This has led to banning of Google Glass from a number of public places, which is not the ideal solution. While there is some work in the past [108], practical solutions for bystander privacy are largely an open research challenge.

Edge Computing for MR Security Edge computing has gained significant interest from both academia and industry as a potential solution to latency trade-offs in MR applications [109]. A few investigations on cloud- and edge-assisted platforms for MR mobility were discussed in 7.1.2. However, edge computing, in its current form, will not solve the data ownership and privacy problems. Rather, it could even make them even more acute due to the storage and processing of data in locations/devices that are not under the control of application or service provider.

Similar to MR technology, progresses on these computing paradigms (whether edge, fog, cloud, or their combinations) are still currently being developed. As a consequence, the security and privacy work over these computing paradigms remain to be in silos; however, it is argued that these efforts need to be in synergy as when we put together various components of, say, an MR mobile environment, novel security and privacy issues arise [110]. Consequently, as both MR and edge computing are currently still being developed, it is opportune to design and develop research platforms or frameworks that allow the co-development of both areas. One potential approach is to dissolve hard boundaries of edge and allow elastic function distribution to different entities that span from user device to traditional cloud depending on the control and ownership of entities. For example, personal laptop, home router, and game controller are more trusted devices compared with edge computing boxes on network base stations. Personal cloud entities like Dropbox are more trusted that cloud storage provided by the MR service. Thus, integration of personal trustworthiness of entities into function orchestration in edge computing will be an interesting future direction.

Shared Standards and Common Ontology Cooperation among developers and manufacturers is required for the use of a standard. This is a common issue with most emerging and currently developing technologies. For example, in the IoT space, there are various existing standards that inadvertently impact the adoption of the technology. Likewise, in MR, there are endeavors in pushing for a shared or open platform particularly for spatial computing, such as the OpenARCloud (https://www.openarcloud.org/oscp).

A common consequence (albeit, mostly, a preliminary one) to standardization is the additional workload and, sometimes, complexity that is loaded to developers in conforming to these standards. An example would be the creation of an ontology of information abstractions in MR. The abstractions limit the information provided to the application, which may require modifications on the processing to provide the same level of service as the original scenario with more information.

Broad Considerations Now, as we desire to be ready for an MR *near*-future, we also ought to consider the various challenges involving the users, developers and device manufacturers, and policymakers.

- **Users.** As the number of users of MR continues to grow, it is imperative that users are made aware of the sensitivity and specificity of the data that is collected from their environment using MR devices. The risks and extent of data collection (not just for 3D MR data) has to be communicated clearly to users. Likewise, users themselves have to take part in the privacy discussion, especially as data protection policies are now being enforced by governments around the world.

  However, the finer-grained (e.g., object-level) user privacy preference specification can extraneously load users. This can become unreasonably taxing, which may defeat the intention of protection. Thus, the perceived utility of the user should also be considered in quantifying the overall quality of experience (QoE).
- **Developers and Device Manufacturers.** Both developers and device manufacturers should also understand the sensitivity of the data that their applications and devices are collecting and handling. For instance, Facebook has recently followed a similar approach unveiling a research-only MR device for the purpose of understanding the breadth and depth of MR challenges before developing and releasing true MR services to everyday users. (Facebook's Project Aria: https://about.fb.com/news/2020/09/announcing-project-aria-a-research-project-on-the-future-of-wearable-ar/) Also, they should be cognizant of the potential reservations that users may have once the users are made aware of the potential risks while using MR services and provide ways for users to either dictate the amount of information they provide or outright prevent collection of information as they desire.
- **Policymakers.** As of 2020, according to the UN Conference on Trade and Development (UNCTAD), 128 out of 194 countries had put legislation that protects user data privacy (https://unctad.org/page/data-protection-and-privacy-legislation-worldwide). Among these is the EU-General Data Protection Regulation (EU-GDPR) whose implementation in May 2018 caused major changes in the privacy policies for most applications worldwide even for non-EU consumers. Despite these significant progresses in legislation and policy enforcement, the actual necessary privacy and security technology remain to either be formulated or developed. Having a shared ontology and taxonomy of the security and privacy terminologies can help in the facilitation, translation, and enforcement of these policies through actual security- and privacy-enhancing technologies. Of course, there are also inherent differences in what is acceptable privacy and security risk that varies between societies and individuals. These variations can potentially complicate the enforcement of the policies and, thus, the implementation of the necessary protection technology.

Nonetheless, it is important to note that these broad challenges are not unique to MR. Likewise with other emerging technologies, we—whether we are from academia, government, or industry—ought to look towards how well-established technologies interact with society in striking a balance between societal and technological progresses.

7.4 Towards Everyday MR Services

MR allows users to interact with machines and each other in a totally different manner: for example, using gestures in the air instead of swiping on screens or tapping on keys. The output of our interactions, also, will no longer be confined within a screen. Instead, outputs will now be mixed with our real-world experience, and soon, we may not be able to tell what is real and what is synthetic. Recently released MR devices such as the Microsoft's HoloLens and the Magic Leap demonstrates what they can do. They allow users to interact with holographic augmentations in a more seamless and direct manner. In prospect, these MR services are likely to be integrated with existing platforms, such as smart phones (which is already happening), wearable devices, and other IoT devices, and will extend, if not enhance, the capabilities of these devices and platforms.

However, (**i**) as MR devices allow for these various novel services, and (**ii**) as novel technologies are integrated to the MR device to provide such services, the security and privacy risks inherent to these services and technologies are then inherited by the MR system. Moreover, the combination of the services and technologies in MR may give rise to novel security and privacy risks.

Summary In this chapter, we have collected, categorized, and reviewed various security and privacy works on MR. Table 7.3 summarizes and presents an overview of all the approaches that have been discussed for every aspect. It also presents a comparison based on which security and privacy properties they address (as defined in Table 7.2). From the summary table, we can see that most of the approaches are focused on input protection. This is understandable as it is more apparent to implement protection at the input side to prevent the capture of sensitive information or, in other input protection approaches, perform post-capture protection before providing the information to the application or service. However, we reiterate that to provide holistic protection, it is also equally important to look into the risks on the other aspects and design protection mechanisms to mitigate these risks. A crucial example is how output-side adversaries can potentially impose harm to users by delivering malicious outputs.

Moreover, the summary table also shows the distribution of the properties addressed by the different approaches. Input protection approaches are privacy-leaning, while the approaches for the other aspects, especially that of device aspect, are security-leaning. This further highlights the need for a synergistic approach to security and privacy in MR. That is, in order to provide protection that addresses most, if not all, of these properties, we need to provide protection for all aspects.

Furthermore, we have highlighted the remaining gaps for every aspect, as well as the broader challenges that MR face. Among the challenges were the necessary risk assessment of other MR devices and platforms, recommendations in the **research methodology** for evaluating performance of protection measures, and the broader considerations when it comes to policy and enforcement. For the latter, the apparent recommendation was the establishment of a shared ontology or taxonomy of security and privacy terminologies and definitions that can (i) facilitate the translation of policy to technology (whether the technology will be applied on the device, edge, or remote cloud or whether it will be), (ii) facilitate in the development of information abstractions for finer-grained policy enforcement, and (iii) allow users to have a common understanding of the risks and the permissions they are giving out.

***Near*-Future with MR** Arguably, ensuring security and privacy for future technologies can ensure the widespread user adoption of these technologies. MR presents a *near* future of new and immersive experiences that will inherently introduce security and privacy risks. Moreover, as MR devices are just starting to ship out commercially, there may still be unknown security and privacy risks.

Figure 7.15, a modified version of Fig. 7.1, shows a system-level recommendation on how to proceed with the provision of a holistic security and privacy protection for MR. Specifically, an intermediary protection layer as seen in Fig. 7.7 is inserted along the MR processing pipeline. Similar to the approaches discussed in 7.3, this proposed layer can provide heterogeneous methods of protection such as abstraction and data transformations. The chosen methods are dictated by the user privacy preferences. Extending the functions of this intermediary layer, we also integrate a means of calculating the risk, say, of user-identifiable information from the data collected by the MR device. Concurrently, a shared ontology can be used to define both the information abstractions and the risks. The same ontology can also be used to design the language of the user privacy preferences.

However, this proposed intermediary layer of protection is ultimately just a paradigm. As we proceed to implement this proposed paradigm, various implementations can arise. It is also important to note that as we proceed with these implementations, novel security and privacy challenges can also arise. Thus, it is important that most subsequent developments be pursued with synergy.

Acknowledgments We would like to thank Kwon Nung Choi for working on the packet analysis shown in Fig. 7.4. We also would like to thank Facebook Reality Labs for partially funding this work through Facebook Research Awards 2020.

Table 7.3 Summary of MR approaches that have been discussed, and which security and privacy properties are addressed by each approach and to what extent

Approach	Integrity	Non-Repudiation	Availability	Authorization	Authentication	Identification	Confidentiality	Anonymity	Unlinkability	Undetectability	Deniability	Awareness	Compliance
Interaction Protection Approaches													
Darkly [20][a]		✓		✓✓			✓	✓		✓		✓✓	✓
Context-based sanitization [24][a]				✓			✓	✓		✓		✓✓	✓
PlaceAvoider [23][b]				✓			✓	✓		✓		✓✓	✓
3D humanoids replace humans [111][a]							✓	✓		✓		✓	✓
OpenFace/RTFace [112][b]							✓	✓		✓		✓	✓
Capture-resistant spaces [27][c]				✓			✓	✓		✓			✓
See-through vision [113][a]				✓			✓	✓		✓		✓	✓
World-driven access control [28][a]				✓			✓	✓		✓		✓	✓
iRyp [31][a]				✓			✓	✓		✓		✓	✓
I-pic [29][b]				✓			✓	✓		✓		✓	✓
PrivacyCamera [30][b]				✓			✓	✓		✓		✓	✓
Cardea [25][b]				✓			✓	✓		✓		✓	✓
MarkIt [21,22][b]				✓			✓	✓		✓		✓	✓
PrivacyEye [26][b]				✓			✓	✓		✓		✓	✓
PrePose [34][a]		✓		✓✓			✓✓			✓			✓
Recognizers [32][a]		✓		✓✓			✓✓	✓		✓		✓	✓
SafeMR [37,38][a]		✓		✓✓			✓✓	✓		✓		✓	✓
AR-Silos [36][a]		✓		✓✓			✓✓	✓		✓		✓	✓
SemaDroid [104][c]		✓		✓✓			✓			✓		✓	✓
HE-sift [41][c]	✓		✓				✓✓			✓	✓		
Leveled HE-sift [44][c]	✓		✓				✓✓			✓	✓		

CryptoImg [45][c]	✓						✓✓			✓	✓		
SecSIFT [42, 114][c]	✓						✓✓			✓	✓		
P3 [46][c]	✓						✓✓			✓	✓		
Cloth Try-on [47][a]				✓			✓✓	✓	✓	✓✓	✓		
k-anonymous faces [48–50][b]							✓✓	✓✓	✓✓	✓✓	✓		
GAN-based face de-identification [51–53] [b]							✓✓	✓✓	✓✓	✓✓	✓		
Privacy-preserving pose estimation [56][a]							✓✓	✓✓	✓✓	✓✓	✓		
Conservative Plane Releasing [58][a]							✓✓	✓✓	✓✓	✓✓	✓		
SSAC [103][a]	✓	✓	✓	✓✓			✓					✓	✓✓
Output Protection Approaches													
Arya [60–62][a]	✓	✓	✓	✓✓			✓					✓	✓✓
SurroundWeb [33, 35][a]				✓			✓	✓		✓✓		✓	✓✓
EyeGuide [64][a]				✓✓			✓✓			✓✓			
VR Codes [65][a]			✓				✓			✓✓			
Psycho-visual modulation [67][a]			✓				✓			✓✓			
Visual cryptography using AR [70][a]							✓✓			✓✓			
Secret messages using AR [71][a]							✓✓			✓✓			
Mobile visual cryptography [115][a]			✓	✓✓			✓	✓					
Interaction Protection Approaches													
Emmie [74, 75][a]				✓			✓			✓		✓	✓✓
Kinected Conference [79][a]							✓	✓		✓		✓	✓✓
SecSpace [80][a]		✓	✓			✓	✓					✓✓	✓✓
Candid Interactions [81][c]		✓	✓			✓	✓					✓	✓

(continued)

Table 7.3 (continued)

Approach	Integrity	Non-Repudiation	Availability	Authorization	Authentication	Identification	Confidentiality	Anonymity	Unlinkability	Undetectability	Deniability	Awareness	Compliance
PIP [82, 116][a]				✓			✓			✓		✓	✓
TOUCHSPACE [83][a]				✓			✓			✓		✓	✓
BRAGFISH [84][a]				✓			✓			✓		✓	✓
LooksGoodToMe [86][b]	✓	✓✓	✓	✓	✓✓	✓✓	✓✓					✓	
HOLOPAIR [87], TAPPAIR [88][b]	✓	✓✓	✓	✓	✓✓	✓✓	✓✓					✓	
Device Protection Approaches													
Seamless and secure VR [117][a]					✓✓	✓	✓						
Mid-air authentication gestures [90][b]				✓	✓✓	✓	✓						
Head and blinking gestures [91][b]				✓	✓✓	✓	✓						
HEADBANGER [92][a]				✓	✓✓	✓	✓						
PSKA [94][c]				✓	✓✓	✓	✓						
SKULLCONDUCT [95][b]				✓	✓✓	✓	✓						
Facial multi-modal authentication [96][b]				✓	✓✓	✓	✓						
GAZETOUCHPASS [97][b]				✓	✓✓	✓	✓						
Polarization [100][a]							✓✓			✓✓			
TAPS Widget [101][c]				✓			✓			✓			
Chameleon-like [102][c]				✓			✓			✓			
EYEDECRYPT [98][a]	✓			✓✓		✓	✓✓						
Preventing keystroke inference [99][a]	✓			✓✓			✓						

The extent of each approach was either ✓✓ significantly addressing or ✓ partially addressing the security and privacy properties.
The approaches have been applied to either an [a]MR context, a [b]proto-MR context, or a [c]non-MR context.

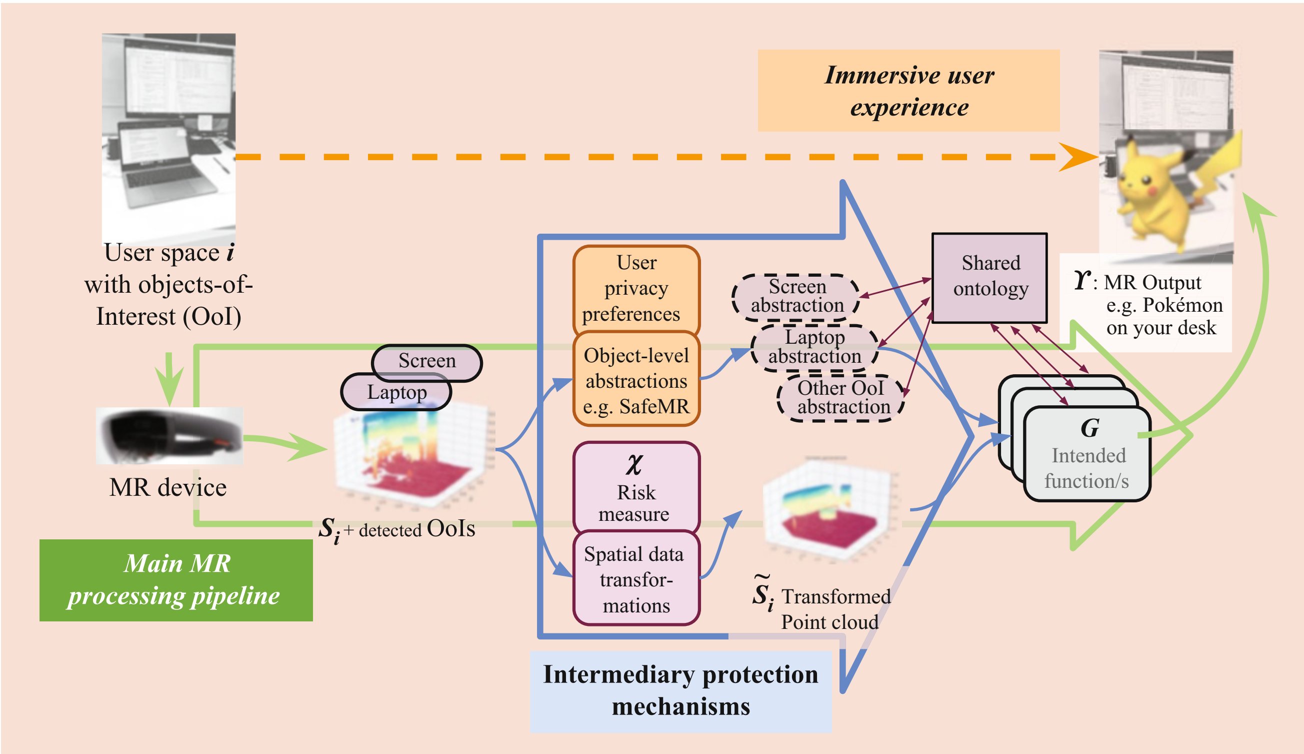


Fig. 7.15 A system diagram showing an *intermediary* layer of protection

References

1. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **77**(12), 1321–1329 (1994)
2. Benford, S., Brown, C., Reynard, G., Greenhalgh, C.: Shared spaces: transportation, artificiality, and spatiality (1996)
3. Azuma, R.T.: A survey of augmented reality. Presence Teleoperators Virtual Environ. **6**(4), 355–385 (1997)
4. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graphics Appl. **21**(6), 34–47 (2001)
5. Rabbi, I., Ullah, S.: A survey on augmented reality challenges and tracking. Acta Graphica znanstveni časopis za tiskarstvo i grafičke komunikacije **24**(1–2), 29–46 (2013)
6. Heimo, O.I., Kimppa, K.K., Helle, S., Korkalainen, T., Lehtonen, T.: Augmented reality-towards an ethical fantasy? (2014)
7. Friedman, B., Kahn Jr, P.H.: New directions: a value-sensitive design approach to augmented reality (2000)
8. Baldassi, S., Kohno, T., Roesner, F., Tian, M.: Challenges and new directions in augmented reality, computer security, and neuroscience-part 1: Risks to sensation and perception (2018). arXiv preprint arXiv:1806.10557
9. Roesner, F., Kohno, T., Molnar, D.: Security and privacy for augmented reality systems. Commun. ACM **57**(4), 88–96 (2014a)
10. de Guzman, J.A., Thilakarathna, K., Seneviratne, A.: Security and privacy approaches in mixed reality: A literature survey. ACM Comput. Surv. **52**(6), 110:1–110:37 (2019d)
11. Qiao, X., Ren, P., Dustdar, S., Liu, L., Ma, H., Chen, J.: Web ar: A promising future for mobile augmented reality-state of the art, challenges, and insights. Proc. IEEE **107**(4), 651–666 (2019)
12. Villari, M., Fazio, M., Dustdar, S., Rana, O., Ranjan, R.: Osmotic computing: a new paradigm for edge/cloud integration. IEEE Cloud Comput. **3**(6), 76–83 (2016)
13. Zhang, W., Han, B., Hui, P.: Jaguar: Low latency mobile augmented reality with flexible tracking. In: Proceedings of the 26th ACM international conference on Multimedia (2018)
14. Liu, L., Li, H., Gruteser, M.: Edge assisted real-time object detection for mobile augmented reality. The 25th Annual International Conference on Mobile Computing and Networking (MobiCom). ACM, New York (2019)
15. Qiu, H., Ahmad, F., Bai, F., Gruteser, M., Govindan, R.: Avr: Augmented vehicular reality (2018)
16. Liu, L., Zhong, R., Zhang, W., Liu, Y., Zhang, J., Zhang, L., Gruteser, M.: Cutting the cord: Designing a high-quality untethered vr system with low latency remote rendering. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (2018)
17. Oh, S.J., Benenson, R., Fritz, M., Schiele, B.: Faceless person recognition: Privacy implications in social media. In: European Conference on Computer Vision. Springer, Cham (2016)
18. Acquisti, A.: Privacy in the age of augmented reality. In: Proceedings of the National Academy of Sciences (2011)
19. Lebeck, K., Ruth, K., Kohno, T., Roesner, F.: Towards security and privacy for multi-user augmented reality: foundations with end users. In: Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP). IEEE, New York (2018)
20. Jana, S., Narayanan, A., Shmatikov, V.: A scanner darkly: Protecting user privacy from perceptual applications. In: Proceedings of the 2013 IEEE symposium on security and privacy. IEEE, New York (2013b)

21. Raval, N., Srivastava, A., Lebeck, K., Cox, L., Machanavajjhala, A.: Markit: Privacy markers for protecting visual secrets. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication (2014)
22. Raval, N., Srivastava, A., Razeen, A., Lebeck, K., Machanavajjhala, A., Cox, L.P.: What you mark is what apps see. In: Proceedings of the 14th annual international conference on mobile systems, applications, and services (2016)
23. Templeman, R., Korayem, M., Crandall, D.J., Kapadia, A.: Placeavoider: steering first-person cameras away from sensitive spaces. In: NDSS (2014)
24. Zarepour, E., Hosseini, M., Kanhere, S.S., Sowmya, A.: A context-based privacy preserving framework for wearable visual lifeloggers. In: 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). IEEE, New York (2016)
25. Shu, J., Zheng, R., Hui, P.: Cardea: Context-aware visual privacy protection from pervasive cameras. arXiv preprint arXiv:1610.00889 (2016)
26. Steil, J., Koelle, M., Heuten, W., Boll, S., Bulling, A.: Privaceye: Privacy-preserving first-person vision using image features and eye movement analysis. CoRR, abs/1801.04457 (2018)
27. Truong, K., Patel, S., Summet, J., Abowd, G.: Preventing camera recording by designing a capture-resistant environment. In: UbiComp 2005: Ubiquitous Computing, pp. 903–903 (2005)
28. Roesner, F., Molnar, D., Moshchuk, A., Kohno, T., Wang, H.J.: World-driven access control for continuous sensing. In: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (2014b)
29. Aditya, P., Sen, R., Druschel, P., Joon Oh, S., Benenson, R., Fritz, M., Schiele, B., Bhattacharjee, B., Wu, T.T.: I-pic: A platform for privacy-compliant image capture. In: Proceedings of the 14th annual international conference on mobile systems, applications, and services (2016)
30. Li, A., Li, Q., Gao, W.: Privacycamera: Cooperative privacy-aware photographing with mobile phones. In: 2016 13th Annual IEEE international conference on sensing, communication, and networking (SECON). IEEE, New York (2016a)
31. Sun, Y., Chen, S., Zhu, S., Chen, Y.: iRyP: a purely edge-based visual privacy-respecting system for mobile cameras. In: Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks (2020)
32. Jana, S., Molnar, D., Moshchuk, A., Dunn, A.M., Livshits, B., Wang, H.J., Ofek, E.: Enabling fine-grained permissions for augmented reality applications with recognizers. In: 22nd USENIX Security Symposium (USENIX Security 13) (2013a)
33. Vilk, J., Molnar, D., Ofek, E., Rossbach, C., Livshits, B., Moshchuk, A., Wang, H.J., Gal, R.: Least privilege rendering in a 3d web browser. In: Microsoft Research Technical Report MSR-TR-2014-25 (2014)
34. Figueiredo, L.S., Livshits, B., Molnar, D., Veanes, M.: Prepose: Privacy, security, and reliability for gesture-based programming. In: 2016 IEEE Symposium on Security and Privacy (SP), pp. 122–137. IEEE, New York (2016)
35. Vilk, J., Molnar, D., Livshits, B., Ofek, E., Rossbach, C., Moshchuk, A., Wang, H.J., Gal, R.: Surroundweb: Mitigating privacy concerns in a 3d web browser. In: 2015 IEEE Symposium on Security and Privacy. IEEE, New York (2015)
36. Jensen, J., Hu, J., Rahmati, A., LiKamWa, R.: Protecting visual information in augmented reality from malicious application developers. In: The 5th ACM Workshop on Wearable Systems and Applications (2019)
37. de Guzman, J.A., Thilakarathna, K., Seneviratne, A.: Safemr: Privacy-aware visual information protection for mobile mixed reality. In: 2019 IEEE 44th Conference on Local Computer Networks (LCN). IEEE, New York (2019c)
38. de Guzman, J.A., Thilakarathna, K., Seneviratne, A.: Demo: Privacy-aware visual information protection for mobile mixed reality. In: 2019 IEEE 44th Conference on Local Computer Networks (LCN). IEEE, New York (2019a)
39. Hsu, C.-Y., Lu, C.-S., Pei, S.-C.: Homomorphic encryption-based secure sift for privacy-preserving feature extraction. In: Media Watermarking, Security, and Forensics III, vol. 7880. International Society for Optics and Photonics, New York (2011)
40. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004)
41. Jiang, L., Xu, C., Wang, X., Luo, B., Wang, H.: Secure outsourcing sift: Efficient and privacy-preserving image feature extraction in the encrypted domain. IEEE Trans. Dependable Secure Comput. **17**(1), 179–193 (2017)
42. Qin, Z., Yan, J., Ren, K., Chen, C.W., Wang, C.: Towards efficient privacy-preserving image feature extraction in cloud computing. In: Proceedings of the 22nd ACM international conference on multimedia (2014a)
43. Qin, Z., Yan, J., Ren, K., Chen, C.W., Wang, C.: Secsift: Secure image sift feature extraction in cloud computing. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **12**(4s), 65 (2016)
44. Zhang, L., Jung, T., Feng, P., Li, X.-Y., Liu, Y.: Cloud-based privacy preserving image storage, sharing and search. arXiv preprint arXiv:1410.6593 (2014)
45. Ziad, M.T.I., Alanwar, A., Alzantot, M., Srivastava, M.: Cryptoimg: Privacy preserving processing over encrypted images. In: 2016 IEEE Conference on Communications and Network Security (CNS) (2016)
46. Ra, M.-R., Govindan, R., Ortega, A.: P3: Toward privacy-preserving photo sharing. In: 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13) (2013)
47. Sekhavat, Y.A.: Privacy preserving cloth try-on using mobile augmented reality. IEEE Trans. Multimedia **19**(5), 1041–1049 (2017)
48. Newton, E.M., Sweeney, L., Malin, B.: Preserving privacy by de-identifying face images. IEEE Trans. Knowl. Data Eng. **17**(2), 232–243 (2005)
49. Gross, R., Sweeney, L., De la Torre, F., Baker, S.: Model-based face de-identification. In: 2006 Conference on computer vision and pattern recognition workshop (CVPRW'06). IEEE, New York (2006)
50. Gross, R., Sweeney, L., de la Torre, F., Baker, S.: Semi-supervised learning of multi-factor models for face de-identification. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition (2008)
51. Du, L., Yi, M., Blasch, E., Ling, H.: Garp-face: Balancing privacy protection and utility preservation in face de-identification. In: IEEE International Joint Conference on Biometrics. IEEE, New York (2014)
52. Brkic, K., Sikiric, I., Hrkac, T., Kalafatic, Z.: I know that person: Generative full body and face de-identification of people in images. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, New York (2017)
53. Wu, Y., Yang, F., Ling, H.: Privacy-protective-gan for face de-identification. arXiv preprint arXiv:1806.08906 (2018)
54. Waegel, K.: [poster] a reconstructive see-through display (2014)
55. Pittaluga, F., Koppal, S.J., Kang, S.B., Sinha, S.N.: Revealing scenes by inverting structure from motion reconstructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019)
56. Speciale, P., Schonberger, J.L., Kang, S.B., Sinha, S.N., Pollefeys, M.: Privacy preserving image-based localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019)
57. de Guzman, J.A., Thilakarathna, K., Seneviratne, A.: A first look into privacy leakage in 3d mixed reality data. In: European

Symposium on Research in Computer Security. Springer, Cham (2019b)
58. de Guzman, J.A., Thilakarathna, K., Seneviratne, A.: Conservative plane releasing for spatial privacy protection in mixed reality. arXiv preprint arXiv:2004.08029 (2020)
59. Casey, P., Baggili, I., Yarramreddy, A.: Immersive virtual reality attacks and the human joystick. In: IEEE Transactions on Dependable and Secure Computing (2019)
60. Lebeck, K., Kohno, T., Roesner, F.: How to safely augment reality: Challenges and directions. In: Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications (2016)
61. Lebeck, K., Ruth, K., Kohno, T., Roesner, F.: Securing augmented reality output. In: 2017 IEEE symposium on security and privacy (SP). IEEE, New York (2017)
62. Lebeck, K., Kohno, T., Roesner, F.: Enabling multiple applications to simultaneously augment reality: Challenges and directions. In: Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications (2019)
63. Ahn, S., Gorlatova, M., Naghizadeh, P., Chiang, M., Mittal, P.: Adaptive fog-based output security for augmented reality. In: Proceedings of the 2018 Morning Workshop on Virtual Reality and Augmented Reality Network (2018)
64. Eaddy, M., Blasko, G., Babcock, J., Feiner, S.: My own private kiosk: Privacy-preserving public displays. In: Eighth International Symposium on Wearable Computers, vol. 1. IEEE, New York (2004)
65. Woo, G., Lippman, A., Raskar, R.: Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In: 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, New York (2012)
66. Yerazunis, W., Carbone, M.: Privacy-enhanced displays by time-masking images. In: Australian Conference on Computer-Human Interaction (2002)
67. Lin, P.-Y., You, B., Lu, X.: Video exhibition with adjustable augmented reality system based on temporal psycho-visual modulation. EURASIP J. Image Video Process. **2017**(1), 7 (2017)
68. Chang, J.J.-Y., Li, M.-J., Wang, Y.-C., Juan, J.S.-T.: Two-image encryption by random grids. In: 2010 10th International Symposium on Communications and Information Technologies. IEEE, New York (2010)
69. Simkin, M., Schröder, D., Bulling, A., Fritz, M.: Ubic: Bridging the gap between digital cryptography and the physical world. In: European Symposium on Research in Computer Security. Springer, Cham (2014)
70. Lantz, P., Johansson, B., Hell, M., Smeets, B.: Visual cryptography and obfuscation: A use-case for decrypting and deobfuscating information using augmented reality. In: International Conference on Financial Cryptography and Data Security. Springer, Berlin (2015)
71. Andrabi, S.J., Reiter, M.K., Sturton, C.: Usability of augmented reality for revealing secret messages to users but not their devices. In: Eleventh Symposium On Usable Privacy and Security (SOUPS 2015) (2015)
72. Benford, S., Greenhalgh, C., Reynard, G., Brown, C., Koleva, B.: Understanding and constructing shared spaces with mixed-reality boundaries. ACM Trans. Comput.-Hum. Interact. (TOCHI) **5**(3), 185–223 (1998)
73. Scott, S.D., Carpendale, M.S.T., Inkpen, K.M.: Territoriality in collaborative tabletop workspaces. In: Proceedings of the 2004 ACM conference on Computer supported cooperative work (2004)
74. Butz, A., Beshers, C., Feiner, S.: Of vampire mirrors and privacy lamps: Privacy management in multi-user augmented environments. In: Proceedings of the 11th annual ACM symposium on User interface software and technology (1998)
75. Butz, A., Höllerer, T., Feiner, S., MacIntyre, B., Beshers, C.: Enveloping users and computers in a collaborative 3d augmented reality. In: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99). IEEE, New York (1999)
76. Wu, M., Balakrishnan, R.: Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In: Proceedings of the 16th annual ACM symposium on User interface software and technology (2003)
77. Morris, M.R., Cassanego, A., Paepcke, A., Winograd, T., Piper, A.M., Huang, A.: Mediating group dynamics through tabletop interface design. IEEE Comput. Graphics Appl. **26**(5), 65–73 (2006a)
78. Morris, M.R., Huang, A., Paepcke, A., Winograd, T.: Cooperative gestures: multi-user gestural interactions for co-located groupware. In: Proceedings of the SIGCHI conference on Human Factors in computing systems (2006b)
79. DeVincenzi, A., Yao, L., Ishii, H., Raskar, R.: Kinected conference: augmenting video imaging with calibrated depth and audio. In: Proceedings of the ACM 2011 conference on Computer supported cooperative work (2011)
80. Reilly, D., Salimian, M., MacKay, B., Mathiasen, N., Edwards, W.K., Franz, J.: Secspace: prototyping usable privacy and security for mixed reality collaborative environments. In: Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems (2014)
81. Ens, B., Grossman, T., Anderson, F., Matejka, J., Fitzmaurice, G.: Candid interaction: Revealing hidden mobile and wearable computing activities. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (2015)
82. Szalavári, Z., Eckstein, E., Gervautz, M.: Collaborative gaming in augmented reality. In: Proceedings of the ACM symposium on Virtual reality software and technology (1998)
83. Cheok, A.D., Yang, X., Ying, Z.Z., Billinghurst, M., Kato, H.: Touch-space: Mixed reality game space based on ubiquitous, tangible, and social computing. Personal Ubiquitous Comput. **6**(5–6), 430–442 (2002)
84. Xu, Y., Gandy, M., Deen, S., Schrank, B., Spreen, K., Gorbsky, M., White, T., Barba, E., Radu, I., Bolter, J., et al.: Bragfish: exploring physical and social interaction in co-located handheld augmented reality games. In: Proceedings of the 2008 international conference on advances in computer entertainment technology (2008)
85. Henrysson, A., Billinghurst, M., Ollila, M.: Face to face collaborative ar on mobile phones. In: Fourth ieee and acm international symposium on mixed and augmented reality (ISMAR'05). IEEE, New York (2005)
86. Gaebel, E., Zhang, N., Lou, W., Hou, Y.T.: Looks good to me: Authentication for augmented reality. In: Proceedings of the 6th international workshop on trustworthy embedded devices (2016)
87. Sluganovic, I., Serbec, M., Derek, A., Martinovic, I.: Holopair: Securing shared augmented reality using microsoft hololens. In: Proceedings of the 33rd annual computer security applications conference (2017)
88. Sluganovic, I., Liskij, M., Derek, A., Martinovic, I.: Tap-pair: Using spatial secrets for single-tap device pairing of augmented reality headsets. In: Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy (2020)
89. Dickinson, B.: 5 authentication methods putting passwords to shame (2016)
90. Aslan, I., Uhl, A., Meschtscherjakov, A., Tscheligi, M.: Mid-air authentication gestures: an exploration of authentication based on palm and finger motions. In: Proceedings of the 16th International Conference on Multimodal Interaction (2014)
91. Rogers, C.E., Witt, A.W., Solomon, A.D., Venkatasubramanian, K.K.: An approach for user identification for head-mounted displays. In: Proceedings of the 2015 ACM International Symposium on Wearable Computers (2015)

92. Li, S., Ashok, A., Zhang, Y., Xu, C., Lindqvist, J., Gruteser, M.: Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns. In: 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, New York (2016b)
93. Chauhan, J., Hu, Y., Seneviratne, S., Misra, A., Seneviratne, A., Lee, Y.: Breathprint: Breathing acoustics-based user authentication. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (2017)
94. Venkatasubramanian, K.K., Banerjee, A., Gupta, S.K.S.: Pska: Usable and secure key agreement scheme for body area networks. IEEE Trans. Inf. Technol. Biomed. **14**(1), 60–68 (2010)
95. Schneegass, S., Oualil, Y., Bulling, A.: Skullconduct: Biometric user identification on eyewear computers using bone conduction through the skull. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2016)
96. Raja, K.B., Raghavendra, R., Stokkenes, M., Busch, C.: Multimodal authentication system for smartphones using face, IRIS and periocular. In: International Conference on Biometrics (ICB), pp. 143–150. IEEE, New York (2015)
97. Khamis, M., Alt, F., Hassib, M., von Zezschwitz, E., Hasholzner, R., Bulling, A.: Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (2016)
98. Forte, A.G., Garay, J.A., Jim, T., Vahlis, Y.: Eyedecrypt—private interactions in plain sight. In: International Conference on Security and Cryptography for Networks. Springer, Cham (2014)
99. Maiti, A., Jadliwala, M., Weber, C.: Preventing shoulder surfing using randomized augmented reality keyboards. In: 2017 IEEE international conference on pervasive computing and communications workshops (PerCom Workshops) (2017)
100. Kohno, T., Kollin, J., Molnar, D., Roesner, F.: Display leakage and transparent wearable displays: Investigation of risk, root causes, and defenses. Technical report (2016)
101. Möllers, M., Borchers, J.: TaPS widgets: interacting with tangible private spaces. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (2011)
102. Pearson, J., Robinson, S., Jones, M., Joshi, A., Ahire, S., Sahoo, D., Subramanian, S.: Chameleon devices: investigating more secure and discreet mobile interactions via active camouflaging. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (2017)
103. Rubio-Medrano, C.E., Jogani, S., Leitner, M., Zhao, Z., Ahn, G.-J.: Effectively enforcing authorization constraints for emerging space-sensitive technologies. In: Proceedings of the 24th ACM Symposium on Access Control Models and Technologies (2019)
104. Xu, Z., Zhu, S.: Semadroid: A privacy-aware sensor management framework for smartphones. In: Proceedings of the 5th ACM Conference on Data and Application Security and Privacy (2015)
105. Ruth, K., Kohno, T., Roesner, F.: Secure multi-user content sharing for augmented reality applications. In: 28th USENIX Security Symposium (USENIX Security 19) (2019)
106. Guzman, J.A.d., Seneviratne, A., Thilakarathna, K.: Unravelling spatial privacy risks of mobile mixed reality data. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5(1), pp. 1–26 (2021)
107. Arthur, C.: Google glass security failings may threaten owner's privacy. In: The Guardian: Dostupnona. https://www.theguardian.com/technology/2013/may/01/google-glass-security-privacyrisk [pristupljeno: 08.08. 2019] (2013)
108. Denning, T., Dehlawi, Z., Kohno, T.: In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2014)
109. Siriwardhana, Y., Porambage, P., Liyanage, M., Ylinattila, M.: A survey on mobile augmented reality with 5g mobile edge computing: Architectures, applications and technical aspects. IEEE Commun. Surv. Tutorials **23**(2), 1160–1192 (2021)
110. Roman, R., Lopez, J., Mambo, M.: Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. Future Gener. Comput. Syst. **78**, 680–698 (2018)
111. Szczuko, P.: Augmented reality for privacy-sensitive visual monitoring. In: International Conference on Multimedia Communications, Services and Security. Springer, Cham (2014)
112. Wang, J., Amos, B., Das, A., Pillai, P., Sadeh, N., Satyanarayanan, M.: A scalable and privacy-aware iot service for live video analytics. In: Proceedings of the 8th ACM on Multimedia Systems Conference (2017)
113. Hayashi, M., Yoshida, R., Kitahara, I., Kameda, Y., Ohta, Y.: An installation of privacy-safe see-through vision. Procedia-Social Behav. Sci. **2**(1), 125–128 (2010)
114. Qin, Z., Yan, J., Ren, K., Chen, C.W., Wang, C., Fu, X.: Privacy-preserving outsourcing of image global feature detection. In: 2014 IEEE global communications conference. IEEE, New York (2014b)
115. Fang, C., Chang, E.-C.: Securing interactive sessions using mobile device through visual channel and visual inspection. In: Proceedings of the 26th Annual Computer Security Applications Conference (2010)
116. Szalavári, Z., Gervautz, M.: The personal interaction panel-a two-handed interface for augmented reality. In: Computer graphics forum, vol. 16(3), pp. C335–C346. Blackwell Publishers Ltd., Oxford (1997)
117. George, C., Khamis, M., von Zezschwitz, E., Burger, M., Schmidt, H., Alt, F., Hussmann, H.: Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. In: NDSS (2017)
118. Mulloni, A., Wagner, D., Schmalstieg, D.: Mobility and social interaction as core gameplay elements in multi-player augmented reality. In: Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts, pp. 472–478. ACM (2008)

Jaybie A. de Guzman has recently concluded his PhD at the UNSW, Sydney, last January 2021. While doing his PhD, he was a research student with the Information Security and Privacy group of CSIRO's Data6. He is currently working on spatial privacy and computer networks and is an Assistant Professor at the Electrical and Electronics Engineering Institute of the University of the Philippines in Diliman, Quezon City, Philippines.

Kanchana Thilakarathna received his PhD from the UNSW, Sydney, in 2015 with the Malcolm Chaikin Prize. He is currently working in cybersecurity and privacy at the University of Sydney. Prior to that, he worked at the information security and privacy group at Data61-CSIRO. He is a recipient of Facebook Research Awards in 2020 and Google Faculty Awards in 2018.

Aruna Seneviratne is currently a Foundation Professor of telecommunications at the University of New South Wales, Australia, where he is the Mahanakorn Chair of telecommunications. He has also worked at many other Universities in Australia, UK, and France and industrial organizations, including Muirhead, Standard Telecommunication Labs, Avaya Labs, and Telecom Australia (Telstra). In addition, he has held visiting appointments at INRIA, France, and has been awarded a number of fellowships, including one at the British Telecom and one at the Telecom Australia Research Labs. His current research interests are in physical analytics: technologies that enable applications to interact intelligently and securely with their environment in real time. Most recently, his team has been working on using these technologies in behavioral biometrics, optimizing the performance of wearables, and verifying the IoT system.

Part III

Hardware and Peripherals

The Optics of Augmented Reality Displays

8

Aaron Bauer and Jannick P. Rolland

Contents

8.1 Introduction to Augmented and Virtual Reality 187

8.2 A Brief History of AR Displays 188

8.3 The Basics of Visual Instrument Design 189
8.3.1 The Human Visual System 189
8.3.2 Optical Design Properties for AR Displays 190

8.4 Optical Components of an AR Display 193
8.4.1 Microdisplays as the Light Engine 193
8.4.2 Radiometric Brightness Analysis for AR Displays to Guide Microdisplay Specifications 194
8.4.3 A Brief Foray into Laser Scanning 195
8.4.4 Imaging Optics and Combiners 196

8.5 Optical Architectures and How They Work 196

8.6 Areas for Improvement 203

8.7 Components and Techniques for AR Displays of the Future 203
8.7.1 Pupil-Steering and Eye-Tracking 204
8.7.2 Freeform Optics 204
8.7.3 Metasurfaces 204

8.8 Conclusion 205

References 205

Abstract

Augmented reality (AR) is the next frontier for visual displays. In the optimal AR display, the mechanics, electronics, and optics must interact seamlessly. In this chapter, optical science concepts are developed to facilitate the reader's understanding of the optics found within current AR technology. Various optical architectures are being used in the current AR display technology, and those architectures will be dissected and discussed. A combination of physical, electrical, and optical constraints limits the capabilities of recent AR displays. As the details of the optical challenges in designing AR displays are examined, emerging technologies that could facilitate a fundamental change in design processes, such as holographic optics, freeform optics, and metasurfaces, will be introduced.

Keywords

Augmented reality · Freeform · Augmented reality design · Optical architectures · Head-worn display · Optical see-through displays

8.1 Introduction to Augmented and Virtual Reality

A historical change in the way that humans consume digital content is underway. Stationary, stand-alone displays, such as televisions and computer monitors, will soon join radio and newspapers as outdated media delivery devices. Similarly, the days of slouching over, with our heads and gazes pointed downward while being consumed by smartphones, may also be numbered. A new generation of display devices that provides never-before-experienced visual stimulation looms on the horizon. Virtual reality (VR) and augmented reality (AR) are poised for a digital display revolution. This chapter aims to discuss the optical properties of the display technologies required for such a revolution because, without a successful hardware implementation, it is more dream than reality.

Let us first distinguish between VR and AR. In a VR application, purely digital content replaces the user's view of the real world. This view contrasts with AR, where the user's view of the real world is either digitally or optically preserved, and digital content augments the user's field of vision. Examples of some AR experiences are illustrated in Fig. 8.1. The differences between VR and AR trace back to the hardware display system requirements, and, as such, the developments of AR and VR hardware are considered separately. Because AR preserves the real-world view, it has

A. Bauer (✉) · J. P. Rolland
The Institute of Optics, University of Rochester, Rochester, NY, USA
e-mail: aaron.bauer@rochester.edu; rolland@optics.rochester.edu

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_8

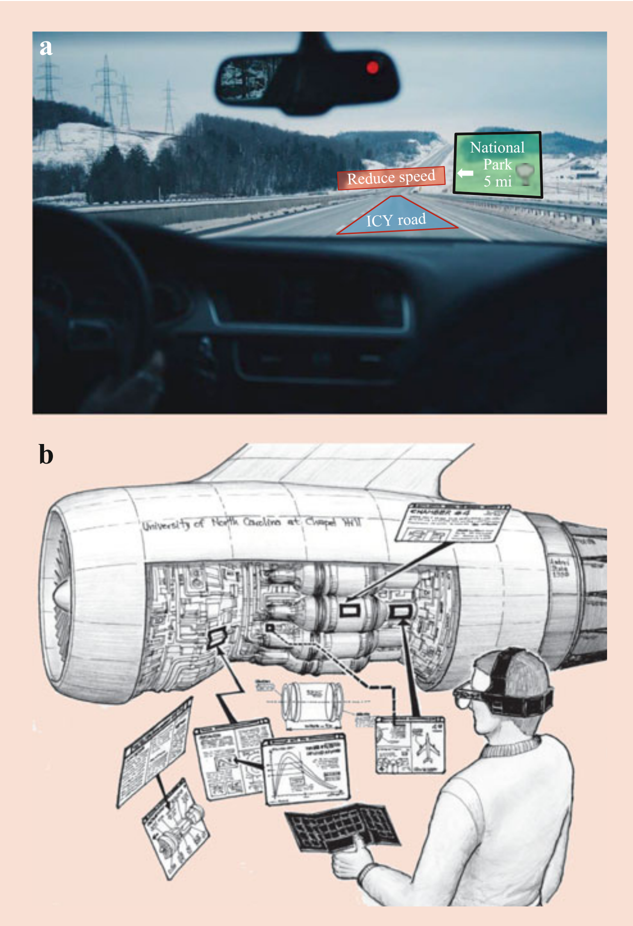


Fig. 8.1 (**a**) Virtual information (in this case, notifications while driving) is overlaid on the user's view of the real world, representing an example of user experience when using an AR display. (**b**) An AR application illustration for jet engine repair. (Courtesy of Andrei State)

the potential to create a more significant impact on our lives, as evidenced by its predicted $70–75 billion market by 2023 [1]. Thus, the optics of AR will be the focus in this chapter, with an emphasis on consumer devices.

8.2 A Brief History of AR Displays

Optical see-through AR displays have been in development since the 1960s. Ivan Sutherland's first computer graphics-driven display [2], coined "The Sword of Damocles" and recently given an artistic makeover shown in an article by Rolland [3], generated interest and drove research in the field of AR hardware. However, much of the early work on AR displays was done in the military, where it could be utilized in simulators and **helmet-mounted visualization** for flight assistance such as targeting and instrumentation readouts. In the 1980s, Honeywell developed a cathode ray tube (CRT) based AR display for use in the Apache military helicopter for integration with an infrared camera. This helmet-based display was named the Integrated Helmet and Display Sighting System (IHADSS), and it had a $40^\circ \times 30^\circ$ monocular field of view (FOV) with a 10 mm eyebox (allowing for 20% vignetting at full field) and integrated head tracking [4]. The success of the IHADSS system prompted other military branches to invest in programs to develop AR displays, such as the Helmet Integrated Display Sight System, or HIDSS, developed by Rockwell Collins for the RAH-66 Comanche military helicopters and, much later, the Q-Sight by BAE Systems, shown in Fig. 8.2 together with the IHADSS. Early military AR programs informed engineers of the need to supplement innovation in the hardware with a deep understanding of the human perception impacts of augmenting vision to mitigate undesirable side effects of prolonged use, such as nausea, dizziness, fatigue, or headaches [5–10]. Thorough reviews of the history of military helmet-mounted displays are found in Foote and Melzer [11] and Zhang et al. [12].

Many of the early AR systems utilized rotationally symmetric optics for the virtual image formation and a beamsplitter to merge the virtual and real worlds. The hefty beamsplitters brought the center of mass of early helmet-mounted AR displays too far forward, causing a safety issue for use in aircraft. Researchers in the late 1980s began to search for lighter-weight alternatives to the customary beamsplitter setup, resulting in the invention of the holographic off-axis combiner [13, 14]. Besides enabling lighter-weight optics, holographic combiners improved the real-world transmissivity and virtual image brightness by employing holograms optimized for the spectrum emitted by the CRT displays. However, the real-world view was slightly discolored from the subtraction of the CRT wavelengths from the real-world light, and the curved off-axis combiner had worse accuracy and parallax performance than the conventional approach. An exemplary ray trace for an early holographic combiner AR display is shown in Fig. 8.3.

Fast forward to 2001, when Microvision commercialized the first personal AR display that utilized a laser scanning system. Around this time, consumer products were being developed that featured miniaturized versions of the components used in the military AR displays. Notably, microdisplays and laser scanning systems replaced the heavy and bulky CRT displays. Both military and consumer development of the next generation of AR displays continued in tandem, but with utilitarian, helmet-based solutions for the military and comfortable, ergonomic solutions for consumers. This contrast in ideology is depicted in Fig. 8.4.

Fig. 8.2 (**a**) Integrated Helmet and Display Sight System by Honeywell and (**b**) the Q-Sight by BAE Systems. (Images from [12])

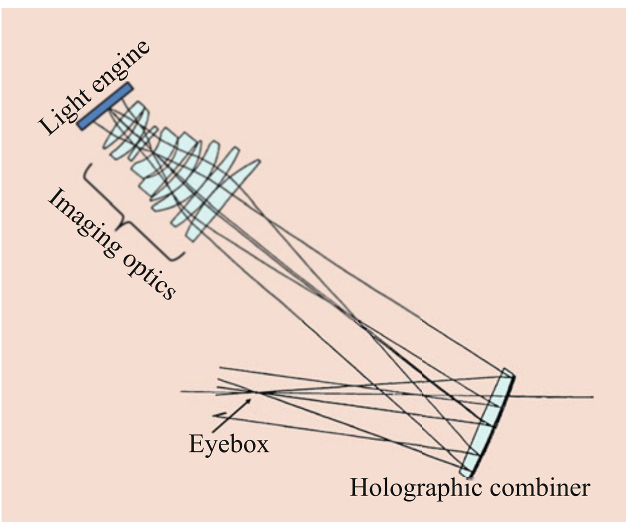


Fig. 8.3 Exemplary ray trace for an early AR display with a holographic combiner, the HGS-1000. (Adapted from [15])

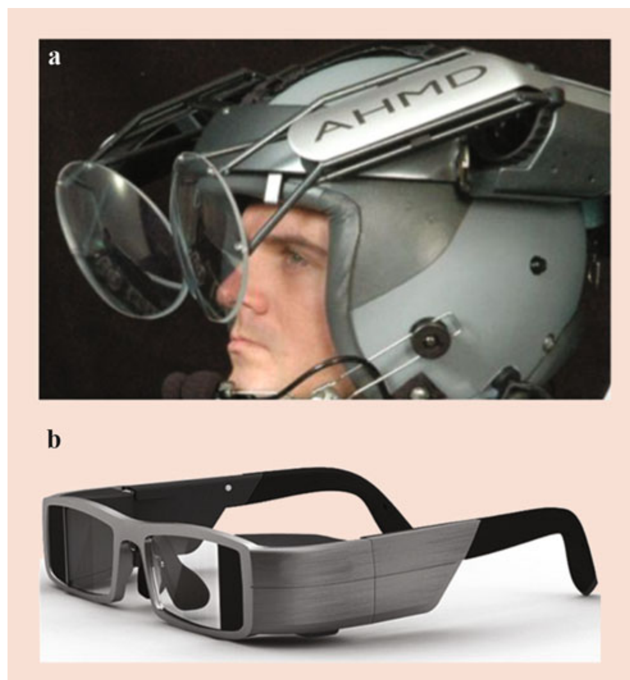


Fig. 8.4 Designed by Optical Research Associates (now Synopsys) and manufactured by Zygo and L3, the (**a**) Advanced Helmet Mounted Display emphasizes FOV and eyebox over system size. (Image from [16]). Consumer AR displays take the opposite path, emphasizing wearability and the eyeglasses form factor, as the (**b**) Lumus PD-18-2 demonstrates

8.3 The Basics of Visual Instrument Design

To understand the optical properties of AR displays, an understanding of how a generic optical device interfaces with the **human visual system** must first be developed. There is a long history of developing optical systems to interface with the human visual system (e.g., telescopes, microscopes, binoculars). The critical optical components of these systems can be identified and applied to the specific task of designing an AR display system.

8.3.1 The Human Visual System

When designing an optical system to mate with an existing optical system, it is critical to deeply understand the

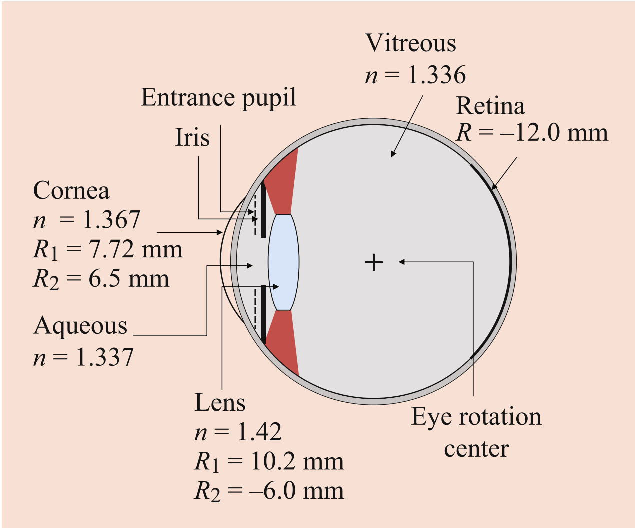


Fig. 8.5 A cross-section of the human eye with the various structures relevant for designing visual instruments labeled. (Illustration adapted from [17] with data from [18])

properties of the existing optical system. For an AR display as the to-be-designed optical system, the existing optical system to which it will mate is the human eye. The eye is a complex structure, but, while its full nature will not be explored here, a basic model of the eye is sufficient for examining the properties pertinent to optical design. The eye has components responsible for image formation and image detection, much like a camera, and can be treated as such when discussing its first-order imaging properties. Figure 8.5 shows a cross-section of an eye with the pertinent features for optical design called out.

The first optical element in the eye, the **cornea** , is located at the air-eye interface. It provides the eye with 75% of its total focusing power due to the significant refractive index difference at the air interface. Located approximately 3.6 mm behind the cornea is the eye's lens [18]. It is attached to muscles that can stretch or contract, allowing for dynamic focusing called accommodation. A typical relaxed eye is focused at infinity, with an accepted near point for a fully accommodated eye of 250 mm. Coincident with the most external surface of the lens is the iris, which acts as the eye's aperture stop. Like the lens, the iris is attached to muscles that can expand or contract, widening or narrowing the iris opening between 1 and 8 mm in diameter. The iris controls the overall amount of light entering the eye and the eye's f-number, which, in turn, affects the magnitude of the eye's imaging aberrations. Proceeding the lens is a large fluid-filled volume called the vitreous with an index of refraction near that of water. There are no additional focusing optics in the vitreous before the **retina** detects the light at the rear of the eye. The eye's center of rotation is approximately 11 mm behind the cornea, on average [19].

From an optical design perspective, the retina can be treated as a non-planar detector whose curvature matches the eyeball. The retina's "pixels" come in two varieties – rods and cones. Rods outnumber the cones by about 20:1 and are mainly responsible for low-resolution peripheral vision. The fovea, located near the retina center, contains about 15,000 densely packed cones spaced ~2.5 μm apart [17] and enables detailed inspection in the high-resolution center of the eye's FOV. To perceive a full crisp image, the eye rapidly scans this high-resolution patch over a target object.

Nyquist sampling theory can be used to estimate the maximum theoretical **resolution of the eye** . One contrast cycle requires a minimum of three pixels center to center to be detected. With a "pixel" size of 2.5 μm (cone spacing), the eye's theoretical maximum resolution is calculated to be 200 cycles/mm. It is expected that the eye can resolve ~70% of the theoretical limit after accounting for the randomness and nonuniformity of the cone spacing, assuming diffraction-limited eye optics (valid for an iris diameter of <2 mm) [17]. When further considering the aberrations of the eye optics, a realistic value for the resolution of the eye is 110 cycles/mm, which equates to roughly one arcminute in angle space [17]. Again, it is noted that this resolution is only at or near the center of the FOV. The resolution degrades rapidly with increasing FOV, as indicated in Fig. 8.6a. Further, the most vivid images originate in the green band of the visible spectrum, as inferred from the eye's photopic spectral response curve in Fig. 8.6b.

8.3.2 Optical Design Properties for AR Displays

Armed with an understanding of how the eye functions as an imaging system, a discussion on designing optical systems for use with the eye can begin. The goal of many optical systems that interface with the human visual system, including AR displays, is to present a magnified image to the eye. This type of optical system falls under the design class of eyepieces. A distinguishing feature of eyepieces is that they are designed with an external pupil. To avoid vignetting of the FOV when the eyepiece is being used, the optical system's external pupil must be coincident with the eye's entrance pupil, which is the image of the iris through the cornea and is located about 3.05 mm behind the cornea. Figure 8.7 illustrates the ray-clipping, or vignetting, that occurs if the eyepiece is not **pupil-matched** with the eye.

The concept of an **eyebox** goes hand-in-hand with pupil-matching. The eyebox of an optical system is the three-dimensional region where the user's eye can be placed such that there is acceptable vignetting of the FOV and the image quality meets specification [21]. The eyebox is typically designed to be larger than the pupil of the eye to allow

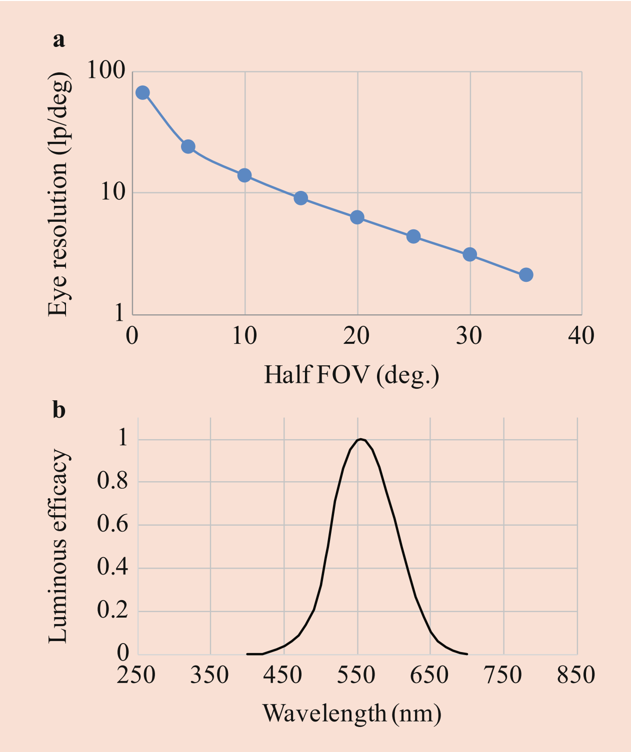


Fig. 8.6 (**a**) Resolution of the human eye plotted versus its half FOV. The highest resolution occurs at the center of the FOV and degrades rapidly as the FOV increases. (Data from [20]). (**b**) The photopic response curve of the eye. The green portion of the visible spectrum sees the greatest spectral response

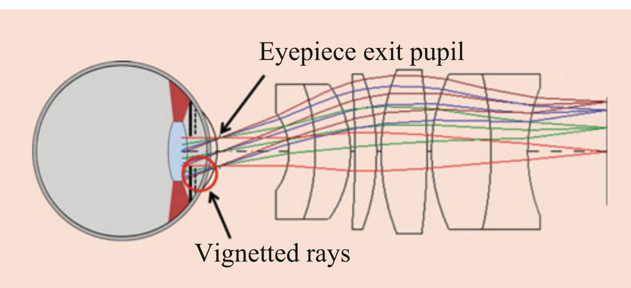


Fig. 8.7 An improperly pupil-matched eyepiece (both in location and size) will result in portions of the FOV being vignetted, as indicated by the red circle

for easy, adaptable use and to accommodate for natural eye rotations for a given head position. It is recommended that the minimum eyebox size for an AR display be 5 mm in diameter, which allows for 1 mm on each side of a 3 mm eye pupil for positioning. It is advised that AR displays with an eyebox below this threshold value utilize eye-tracking to ensure visibility of the images through natural eye movements. The minimum diameter is quite limiting, and it is recommended to aim for a much larger eyebox for an AR display, somewhere in the range of 8–15 mm. In this size range, there is more

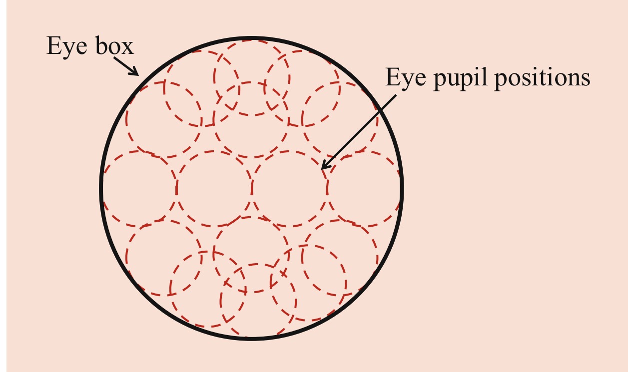


Fig. 8.8 Cross-section of a visual instrument eyebox. The black circle represents the full extent of the eyebox, and the dotted red circles indicate possible positions of the eye pupil when using the visual instrument. The eye receives the instrument's entire FOV when the eye pupil is wholly located within the eyebox

tolerance when positioning the AR display on one's face, and it captures a more extensive range of user interpupillary distances, the majority of which fall between 55–73 mm, in a single version of any particular AR display. Figure 8.8 shows a cross-section of an eyebox with various potential locations of the user's pupil.

Eye clearance is a parameter of an AR display related to the eyebox that affects the user's comfort. Eye clearance is defined as the minimum distance between the eye's entrance pupil and any mechanical feature of the AR display assembly. It is similar to the more common specification of eye relief but considers mechanical features in addition to optical components. If the eye clearance is too short, portions of the AR display will collide with the user's face or nose. If the eye clearance is too large, the AR display will protrude unsightly from the user's face. The separation between the field footprints on the optical surfaces will then increase, resulting in larger and faster optics. Thus, the eye clearance should be minimized for the application to facilitate the remainder of the system design. Eyeglasses, which reside about 17–20 mm from the face, can be used to reference the optimal eye clearance for AR displays. The minimum recommended eye clearance is about 10 mm – just enough room for the user to blink without touching the display.

Once the eye is positioned within the eyebox, the full virtual image can be seen, and the **FOV** determines the size of the virtual image. The FOV is one of the most critical design parameters of an AR display and typically leaves the strongest impression on a user. The larger the FOV, the more immersive the experience. The required FOV of an AR display varies widely based on the application. For text display and notifications, a small FOV of 15° may be sufficient. For immersive gaming, a 90° or larger FOV is sought. In classical eyepiece designs for inspection applications such as telescopes or microscopes, the image quality over the FOV

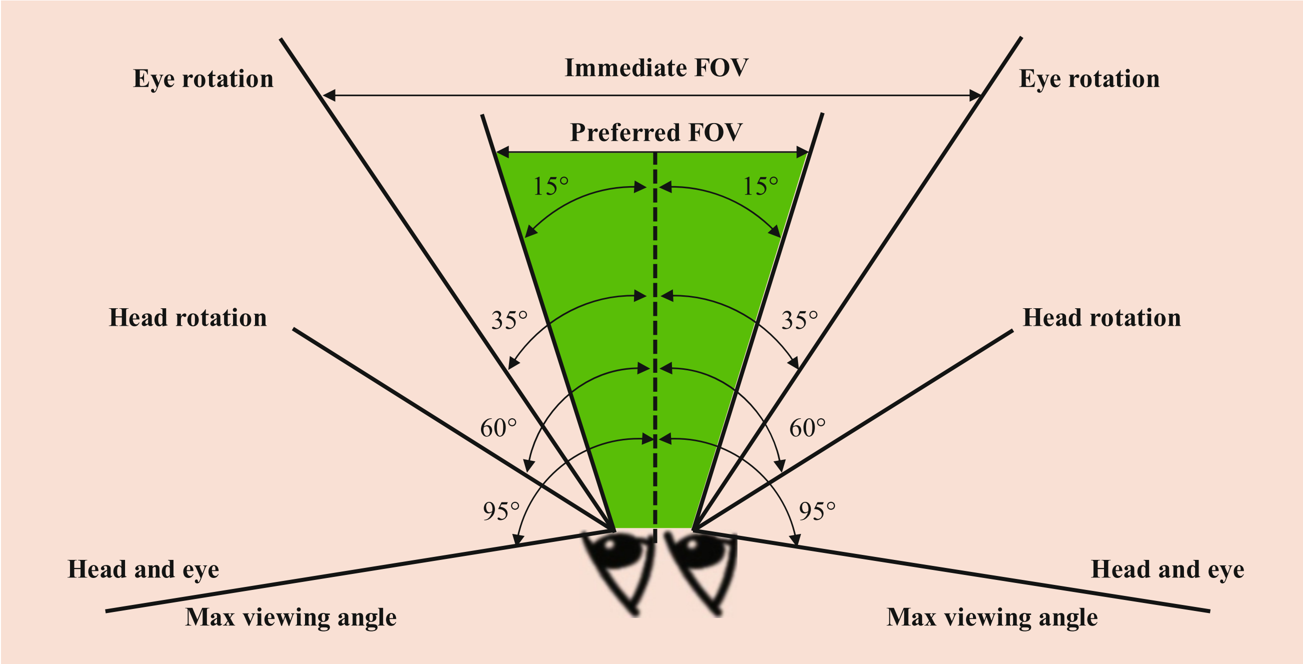


Fig. 8.9 Human visual system FOV with eye and head movements. (Adapted from [22])

is not designed to be uniform; often, the center of the FOV has better aberration correction than the periphery, and the object under observation is moved into the center of the FOV for closer inspection. Though AR display optics are similar to those of an eyepiece, this is an area where they differ. Instead of moving the inspection point to the center of the FOV, the eye rotates to do the inspection. Therefore, the image must maintain high image quality over the portion of the FOV that corresponds to the natural rotation angle of the eye. "Natural," in this case, means that there is a threshold in the FOV at around 15° (half-field) where, instead of the eye swiveling, the user's head will rotate to view objects beyond that point, as illustrated in Figure 8.9 [22]. With this threshold in mind, the FOV should be well-corrected to a 15° half-field. Beyond that point in the FOV, lower image quality may be leveraged and can correspond to the eye's acuity in the periphery.

As the FOV is increased, so does the **image distortion**. In rotationally symmetric visual systems, the distortion is purely radial in either pincushion or barrel forms, as illustrated in Fig. 8.10, and is perceivable when the percent distortion is >1%. It is preferable for the optics to mitigate image distortion. Still, in situations where this is not possible, distortion correction software algorithms can be implemented to pre-distort the source image with the inverse optical distortion, providing an undistorted view to the user [23–25]. AR displays often do not have rotational symmetry, so the distortion is rarely solely radial. Keystone and smile distortion are significant contributors to the overall distortion pattern seen in AR displays and cannot be corrected using

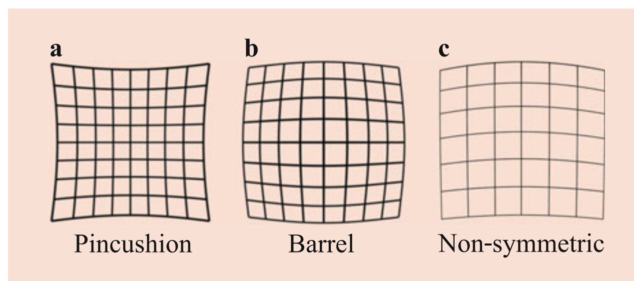


Fig. 8.10 Radial distortion in rotationally symmetric optics consists of either (**a**) pincushion or (**b**) barrel distortion. (**c**) More complex distortions are evident in optical systems where there is asymmetry

the same algorithms as radial distortion. In this case, more flexible algorithms are implemented [26]. A complexity in correcting distortion is the variation of distortion with the location of the eye in the eyebox, called pupil swim, which can be addressed in some cases in the graphics pipeline [27]. In full-color displays, the magnification variation with wavelength, termed lateral color, results in virtual image sizes that vary with wavelength. Like distortion correction, the residual lateral color of the display can be corrected via software, allowing some additional freedom in the optical design.

Finally, a parameter related to the FOV that is vitally important to AR displays with wide FOVs is **visual resolution**. This concept refers to the angle that a single pixel in the virtual image subtends at the eye, assuming pixel-limited operation. When viewing the virtual image, it is preferable to perceive a continuous image instead of a pixelated image. If the number of pixels that make up the image is low or conversely the FOV of the AR display is wide, the visual resolution can be low, leading to a noticeable "screen-door" effect of the virtual image where the eye resolves the individual pixels. For an unresolvable pixel, the visual resolution should be at or below the eye's maximum resolution, or one arcminute. This issue is common with VR headsets, where the 100°+ FOV is subtended by only a few thousand pixels, yielding a visual resolution of around three arcmin/pixel (for a 100° FOV and 2000-pixel display), corresponding to about 20/60 vision (20/200 or 10 arcmin resolution is considered legally blind). Solutions include increasing the number of pixels that subtend the image or distributing the pixels non-uniformly to place them preferentially where the eye is looking, which requires advances in the digital display industry.

8.4 Optical Components of an AR Display

From an optics standpoint, an AR display can be broken into essentially three basic components: the light engine, the imaging optics, and the combiner. The light engine is responsible for generating and spatially modulating the light that forms the virtual image. Once the light has been generated, it is sent through the imaging optics to yield collimated (or nearly collimated) light beams that a relaxed eye can focus into virtual images. The last surface before the light reaches the eye is the combiner. Its primary function is to allow light from the environment to propagate to the eye while providing a surface to direct the light from the imaging optics into the eye. If the combiner has optical power, it also plays a role in the imaging optics. A schematic of the three components shown on a generic AR display platform is shown in Fig. 8.11. This section will break down each of those three components.

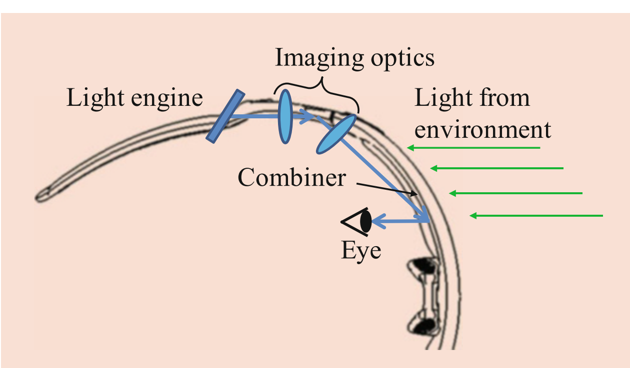


Fig. 8.11 Schematic of a generic AR display and its three basic components

8.4.1 Microdisplays as the Light Engine

As discussed, one optical component of an AR display is the **light engine**. This component can take essentially two forms: (1) a miniature digital display (termed microdisplay) or (2) a single source that is rapidly scanned over the eye's FOV. First, microdisplays and the vital role they play in the first-order optical calculations of an AR display will be discussed. The FOV and the focal length of the imaging optics are related to the physical size of the microdisplay by the familiar equation,

$$h = f \tan\theta, \tag{8.1}$$

where h is the semi-diagonal of the microdisplay, f is the focal length of the imaging optics, and θ is the half diagonal FOV. For a given FOV, the physical size of the optical system scales with the focal length; therefore, a compact optical system prefers a small focal length. According to Eq. (8.1), a smaller focal length yields a smaller microdisplay, which is advantageous for packaging. However, given a constant eye-box size, the desire for a compact optical system is countered by the ability to correct the increased imaging aberrations for the faster f-number that results from a smaller focal length. Operating at a faster f-number can require additional optical elements for aberration correction, so the packaging benefit of a small focal length is diminished. Another main tradeoff with microdisplay size is resolution. A smaller microdisplay will tend to limit the angular resolution at the system level as, for a given technology, the smallest pixel pitch achievable will limit how many pixels may be packed into a given microdisplay size.

There are essentially two classes of **microdisplays**, differentiated based on the source of the display light – self-emissive and externally illuminated. In self-emissive microdisplays, the pixels operate without the use of an external light source. The main advantages of a self-emitting microdisplay are the simplicity of its operation and its compact packaging. With space at a premium for an AR display, these advantages are critical. However, self-emissive displays are generally dimmer than externally illuminated displays. The most common self-emissive microdisplay uses organic light-emitting diodes (OLED). Both full-color and monochromatic OLED microdisplays are commercially available. The monochrome versions are typically green and can achieve higher brightness than the full-color options. The organic material can be optimized for monochromatic light, and the pixels do not require color filtering [28]. There have been early demonstrations of full-color OLED microdisplays that eliminate color filtering, resulting in a significantly brighter image [29].

A new technology destined to compete with **OLEDs** for self-emissive microdisplay supremacy is micro light-emitting diodes (LED). Operating with the same principles as conventional LEDs, micro-LEDs share the benefits of LEDs but in a much smaller footprint. Micro-LED microdisplays show promise to provide a bright solution with excellent color representation that draws less power and lasts longer [30]. At the time of this writing, micro-LEDs are in the early stages of development. While there are a handful of companies in the field, there is not yet a commercially available micro-LED microdisplay [31, 32]. Similarly, superluminescent LED microdisplays are an emerging technology that could impact the AR/VR market [33].

Externally illuminated microdisplays consist of an illumination engine and a pixelated display. The illumination and display base technologies have been mature for quite some time and have been the most widely used microdisplay type in early AR displays. The illumination is typically provided by conventional LEDs, which provide high brightness, low power consumption, and a long lifetime. The pixelated display is most commonly a liquid crystal on a silicon (**LCoS**) display. A typical illumination schematic for an externally illuminated microdisplay is shown in Fig. 8.12, though more compact illumination methods can be leveraged [34, 35]. The pixel size of LCoS displays is generally smaller than OLEDs, which allows for a smaller display board for a given resolution.

Full color can be achieved using LCoS microdisplays in two ways. The first method, called color filtering, uses a white LED to illuminate a single LCoS display combined with individual color filters placed over the pixels. The advantage of this approach is that only a single illumination source is required. The disadvantages include large pixels consisting of sub-pixels, loss of efficiency from the filtering, and color bleeding. The second method is called color sequential, in which three separate illumination sources are utilized for red, green, and blue. The primary colors sequentially illuminate the LCoS display and are combined into a single path using a combiner optic such as an X-prism. This approach allows for smaller individual pixels, smaller displays, and better efficiency (no filtering).

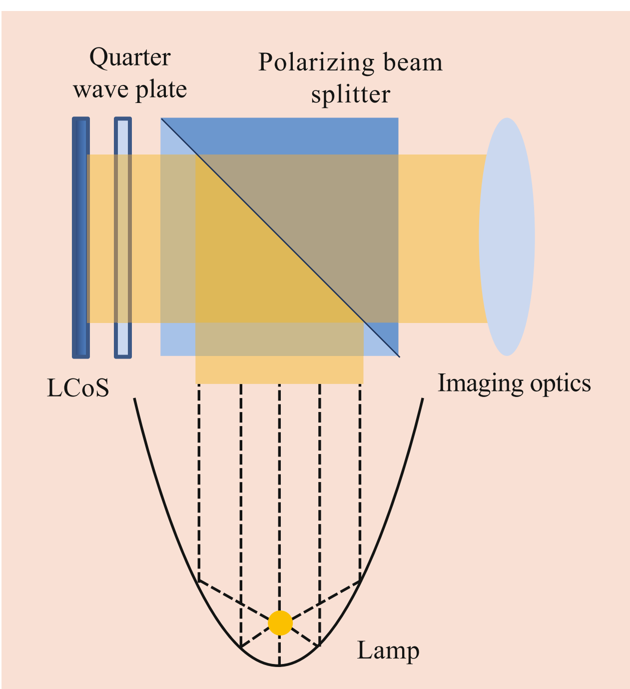


Fig. 8.12 A common illumination schematic for an LCoS microdisplay. Red, green, and blue LEDs can be used in tandem to create full-color illumination

8.4.2 Radiometric Brightness Analysis for AR Displays to Guide Microdisplay Specifications

Directly related to the brightness of the light source is the overall brightness of the virtual image produced by the AR display. In a completely dark room, an AR display image need not be extremely bright for the user to see the image distinctly. However, when used in full sunlight, an insufficiently bright virtual image will look washed out relative to the sun-soaked background of the environment. The question then becomes, how bright does the AR display image need to be? Here, a fundamental radiometric analysis of the AR display and environment brightness will be performed to set a lower bound for the required brightness of the AR display image and the microdisplay.

The relevant metric for this discussion is the ratio between the brightness or, more appropriately, the luminance of the AR display, denoted as $L_{\nu,\mathrm{display}}$, and the luminance of the environment, $L_{\nu,\mathrm{environ}}$. This quantity is called the **contrast ratio** (CR), and it can be written as

$$CR = 1 + \frac{L_{\nu,\mathrm{display}}}{L_{\nu,\mathrm{environ}}} \tag{8.2}$$

The minimum acceptable contrast ratio for AR is 1.2 [15].

The luminance of the environment is driven exclusively by sunlight – irradiating the Earth with approximately 1000 W/m^2 of visible and infrared radiation [36]. However, only visible light should be considered for the CR, so it is necessary to determine what percentage of the 1000 W/m^2 arrives in the form of visible light. The sun is considered to be a blackbody that obeys Planck's radiation equation,

$$L_\lambda = \frac{2hc^2}{\lambda^5}\frac{1}{\left(e^{hc/\lambda kT} - 1\right)}, \tag{8.3}$$

where L_λ is the spectral radiance, h is Planck's constant, c is the speed of light, λ is the wavelength of the light, k is Boltzmann's constant, and T is the temperature of the blackbody. The average temperature of the sun is 6000 K, generating the spectral radiance curve as a function of wavelength, plotted from 200 nm to 1300 nm in Fig. 8.13. By integrating Eq. (8.3) from $\lambda = 400$ nm to $\lambda = 700$ nm, the percentage of light emitted in the visible spectrum can be found to be approximately 42%. Thus, of the 1000 W/m^2 total irradiance delivered to the Earth, 420 W/m^2 is in the visible spectrum. Next, of that 420 W/m^2, only a fraction is reflected or emitted by the Earth. Studies in photometry have shown that a typical outdoor scene has a reflectivity, ρ, of around 0.18 [36]. The exitance (W/m^2) of the illuminated environment, M_{environ}, can be related to the irradiance of the sun, E_{sun} (W/m^2), and the radiance (W/sr·m^2) of the environment, L_{environ}, by

$$L_{\text{environ}} = \frac{M_{\text{environ}}}{\pi} = \frac{\rho E_{\text{sun}}}{\pi}, \tag{8.4}$$

assuming Lambertian scattering from the environment. To compare Eq. (8.4) with AR display specifications for luminance, L_{environ} must be converted to photometric units (lm/sr·m^2) using the relationship

$$L_{v,\text{environ}} = \int_{400\text{ nm}}^{700\text{ nm}} L_{\text{environ}} K_\lambda \, d\lambda, \tag{8.5}$$

where K_λ is the photopic spectral luminous efficacy. Combining Eqs. (8.4) and (8.5), the luminance of the environment illuminated by the sun can be calculated to be approximately $L_{v,\text{environ}} = 5660$ lm/sr m^2. Using Eq. 8.2 and a contrast ratio of 1.2, a lower bound for the luminance required of an AR display is $L_{v,\text{display}} = 1130$ lm/sr m^2 or 1130 nits. Assuming an 85% total transmission of the optics within the AR display (equal to about five anti-reflection coated lenses), 1130 nits translates to a minimum acceptable luminance for a microdisplay of 1330 nits for outdoor use. Most commercially available full-color OLEDs cannot reach this luminance level and thus come short of producing an acceptably bright image for daytime use. However, recent developments of color filter-free OLED microdisplays show promise [29].

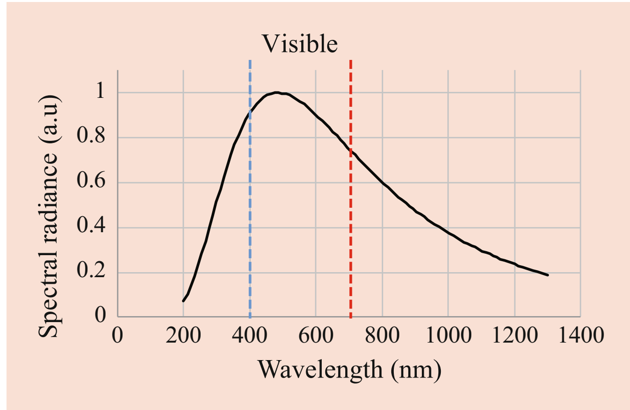


Fig. 8.13 Blackbody radiation curve for the sun at a temperature of 6000 K

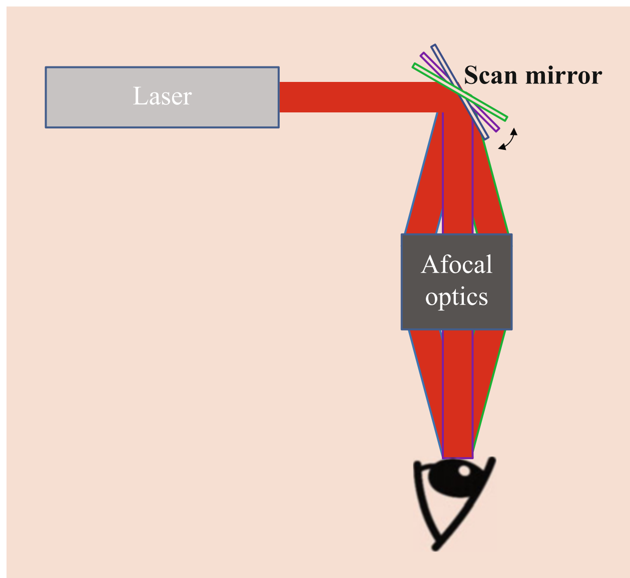


Fig. 8.14 One-dimensional monochromatic laser scanning system architecture. By scanning the mirror rapidly, the eye sees all field points simultaneously. This concept can be extended into two dimensions using an additional scan mirror or implementing two axes of rotation on a single mirror

8.4.3 A Brief Foray into Laser Scanning

An alternative to using a microdisplay is to employ a **laser beam scanning system**. As an exemplary type of scanning system, consider a scanned laser system. In this type of system, a stationary collimated laser is incident on a two-dimensional (X/Y) scanning system, typically consisting of a rotating mirror(s). The rotation of the mirrors is such that the laser can be redirected into a range of angles that is directly related to the FOV. When combined with focusing optics, the scanning creates a series of "pixels" that comprise a complete two-dimensional image. For an AR display, the focusing optics are the eye's optics. The two-dimensional scans occur rapidly enough that, even though only a single location on the retina receives light at any given time, the eye perceives a complete two-dimensional image. This concept is illustrated for one-dimensional scanning in Fig. 8.14. A broadband laser in an adapted architecture or multiple single-wavelength lasers combined may be used to create a full-color image.

The advantages of using a scanned laser system are multiple. First, lasers are capable of high luminance. Second, lasers are power efficient. Third, the polarization of laser

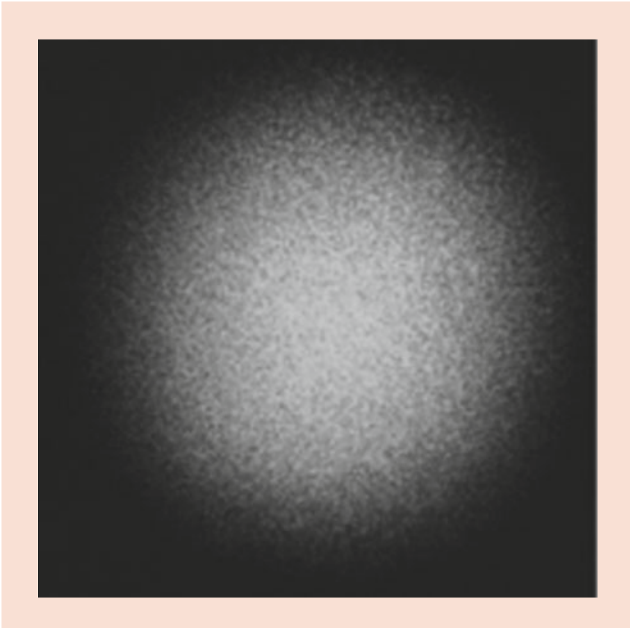


Fig. 8.15 Laser speckle caused by interference on a diffuse surface

light can be beneficially leveraged elsewhere in the optical system. Last, the narrow spectrum of each color provides unique opportunities for manipulation using wavelength-specific mechanisms, such as holograms. As discussed in a later section, some AR display architectures are based on diffraction, for which narrow spectra are advantageous. The drawbacks, too, are multiple. A significant issue is that the narrow diameter of the laser beam waist (1–2 mm) imposes a limit on the size of the eyebox that can be created. A 1–2 mm eyebox is, unfortunately, not sufficient for an AR display. Techniques have been used to circumvent this obstacle, namely pupil replication and eye-tracking. **Pupil replication** multiplies the number of pupils the optics produce to create a sizeable synthetic eyebox [37–39]. The pupils require two-dimensional replication by preferably 8x, which reduces each pupil's intensity by a factor of 64 – luckily, lasers have luminance to spare. **Eye-tracking** allows for a smaller pupil to be used, but the AR display needs to be coupled with a precise eye-tracking system; the challenge goes on. Additionally, coherent laser light produces interference patterns called speckle, shown in Fig. 8.15, that can degrade the quality of the observed image. However, progress has been made towards speckle-free images using [40].

8.4.4 Imaging Optics and Combiners

In the case of a microdisplay-driven AR display, the role of the **imaging optics** is to collimate the diverging light generated by the microdisplay. Sounds simple, right? But add a wide FOV coupled with a large eyebox into the mix, all while keeping the optics compact, and one can begin to understand the challenges of designing the imaging optics. In optical design terms, the imaging optics fall into the category of an eyepiece, with the added wrinkle of conforming the optical layout to a head-worn device that often follows a nonlinear path. The imaging optics must be located either along the glasses' temples or above the brow outside the FOV of the user's eye to maintain an unobscured view of the real world. It is not a linear path to the user's eyes from either location, so the light path must be folded. There are a variety of architectures for constructing the imaging optics, and each option has advantages and disadvantages. These architectures will be expanded upon later in this section.

The **combiner** is the optical component that optically combines the virtual and real-world images. In a typical AR display, the combiner resides at the location of the lens in a conventional pair of eyeglasses. The mechanism by which the combiner functions varies depending on the optical architecture of the imaging system. Some architectures require the combiner to play a role in the collimation of the light before it enters the eye. In contrast, others use the combiner as a simple beamsplitter to redirect already collimated light into the eye. The role of the combiner in various architectures will be detailed below.

8.5 Optical Architectures and How They Work

The optical riddle of designing a wide FOV, large eyebox, and compact AR display that produces bright and crisp images while not occluding the user's view has not yet been solved, though there have been many efforts in pursuit of a solution [41]. The simplest optical architecture couples a conventional coaxial eyepiece with a **plate beamsplitter combiner**, as illustrated in Fig. 8.16. The coaxial eyepiece is responsible for collimating the light source, and it can produce the wide FOV and eyebox necessary for viewing, but the plate beamsplitter combiner leaves much to be desired aesthetically. This architecture is not conducive for integration into a pair of eyeglasses due to the linear path of the eyepiece and the size and tilt angle of the combiner needed to direct the light into the eye. However, given its conceptual simplicity, this architecture is ideal for a benchtop setup in a research setting for perception studies [42–44].

A similar architecture, coined "the birdbath" due to its resemblance to a birdbath, was employed in several commercial AR displays, most notably in Google Glass. The combiner, in this case, is a flat beamsplitter embedded within a piece of glass. The light emitted from the microdisplay propagates within the glass piece with lower divergence than free-space propagation, minimizing the size and curvature

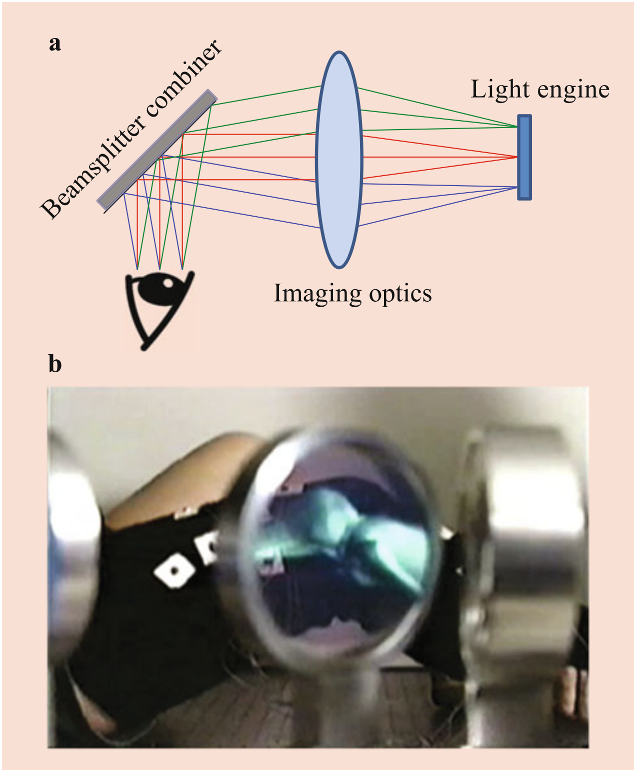


Fig. 8.16 The simplest type of AR display consists of an eyepiece to collimate the light from the light engine and a plate beamsplitter to serve as the combiner of the virtual image and the real world, as shown in the (**a**) ray trace. This architecture is best suited for (**b**) in-the-lab test setups

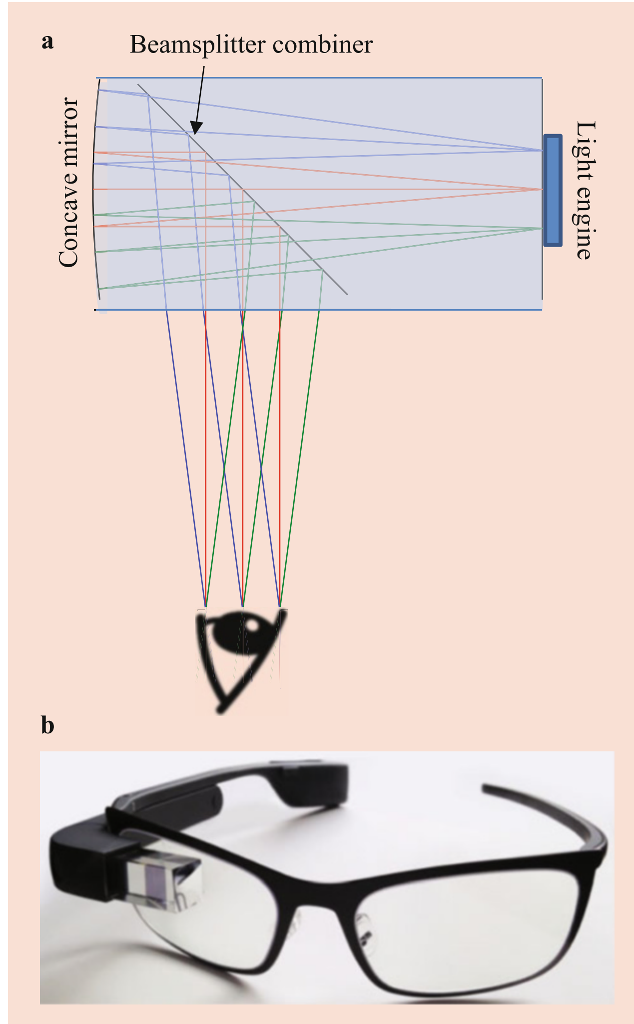


Fig. 8.17 (**a**) Ray trace of the birdbath architecture. A concave mirror collimates the light, and a flat beamsplitter serves as the combiner. This architecture was commercialized in numerous AR displays, most notably in the (**b**) Google Glass

of the downstream optics. The sole imaging optic is a concave mirror, polished into the end of the glass piece. It is most leveraged when its center of curvature coincides with the aperture stop (eyebox), enabling coma and astigmatism correction. Only a small FOV and eyebox are achieved in this architecture, as the single optical surface used for imaging is unable to correct for spherical aberration and field curvature. However, the optics can remain quite small with the small FOV and eyebox of this architecture. Embedding the combiner in glass hinders a seamless integration into a pair of eyeglasses; thus, this architecture has a unique and noticeable aesthetic. An exemplary birdbath architecture ray trace is shown in Fig. 8.17, together with one commercial implementation.

The next architecture group integrates an eyeglasses-like lens as its combiner – the **free-space reflective combiner**. When designing an AR display in this architecture, it becomes immediately apparent that the typical orientation angle of a standard eyeglasses lens cannot reflect the light into the eye if the light engine is located outside the eye's FOV, such as at the temple of the glasses or above the brow. Instead, it reflects the light into the user's nose or cheeks, respectively, as illustrated in Fig. 8.18 for the case of a temples-mounted light engine. One approach to overcome this limitation is to reposition the combiner. The reflected light from the side or top-mounted light engine can be appropriately directed into the eye by modifying the combiner tilt. This path was taken by McGuire in the design of the free-space reflective combiner AR display shown in Fig. 8.19 [45]. Unlike in the previous architectures, there is ample space for the imaging optics to collimate the light, enabling a moderate FOV and eyebox. Many of the lenses in McGuire's design are used off-axis or in a tilted configuration to compensate for the system asymmetry introduced by using a tilted and optically powered combiner surface. The FOV and eyebox in this architecture are limited not by aberrations but by the physical size of the necessary imaging optics. This design, and those like it, comes closer to emulating a pair of eyeglasses, but with one crucial difference. The modified angle of the combiner creates an outward bulge on the temple side of the lens

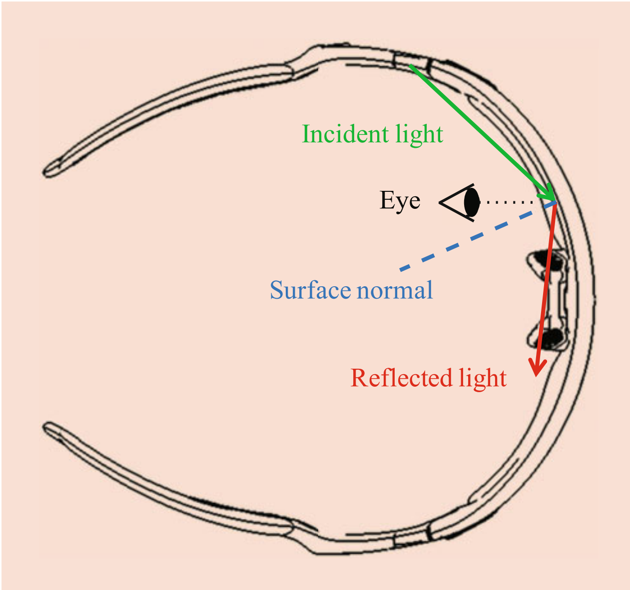


Fig. 8.18 A typical eyeglasses lens orientation in the free-space reflective combiner architecture reflects the light into the nose rather than into the eye with a side-mounted light engine

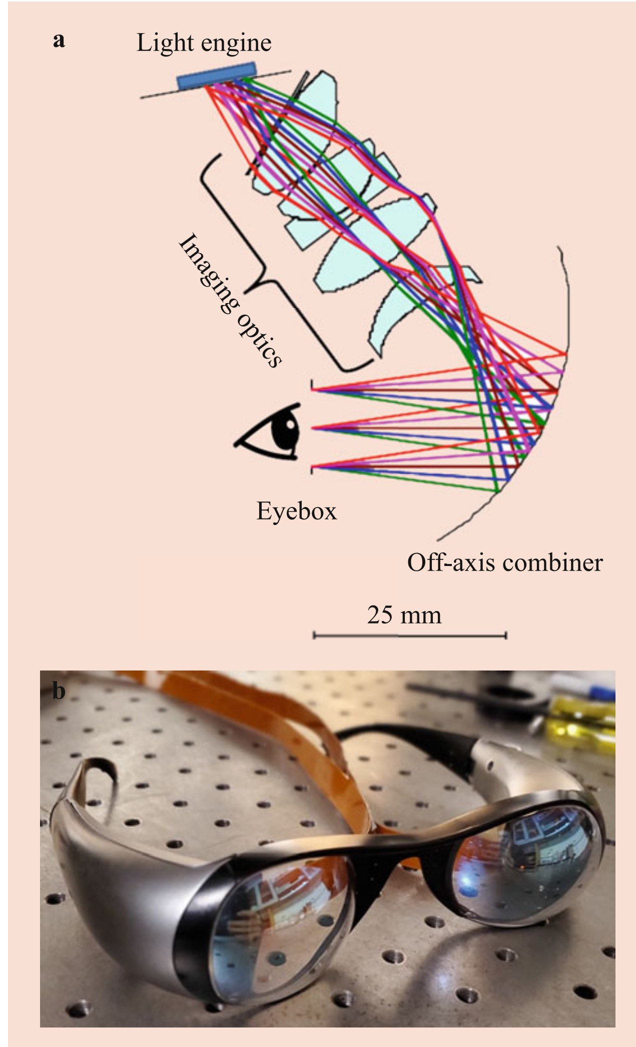


Fig. 8.19 (**a**) Ray trace and (**b**) prototype for the free-space reflective combiner AR display documented by McGuire [45]. It has a FOV of 20° and a 10 mm eyebox diameter. (Ray trace adapted from [45])

and an inward pull on the nose side of the lens, giving the impression bug-eyes on the wearer, hence the colloquial term for this type of combiner – "bug-eye" reflectors. Another design in the free-space reflective combiner architecture also suffering from the "bug-eye" is shown in Fig. 8.20 [46]. In these examples, a beamsplitter coating is applied to the combiner to allow simultaneous reflection of the virtual image and transmission of the real world. Both designs have uncorrected distortion that requires digital prewarping of the source image [26]. Additional designs using the free-space reflective combiner can be found in [47–53].

While the bug-eye form alters the combiner's angle, another method to direct the light into the eye using the free-space reflective combiner architecture is to reduce the angle at which the light is incident on the combiner. For the case of the light engine being located at the temple of the glasses, there is little space or freedom to decrease the incidence angle on the combiner by moving the light engine, so an alternative layout is needed. Bauer and Rolland implemented a mirror that lays along the nose as an intermediary between the combiner and the imaging optics at the temples [54]. The nose mirror is flushed against the nose, becoming similarly unobtrusive to the user, even though it is within the eye's FOV. Light from a microdisplay located at the temples is directed towards the nose mirror, which subsequently reflects the light to the combiner. The bug-eye form is avoided by directing the light to the combiner from the nose side at a lower incidence angle. In their design, Bauer and Rolland implement **freeform surfaces**, or surfaces whose shapes have no rotational or translational symmetry, to combat the aberrations generated by using the tilted nose mirror and combiner. This arrangement eliminates the bug-eye look and uses a minimum number of optical elements. The combiner was coated to be 70/30 in reflection and transmission, respectively. This layout's disadvantage is the proximity of the nose mirror to the wearer's face, limiting its size and, by extension, the FOV and eyebox. This arrangement is optimal with the virtual display in portrait mode (the long dimension of the FOV is vertical) to maximize the FOV and eyebox given the nose-mirror size constraint. The layout of that system is shown in Fig. 8.21.

Similar to using a nose mirror to lower the incidence angle at the combiner, another alternative is to position the microdisplay at the brow above the eye and aim it down towards the combiner. In this case, both the horizontal and vertical tilt of the combiner can be used to direct the light into

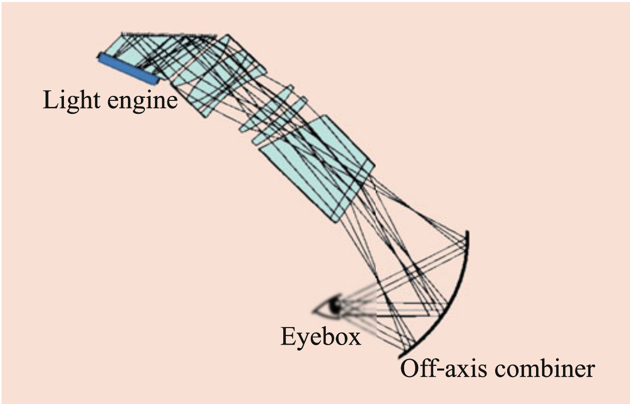


Fig. 8.20 A 60° FOV AR display of the free-space reflective combiner type with a 10 mm eyebox diameter. It uses a toroidal combiner working off-axis and various tilted and decentered elements for aberration correction. (Adapted from [46])

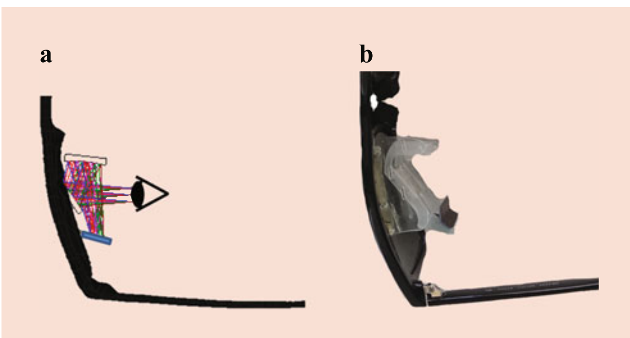


Fig. 8.21 (**a**) A ray trace of a prototype AR display in the free-space reflective combiner architecture and (**b**) its prototype implementation. This design uses a mirror near the nose and freeform mirrors to overcome the challenge of bug-eye AR display [54]

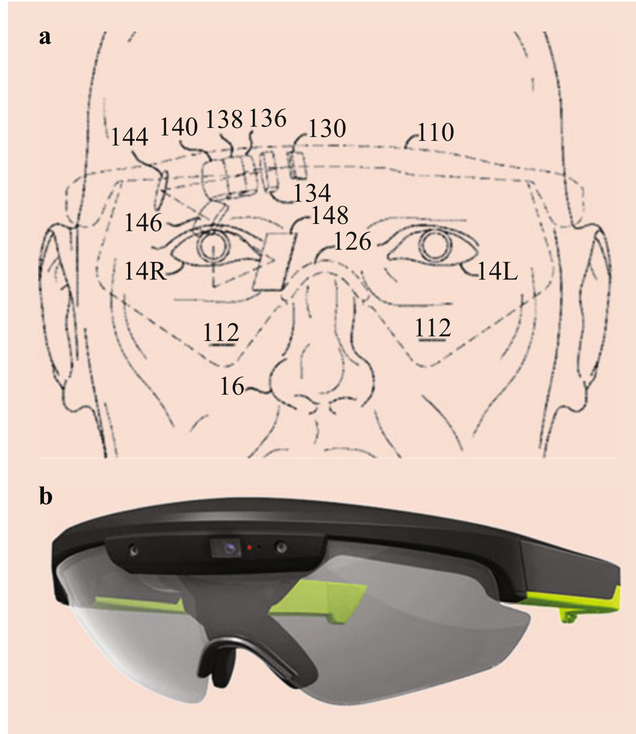


Fig. 8.22 (**a**) Optical layout of the Everysight Raptor with the light engine and optics located above the brow [55]. (**b**) The Everysight Raptor commercial AR display

the eye. The Everysight Raptor, shown in Fig. 8.22, utilizes a brow-mounted display and optics. A mirror near the nose is used to avoid the bug-eye form, and a slight upward tilt of the combiner in the vertical direction directs the light to the eye.

A third method to overcome the bug-eye challenge in the free-space reflective combiner architecture is to leave the light engine and combiner incidence angle unaltered and, instead, change the combiner's optical properties to allow for anomalous reflection. For example, the combiner could be outfitted with a diffraction grating or hologram that obeys the diffraction equation rather than the law of reflection [56–58]. The advantage of this method is clear – it provides the freedom to position the combiner at an angle that is optimal for the aesthetics and mechanics of the eyeglasses while maintaining its optical function. However, this method also has significant disadvantages and challenges that have yet to be overcome. Diffractive structures, such as **holograms**, are highly wavelength selective leading to variable behavior over a 400–700 nm range of wavelengths for a typical RGB light engine. The spectral bandwidth of the light engine must be narrow to get the desired optical response. Multiplexed holograms can be used to widen the operational spectrum, but they can result in unintended ghost images. Laser light can obtain the desired results due to its monochromaticity, but it suffers from the eyebox limitations discussed in an earlier section.

Despite these challenges, this technology has been recently commercialized in AR displays. North developed the Focals after acquiring the patents and technology from Intel, whose product, Vaunt, was discontinued. Images of the Focals by North are shown in Fig. 8.23 with a ray trace of an exemplary laser scanning system with a holographic combiner to illustrate the anomalous reflection. Intel first acquired the intellectual property for Vaunt from the Swiss startup, Composyt Light Labs, that pioneered a method to overcome the limited eyebox associated with a laser scanning AR display [59]. Because the holograms are highly wavelength selective, multiple close variants of a color (or wavelengths) can be used without the user noticing the difference. Three red laser diodes were used in a scanning configuration with three multiplexed holograms on the combiner that provided slightly offset eyeboxes at the eye location. Each eyebox was small (∼1 mm), but this strategy increased the area of the eyebox by a factor of three.

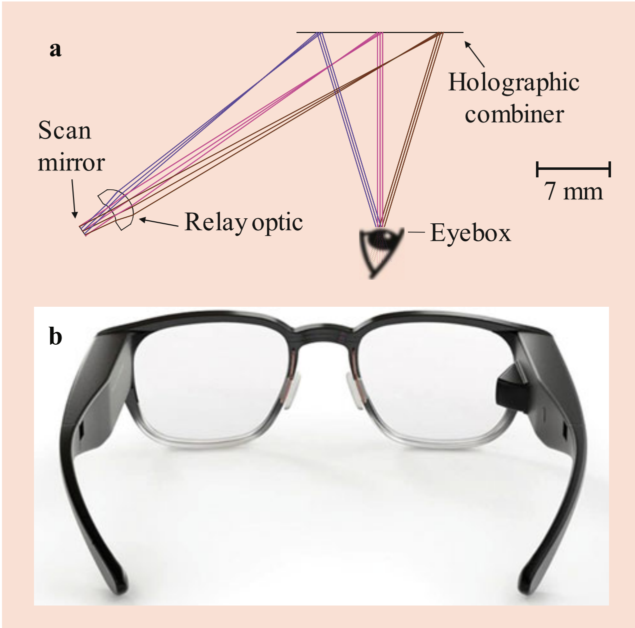


Fig. 8.23 (**a**) A ray trace for a generic pair of laser-scanned AR glasses. (**b**) The Focals by North look comparable to a standard pair of eyeglasses

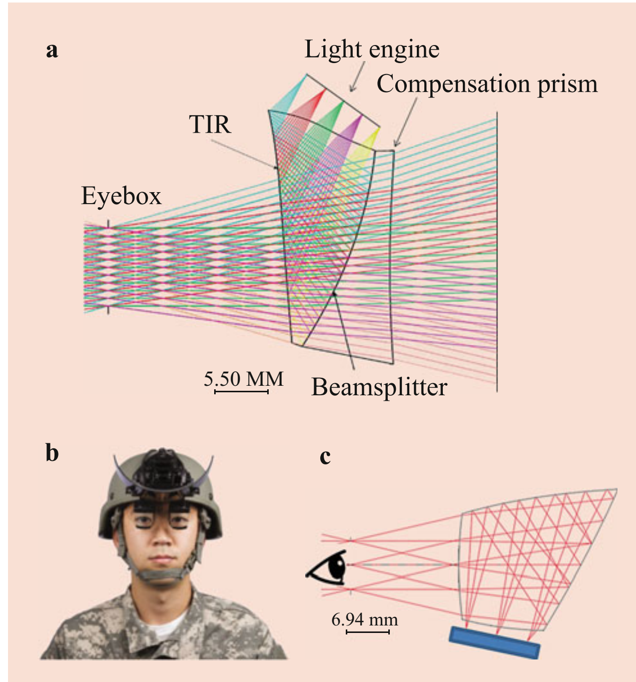


Fig. 8.24 (**a**) A ray trace of the TIR prism architecture. The compensation prism is necessary to have crisp imaging of the real world [62]. This design has a 53.5° FOV and an 8 mm eyebox diameter. (**b**) The TIR prism has been commercialized and is often used in military applications, such as the SA Photonics product shown here. (**c**) An alternative folding geometry has better intrinsic aberration behavior. (Adapted from [63])

Freeform optics were introduced in the discussion above as a tool to correct rotationally variant aberrations in the free-space reflective combiner architecture. Here, the **total internal reflection (TIR) prism** architecture exists solely because of **freeform optics** and was first introduced by Canon in the patent literature [60, 61]. Like the birdbath architecture, the TIR prism uses a piece of glass together with its boundary surfaces to collimate and direct the light into the eye. Referencing Fig. 8.24, light from the light engine enters the prism and is soon incident on one of the prism-air interfaces at an angle steep enough to trigger TIR, enabling a perfect reflection from an uncoated substrate. After the TIR surface, the light traverses the prism and interacts with the opposite side, which is beamsplitter coated and serves as the combiner. After that reflection, the light traverses the prism one more time and interacts once again with the prism-air interface nearest the eye. Because the TIR condition allowed this surface to be uncoated, the light now refracts if the angle of incidence is below the critical angle ($< 41.8°$ for $n = 1.5$ substrate). The result is a monolithic imaging system that provides a wide FOV in the order of 50° and a reasonable eyebox size in the order of 10 mm. The prism surfaces are necessarily freeform shaped due to the need for rotationally variant aberration correction resulting from the tilted prism surfaces. The TIR prism system only becomes fully see-through when a separate prism is added behind the combiner surface to compensate for the freeform shapes of the imaging prism that aberrate the real-world image when viewed alone.

The advantages of the TIR prism are that it allows a wide FOV and eyebox size, and it is a monolithic unit (excluding the compensation prism). Compared to the architectures shown thus far, the TIR prism supports the largest FOV and eyebox relative to the optics size and element count. It is, however, unable to seamlessly integrate into a pair of eyeglasses due to its unique form factor. The best compromise is to put the prism behind a non-functioning semi-transparent lens, like a visor. The TIR prism also has challenges supporting full-color operation because the prism is the sole dispersive material used in the design. Paths to manage the chromatic aberrations include using additional optics between the light engine and the prism, adding a diffractive optical element to a prism surface for its unique dispersion [64], or gradient-index optics [65]. There has been some commercialization of the TIR prism, and it has been primarily marketed towards military use, as shown in Fig. 8.24, where aesthetics are a lower priority.

The correction of the rotationally variant aberrations generated using tilted powered surfaces in the TIR prism has the potential to be improved. The role of the freeform optics is to correct these aberrations, but as Fuerschbach et al. [66] and Bauer et al. [67] show, not all combinations of rotationally variant aberrations are readily corrected by freeform optics.

The TIR prism represents a case where, if a change were made to the prism's reflection direction, freeform optics could be more effective. Takaki et al. [63] explored a variation on the TIR prism also studied by Chen and Herkommer [68] whereby the TIR surface is removed. A ray trace of the layout is shown in Fig. 8.24b. The advantage of this later prism geometry is that it achieves better aberration correction (e.g., capable of more FOV and eyebox) with significantly less freeform departure, making the prism easier to manufacture. The disadvantage of this alternative prism geometry is that it may interfere with one's peripheral vision and can require a thicker substrate than the TIR prism.

The next architecture is reminiscent of both the birdbath and the TIR prism. It combines the TIR surfaces of the prism and the embedded combiner of the birdbath architecture. The concept is illustrated in Fig. 8.25. A microdisplay emits light that is sent through imaging optics before entering a prism. A tilted, reflective prism surface redirects the light within the prism at an angle steep enough to TIR three times as it propagates along the prism. Finally, a beamsplitter-coated combiner breaks the TIR condition, and the light exits the prism before entering the user's eye. Reimaging (with an intermediate image) is utilized in this architecture to support the large optical path that the light must travel. This architecture's advantage is that much of the aberration correction can be accomplished using the imaging optics before entering the prism, allowing the prism to be primarily used to transport the light to the eye. The main disadvantage of this architecture is that the prism can be bulky to accommodate a moderate FOV and eyebox. This architecture has been commercialized by the Epson Moverio AR display, also shown in Fig. 8.25.

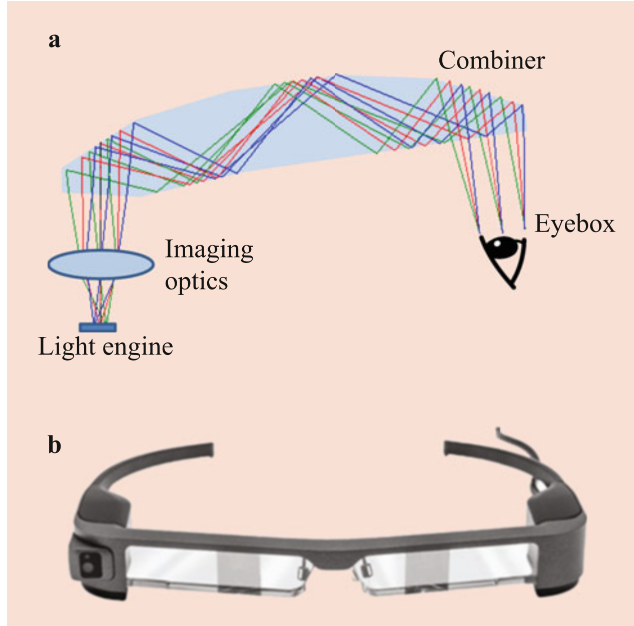


Fig. 8.25 (**a**) Schematic of the architecture commercialized by Epson in the (**b**) Moverio BT-300 AR display. Its FOV is 23°. Note that the ray trace shown is for illustration purposes only and is not to scale

TIR is taken to the next level in the **waveguide** architecture, where the imaging optics collimate light that is subsequently coupled into a substrate slab at an angle greater than the critical angle. The light propagates down the length of the substrate by TIR until the TIR condition is broken. When the light reaches the lateral point at which the eyebox is located, an out-coupler breaks the TIR condition, ejecting the light towards the eye. Most versions of this architecture utilize a planar waveguide slab, which will dominate the discussion here. However, the holy grail for the aesthetics of this architecture is to move to a curved waveguide [69, 70].

The method in which the light is in-coupled and out-coupled from the waveguide differentiates the implementations of this architecture. The most basic in-coupler and out-coupler is an angled mirror [71, 72], as shown for the out-coupler in Fig. 8.26. An angled reflection of the incoming light directs the light into the waveguide at an angle greater than the critical angle for TIR. A similar surface, or surfaces, at the opposite end of the waveguide breaks the TIR condition and reflects the light out of the waveguide. Another common method is to use diffractive structures on the surface of the waveguide to diffract the light into angles greater than the critical angle, as shown for the in-coupler in Fig. 8.26. The diffractive structure could take a variety of forms, whether it be a simple ruled grating [73], a surface relief hologram [74], a volume hologram [75], or a metasurface, as further discussed later in this chapter [76]. Regardless of the specific diffractive coupling method implemented, the goal is to be as efficient as possible without introducing harmful diffraction effects. Optical see-through is accomplished by simply looking through the transparent waveguide and diffractive out-coupler. In the case of a diffractive out-coupler, it can be engineered to give the desired response for specific wavelengths of light within the waveguide while not significantly impacting the light from the real world. However, this aspect must be carefully studied. Full-color imaging is another area where the different versions of the architecture can differentiate themselves. The angled mirror-coupling method is achromatic, but the diffractive techniques are wavelength-dependent and require wavelength-specific engineering. Separate diffractive structures for each primary color have been implemented, with, for example, separate waveguides for green versus the combined red and blue to mitigate crosstalk between the colors, as illustrated in Fig. 8.27 [77].

The out-coupling in waveguide designs has evolved to consist of cascaded implementations of their respective out-coupling features for the purpose of eyebox expansion. For example, in the case of the angled reflectors, the eyebox can be expanded in the horizontal direction by implementing a series of increasingly reflective beamsplitters, as shown in Fig. 8.26. At the first beamsplitter, a percentage of the

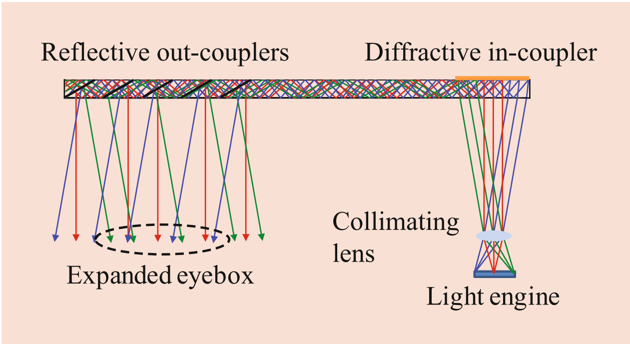


Fig. 8.26 A planar waveguide using a diffractive in-coupler and angled reflectors for the out-coupler. The out-coupler consists of cascaded beamsplitters to expand the eyebox. The Lumus glasses shown in Fig. 8.4 use a similar concept for the out-coupler

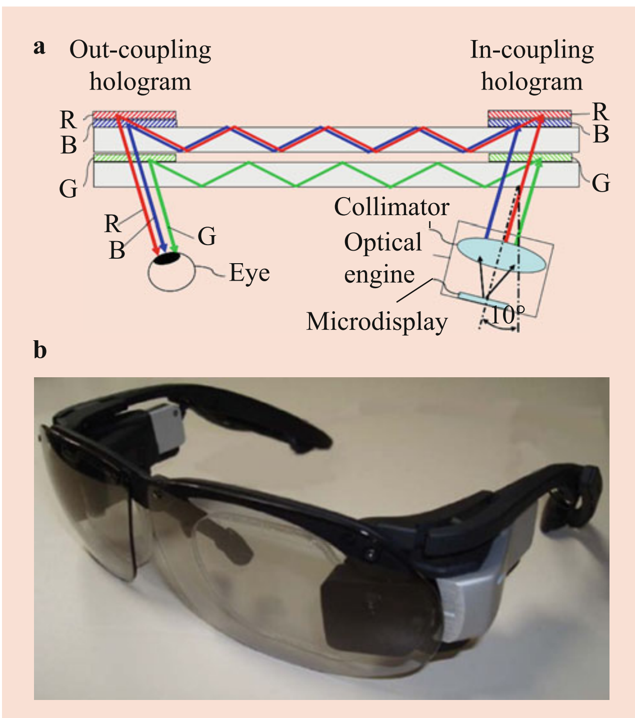


Fig. 8.27 (**a**) The ray trace of a Sony waveguide AR display uses multiple waveguide substrates to reduce crosstalk from the coupling features and its (**b**) prototype. (Images from [77])

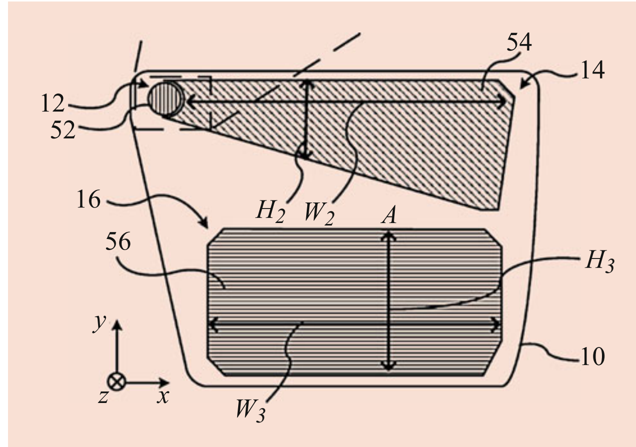


Fig. 8.28 The combiner construction for two-dimensional eyebox expansion was proposed by Microsoft [78]

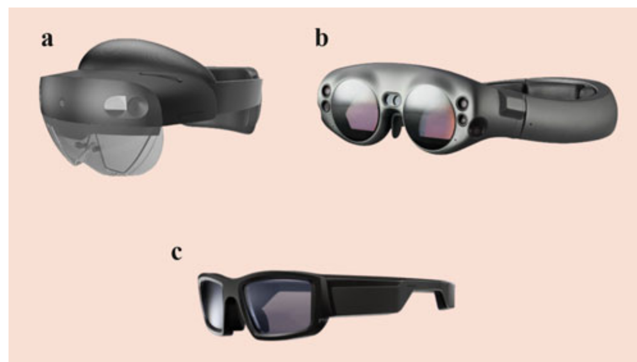


Fig. 8.29 Examples of commercialized waveguide-based AR displays. (**a**) Microsoft Hololens 2, (**b**) Magic Leap One, (**c**) Vuzix Blade

incoming light is reflected towards the eye. The remainder is transmitted to a series of cascades beamsplitters, where a greater percentage of the remaining light is reflected at each successive surface to maintain a uniform brightness across the eyebox [71]. The thickness of the waveguide necessary to have embedded beamsplitters is a main disadvantage of this technique.

A similar technique can be used with the diffraction-based out-couplers to expand the eyebox in both the vertical and horizontal directions. A patent application drawing from Microsoft (Fig. 8.28) illustrates how this could work in practice [78]. Referencing Fig. 8.28, collimated light is incident on component 12, where a grating induces TIR towards the fold region 14. The grating in the fold region diffracts a portion of the light downwards towards the out-coupler 16, while the zeroth-order maintains some intensity to continue horizontally through the fold region. This process happens throughout the fold region to expand the eyebox in the horizontal direction. When the light encounters the out-coupler, the grating diffracts some light from the waveguide to the eye, while the zeroth-order maintains some intensity. This process occurs throughout the height of the out-coupling region, thus expanding the eyebox vertically.

Many commercial AR displays utilize the waveguide architecture as, to date, it has shown the most promise of simultaneously achieving a large eyebox and moderate FOV while having the aesthetics approaching a pair of eyeglasses. Areas to improve for the waveguide architecture are increasing the FOV, limiting ghost reflections from the coupling features, and power consumption from the eyebox expansion. Some

Table 8.1 Summary of the various coupling technologies employed in commercial products in the waveguide architecture. (Adapted from [79])

Coupling technique	Refract mode	Efficiency modulation	Dispersion	Color uniformity	Tunable	Polarization maintaining	Mass producible	Product
Embedded mirrors	Reflect	Complex coating	Little to none	Good	No	Yes	Slicing, coating, polishing	Lumus Ltd. DK50
Micro-prisms	Reflect	Coatings	Little to none	Good	No	Yes	Injection molding	Optinvent Sarl, ORA
Surface relief slanted grating	Diffract	Depth, duty cycle, angle	Strong	Compensation needed	With liquid crystals	No	Nano-imprint lithography	Microsoft Hololens
Surface relief blazed grating	Diffract	Depth	Strong	Compensation needed	With liquid crystals	No	Nano-imprint lithography	Magic Leap One
Surface relief binary grating	Diffract	Depth, duty cycle	Strong	Compensation needed	With liquid crystals	No	Nano-imprint lithography	Magic Leap One
Multi-level surface relief grating	Diffract	Depth, duty cycle	Strong	Compensation needed	With liquid crystals	Possible	Nano-imprint lithography	Wave Optics
Polymer hologram	Diffract	Index change	Strong	Compensation needed	With shear	No	Nano-imprint lithography	Sony Ltd.
Volume hologram	Diffract	Index change	Strong	OK	Electrically	No	Exposure	Digilens Corp
Photopolymer hologram	Diffract	Index change	Minimal	OK	No	No	Multiple exposure	Akonia Corp (now Apple)
Resonant waveguide grating	Diffract	Depth, duty cycle	Can be mitigated	NA	With liquid crystals	Possible	Roll to roll nano imprint lithography	CSEM / Resonant Screens
Metasurface	Diffract	Materials, patterns, coating	Strong, but can be multiplexed	Compensation needed	With liquid crystals	Possible	Nano-imprint lithography	Metalenz Corp.

examples of commercialized waveguide AR displays are shown in Fig. 8.29. Kress provides a thorough review of many waveguide technologies, including a table summarizing the various coupling methods (adapted here in Table 8.1, for convenience) [79].

8.6 Areas for Improvement

The iPhone equivalent of AR displays has yet to be developed and will require a slew of improvements before one can be realized. The most critical area for improvement is the aesthetic of the device while not sacrificing the optical performance. As seen in the previous section, few known architectures have the potential to combine eyeglass-like aesthetics, impressive FOV, and large eyebox. In pursuit of that goal, expanding the FOV can enable killer applications that can highlight the capabilities of an AR display. Expanding the eyebox will not only make the device less frustrating and easier to use, but it will also allow for the accommodation of a larger group of users. Interpupillary distance is the distance between a person's eyes and can vary by tens of millimeters over the population. A large eyebox solves the critical problem of requiring a custom AR display for each person and will allow companies to produce only a few models that fit most of the population. The **vergence-accommodation conflict** poses a challenge for long duration use of an AR/VR display without side effects. It refers to the problematic disparity between the apparent location of a virtual image and the focusing of the eye required to view the image. There are efforts to tackle this obstacle from both the hardware and psychological side [80–85].

A common term relevant in technology development for the military is SWaP, or Size, Weight, and Power. Soldiers know that when wearing or using a piece of equipment for a long duration, the size and weight of the device is critical. Discomfort can easily and significantly interfere with daily tasks, especially for devices worn on the head. The same principle is true for non-military devices as well. A successful AR display will be one that is light enough to be comfortably worn for hours with sufficient battery life to be used continuously throughout the day.

8.7 Components and Techniques for AR Displays of the Future

New techniques and components must be implemented into the existing architectures or leveraged to create new architectures altogether to meet the demands for a successful wearable AR display as outlined above. Three potential game-changing technologies are pupil-steering/eye-tracking, freeform optics, and metasurfaces.

8.7.1 Pupil-Steering and Eye-Tracking

The idea of **pupil-steering** is tied to the principle of eye-tracking and could substantially impact the electronics and optics of AR displays. Pupil-steering refers to dynamically moving the eyebox position in real time to be coincident with the user's eye. The optical design benefit of pupil-steering is that the eyebox would only need to be large enough to match the eye's entrance pupil (about 4 mm). A smaller eyebox yields a slower f-number and less image aberrations for the imaging optics located prior to the steering component. The imaging optics would require fewer and smaller optical elements to generate a crisp image, thus decreasing the size and weight of the whole system. The electronic benefit of using pupil-steering is that the light generated by the light engine is spread over less area at the eyebox, drawing comparatively less power and extending the battery life and lifetime of the light source.

Pupil-steering addresses nearly every item in the AR display improvements section above. However, implementing pupil-steering comes with its own significant challenges. A dynamic eyebox means that somewhere in the system, there must be a dynamic element responsible for moving the eyebox. The mechanism to accomplish this task is a current research topic [86–91]. Additionally, the eyebox cannot be moved to the eye if the eye cannot be located with submillimeter accuracy using **eye-tracking**. The integration of eye-tracking in head-worn displays by design was first proposed by Rolland and Vaissie [92]. Methods for accomplishing precise eye-tracking within portable devices are under active investigation [93–103]. Eye-tracking can be extended to the concept of a **foveated display**, where a high-resolution inset within a larger but lower-resolution image follows the eye's gaze, lessening the tradeoff between FOV and resolution [104].

8.7.2 Freeform Optics

Earlier in this chapter, **freeform optics** were briefly introduced as optical surfaces whose shapes do not have rotational or translational symmetry [105]. Because the optical path of AR displays must bend around the shape of the head and conform to the mechanics of a pair of eyeglasses, using powered optics to image and bend the light simultaneously is highly leveraged. In this case, the rotational symmetry of the optical system is broken, resulting in aberrations that are rotationally variant. Freeform optics have the unique capability to correct the aberrations that are rotationally variant, enabling the most compact solutions [106–108]. In pursuit of a successful AR display, using the minimum number of optical elements is critical, and freeform optics are the surface shapes that allow for this reduction.

Design techniques with freeform optics [63, 67, 109–113] have recently matured to the point where the bottlenecks are now the optical testing and cost-effective mass production of the surfaces [114–122]. Interferometric techniques that have been perfected for spherical surfaces are not applicable to freeform surfaces with steep slopes that cause interference fringes to be too dense on the sensor. The metrology of freeform surfaces is a current research topic [123–127]. The fabrication of freeform surfaces is most commonly performed with sub-aperture techniques, which are not conducive to the mass production necessary for a consumer market product. Molding and replication techniques are being investigated in combination with high-speed and high-accuracy diamond machining of molds to yield mass-production solutions [122, 128].

8.7.3 Metasurfaces

The emerging technology of metasurfaces has the potential to significantly impact the development of AR displays. The term "metasurface" was given to a class of optical elements that manipulate the behavior of light using sub-wavelength elements patterned on a substrate. The particular metasurfaces that concern AR displays are designed to manipulate the direction, polarization, or focusing of light in the visible spectrum. At a high level, metasurfaces operate by altering the phase of an incident beam of light through the interaction and response of the light's electromagnetic field with sub-wavelength elements that decorate the substrate [129–132]. The elements are patterned in such a way as to provide the necessary phase to accomplish the desired task. Whereas a traditional refracting lens focuses light because of its shape and material creating a parabolic phase distribution across the aperture, a metasurface can accomplish the same task on a flat substrate by using sub-wavelength elements to induce the same phase variation across the aperture of the part.

In practice, metasurfaces spatially dictate the phase change across a surface with similarity with what freeform surfaces do but with distinctions. The differences may offer some opportunities for metasurfaces that are advantageous over conventional refractive or reflective optics. Metasurfaces are traditionally manufactured on a flat substrate of uniform thickness and can lead to lightweight structures. Like a diffractive structure, a metasurface functions uniquely across various wavelengths, so there is the potential to multiplex patterns on the same substrate to influence multiple wavelengths simultaneously. With traditional optics, multiple lenses with different glasses or off-axis reflective optics are required. Finally, lenses are manufactured by grinding and polishing with mass production potential via molding or replication. On the other hand, metasurfaces are fabricated on machines that operate similar to those

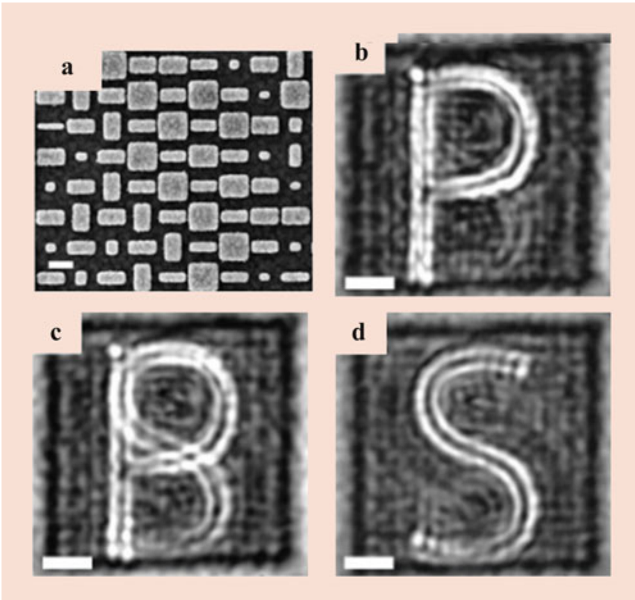


Fig. 8.30 Polarized light interacts with a metasurface hologram with the unit cell shown in (**a**), creating images at the observation plane for (**b**) P-polarized, (**c**) 45-degree polarized, and (**d**) S-polarized light. These images illustrate the polarization selectivity of metasurfaces. (Adapted from [135])

that produce billions of computer chips, providing an en masse fabrication potential [133, 134]. The phase response of metasurfaces is polarization specific, which is an exploitable property. Metasurfaces can produce a phase response to one polarization state of light and a separate phase response to a different polarization state of light. Cheng et al. demonstrated the polarization selectivity in metasurface holograms, as shown in Fig. 8.30 [135].

In their current state, metasurfaces face plenty of challenges to overcome before they can compete with traditional optics. In design, metasurfaces are simulated using finite-difference time-domain software, making designing a metasurface a computationally heavy task. This aspect limits the physical size of the part that can be feasibly simulated and the extent to which the designs can be optimized. Furthermore, the electron beam lithography machines on which metasurfaces are typically fabricated currently limit the maximum part size. Many of the demonstrated designs are <1 mm in diameter and have focal lengths in the hundreds of microns range, which do not have much practical use in AR displays. The phase response of metasurfaces is angle and wavelength-dependent, so designing metasurfaces over a large range of incidence angles and wavelengths is challenging. While multiplexing can yield multiple responses for separate ranges of incidence angles and wavelengths, one must consider how the various patterns interfere. Efficiency in the desired order of diffraction and the simultaneous suppression of ghost images, as applicable for any diffractive structure, is another area where metasurfaces must improve to be competitive with traditional optics. All these issues are areas of active research [129, 133, 136–141]. **Metaform** optics – optics consisting of a metasurface written directly on a freeform substrate – can open the door to new design areas where the benefits of each technology can be harnessed [142].

8.8 Conclusion

As has been demonstrated throughout this chapter, the optics of AR displays are critical to the user experience of an AR display. Properties such as FOV and resolution influence the perceptual impact of the display, whereas parameters such as the eyebox diameter, size, and weight impact the comfort level when wearing the display. It is also critical to recognize that aesthetic of the display is a critical component to entering the consumer market, together with keeping the cost affordable for the masses. As discussed, there are significant challenges to overcome, which set a rich creative and technical playground for research on the numerous aspects of AR displays. Maybe metasurfaces, freeform optics, metaforms, or even a currently unknown technology will enable a new generation of optical architectures for AR displays that will become the gold standard. We cannot predict what will happen in the next decade of AR display development, but we know the ride will be simultaneously humbling and rewarding.

References

1. Digi-Capital: $80+ billion AR/VR market could diverge before converging (2019). Retrieved 10/24/2019, 2019, https://www.digi-capital.com/news/2019/05/80-billion-ar-vr-market-could-diverge-before-converging/
2. Sutherland, I.E.: A head-mounted three-dimensional display. In: Fall Joint Computer conference, AFIPS Conference Proceedings, pp. 757–764 (1968)
3. Rolland, J.P.: AR/VR displays: engineering the ultimate augmented reality display: paths towards a digital window into the world. Laser Focus World, June 1, 2018
4. Rash, C.E., Martin, J.S.: The Impact of the U.S. Army's AH-64 Helmet Mounted Display on Future Aviation Helmet Design, p. ADA202984. Defense Technical Information Center (1988)
5. Davis, S., Nesbitt, K., Nalivaiko, E.: Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. In: Proceedings of the 11th Australasian Conference on Interactive Entertainment, vol. 167, pp. 3–14 (2015)
6. Rebenitsch, L., Owen, C.: Individual variation in susceptibility to cybersickness. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, pp. 309–317. ACM, Honolulu (2014)
7. Reason, J.T.: Motion sickness adaptation: a neural mismatch model. J. R. Soc. Med. **71**, 819–829 (1978)
8. Kolasinski, E.M., Gilson, R.D.: Simulator sickness and related findings in a virtual environment. Proc. Human Factors Ergon. Soc. Annu. Meet. **42**, 1511–1515 (1998)

9. Johnson, D.M.: Introduction to and Review of Simulator Sickness Research, p. 1832. U.S. Army Research Institute for the Behavioral and Social Sciences (2005)
10. Jones, J.A., Swan, J.E., Singh, G., Kolstad, E., Ellis, S.R.: The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In: Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization, pp. 9–14. ACM, Los Angeles (2008)
11. Foote, B., Melzer, J.: A history of helmet mounted displays. In: Proc. SPIE 9470, Defense + Security (2015)
12. Li, H., Zhang, X., Shi, G., Qu, H., Wu, Y., Zhang, J.: Review and analysis of avionic helmet-mounted displays. Opt. Eng. **52**, 110901 (2013). https://doi.org/10.1117/1.OE.52.11.110901
13. Sweatt, W.C.: Describing holographic optical elements as lenses. J. Opt. Soc. Am. **67**, 803–808 (1977)
14. Chen, C.W.: Helmet Visor Display Employing Reflective, Refractive and Diffractive Optical Elements. US 5,526,183 (1996)
15. Wood, R.B., Hayford, M.J.: Holographic and classical head up display technology for commercial and fighter aircraft. In: Proc. SPIE 0883, Holographic Optics: Design and Applications, pp. 36–52 (1988)
16. Sisodia, A., Riser, A., Bayer, M., McGuire, J.P., Advanced helmet mounted display (AHMD) for simulator applications. In: Proc. SPIE 6224, Helmet- and Head-Mounted Displays XI: Technology and Applications (2006)
17. Walker, B.H.: Optical Design for Visual Systems. SPIE Press (2000)
18. Navarro, R., Santamaría, J., Bescós, J.: Accommodation-dependent model of the human eye with aspherics. J. Opt. Soc. Am. A. **2**, 1273–1280 (1985)
19. Donders, F.C., Doijer, D.: The location of the pivot point of the eye. Part II. Strabismus. **24**, 184–188 (2016)
20. Gross, H., Blechinger, F., Achtner, B.: Human eye. In: Handbook of Optical Systems, vol. 4, pp. 1–87 (2008)
21. Cakmakci, O., Hoffman, D.M., Balram, N.: 3D Eyebox in augmented and virtual reality optics. SID Symp. Digest Tech. Pap. **50**, 438–441 (2019)
22. Guidelines on ergonomic criteria for bridge equipment and layout. International Maritime Organization MSC/Circ.982 (2000)
23. Hartley, R., Kang, S.B.: Parameter-free radial distortion correction with center of distortion estimation. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1309–1321 (2007)
24. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000)
25. Guo, Q., Tang, H., Schmitz, A., Zhang, W., Lou, Y., Fix, A., Lovegrove, S., Strasdat, H.M.: Raycast calibration for augmented reality HMDs with off-axis reflective combiners. In: 2020 IEEE International Conference on Computational Photography (ICCP), pp. 1–12 (2020)
26. Bauer, A., Vo, S., Parkins, K., Rodriguez, F., Cakmakci, O., Rolland, J.P.: Computational optical distortion correction using a radial basis function-based mapping method. Opt. Express. **20**, 14906–14920 (2012)
27. Krajancich, B., Kellnhofer, P., Wetzstein, G.: Optimizing depth perception in virtual and augmented reality through gaze-contingent stereo rendering. ACM Trans. Graph. **39**, Article 269 (2020)
28. Gohri, V., Espuno, L., Haas, G., Doyeux, H., Sarrasin, D., Billoint, O., Demars, P.: High brightness green OLED microdisplay with high resolution and low power consumption. In: Proceedings of the XXXI International Display Research Conference Eurodisplay 2011, pp. 19–22. SID, Arcachon (2011)
29. Anandan, M., Ghosh, A.P.: Apparatus and Method for Direct Patterning of an Organic Material Using an Electrostatic Mask. US Patent No. 10,903,427 (2021)
30. Jiang, H.X., Lin, J.Y.: Nitride micro-LEDs and beyond – a decade progress review. Opt. Express. **21**, A475–A484 (2013)
31. Ding, K., Avrutin, V., Izyumskaya, N., Özgur, Ü., Morkoç, H.: Micro-LEDs, a manufacturability perspective. Appl. Sci. **9**, 1206 (2019)
32. Huang, Y., Hsiang, E.-L., Deng, M.-Y., Wu, S.-T.: Mini-LED, micro-LED and OLED displays: present status and future perspectives. Light: Sci. Appl. **9**, 105 (2020)
33. Rossetti, M., Castiglia, A., Malinverni, M., Mounir, C., Matuschek, N., Duelk, M., Vélez, C.: RGB superluminescent diodes for AR micro-displays. SID Symp. Digest Tech. Pap. **49**, 17–20 (2018)
34. Li, Y.-W., Lin, C.-W., Chen, K.-Y., Fan-Chiang, K.-H., Kuo, H.-C., Tsai, H.-C.: Front-lit LCOS for wearable applications. SID Symp. Digest Tech. Pap. **45**, 234–236 (2014)
35. McGuire, J.: Personal Display Using an Off-Axis Illuminator. US2007/0177275A1 (2007)
36. Grievenkamp, J.E.: Field Guide to Geometrical Optics. SPIE Press, Bellingham (2004)
37. Cameron, A.: The application of holographic optical waveguide technology to the Q-Sight family of helmet-mounted displays. In: Proc. SPIE 7326, Defense, Security, and Sensing (2009)
38. Urey, H.: Diffractive exit-pupil expander for display applications. Appl. Opt. **40**, 5840–5851 (2001)
39. Urey, H., Powell, K.D.: Microlens-array-based exit-pupil expander for full-color displays. Appl. Opt. **44**, 4930–4936 (2005)
40. Shi, L., Li, B., Kim, C., Kellnhofer, P., Matusik, W.: Towards real-time photorealistic 3D holography with deep neural networks. Nature. **591**, 234–239 (2021)
41. Cakmakci, O., Rolland, J.: Head-worn displays: a review. J. Display Technol. **2**, 199–216 (2006)
42. Rolland, J.P., Gibson, W., Ariely, D.: Towards quantifying depth and size perception in virtual environments. Presence Teleop. Virt. **4**, 24–49 (1995)
43. Erickson, A., Kim, K., Bruder, G., Welch, G.: A review of visual perception research in optical see-through augmented reality. In: International Conference on Artificial Reality and Telexistence, Eurographics Symposium on Virtual Environments, pp. 1–9 (2020)
44. Gao, Y., Peillard, E., Normand, J.-M., Moreau, G., Liu, Y., Wang, Y.: Influence of virtual objects' shadows and lighting coherence on distance perception in optical see-through augmented reality. J. Soc. Inf. Disp. **28**, 117–135 (2020)
45. McGuire, J.P.: Next-generation head-mounted display. In: Proc. SPIE 7618, Emerging Liquid Crystal Technologies V, pp. 761804–761808 (2010)
46. Rolland, J.P.: Wide-angle, off-axis, see-through head-mounted display. Opt. Eng. **39**, 1760–1767 (2000)
47. Ferrin, F., Droessler, J.G.: Head Gear Display System Using Off-axis Sources. US 5,576,887 (1995)
48. Togino, T.: Visual Display Apparatus. US 5,436,765 (1993)
49. Wei, L., Li, Y., Jing, J., Feng, L., Zhou, J.: Design and fabrication of a compact off-axis see-through head-mounted display using a freeform surface. Opt. Express. **26**, 8550–8565 (2018)
50. Zheng, Z., Liu, X., Li, H., Xu, L.: Design and fabrication of an off-axis see-through head-mounted display with an x–y polynomial surface. Appl. Opt. **49**, 3661–3668 (2010)
51. Wang, J., Liang, Y., Xu, M.: Design of a see-through head-mounted display with a freeform surface. J. Opt. Soc. Korea. **19**, 614–618 (2015)
52. Cakmakci, O., Vo, S., Foroosh, H., Rolland, J.: Application of radial basis functions to shape description in a dual-element off-axis magnifier. Opt. Lett. **33**, 1237–1239 (2008)
53. Cakmakci, O., Moore, B., Foroosh, H., Rolland, J.P.: Optimal local shape description for rotationally non-symmetric optical surface design and analysis. Opt. Express. **16**, 1583–1589 (2008)

54. Bauer, A.: Optical Design with Freeform Surfaces, with Applications in Head-Worn Display Design. Ph.D. Thesis (University of Rochester, Rochester, 2016)
55. Shamir, H., Abraham, S., Zlochin, A., Ashkenazi, A.: Binocular Wide Field of View (WFOV) Wearable Optical Display System. US 2019/0121132A1 (2019)
56. Chang, C., Bang, K., Wetzstein, G., Lee, B., Gao, L.: Toward the next-generation VR/AR optics: a review of holographic near-eye displays from a human-centric perspective. Optica. **7**, 1563–1578 (2020)
57. Choi, M.-H., Ju, Y.-G., Park, J.-H.: Holographic near-eye display with continuously expanded eyebox using two-dimensional replication and angular spectrum wrapping. Opt. Express. **28**, 533–547 (2020)
58. Maimone, A., Georgiou, A., Kollin, J.S.: Holographic near-eye displays for virtual and augmented reality. ACM Trans. Graph. **36**, 1 (2017)
59. Tremblay, E., Guillaumee, M., Ziegler, D., Moser, C., Kilcher, L.: Spatially Separated Exit Pupils in a Head Mounted Display. US 10,338,384 B2 (2019)
60. Iizuka, T., Ishino, T.: Optical Element Combination Optical Element and Observation System Using such Optical Elements. US 6,049,429 (2000)
61. Hoshi, H., Taniguchi, N., Morishima, H., Akiyama, T., Yamazaki, S., Okuyama, A.: Off-axial HMD optical system consisting of aspherical surfaces without rotational symmetry. In: Proc. SPIE 2653, Stereoscopic Displays and Virtual Reality Systems III, pp. 234–242 (1996)
62. Cheng, D., Wang, Y., Hua, H., Talha, M.M.: Design of an optical see-through head-mounted display with a low f-number and large field of view using a freeform prism. Appl. Opt. **48**, 2655–2668 (2009)
63. Takaki, N., Bauer, A., Rolland, J.P.: On-the-fly surface manufacturability constraints for freeform optical design enabled by orthogonal polynomials. Opt. Express. **27**, 6129–6146 (2019)
64. Missig, M.D., Morris, G.M.: Diffractive optics applied to eyepiece design. Appl. Opt. **34**, 2452–2461 (1995)
65. Yee, A.J., Song, W., Takaki, N., Yang, T., Zhao, Y., Ni, Y., Bodell, S.Y., Rolland, J.P., Bentley, J.L., Moore, D.T.: Design of a freeform gradient-index prism for mixed reality head mounted display. In: Proc. SPIE 10676, Photonics Europe (2018)
66. Fuerschbach, K., Rolland, J.P., Thompson, K.P.: Theory of aberration fields for general optical systems with freeform surfaces. Opt. Express. **22**, 26585–26606 (2014)
67. Bauer, A., Schiesser, E.M., Rolland, J.P.: Starting geometry creation and design method for freeform optics. Nat. Commun. **9**, 1756 (2018)
68. Chen, B., Herkommer, A.M.: Alternate optical designs for head-mounted displays with a wide field of view. Appl. Opt. **56**, 901–906 (2017)
69. Cakmakci, O., Martinez, O., Carollo, J.: Optical design of a thin curved lightguide and manufacturing using ophthalmic approaches. In: Proc. SPIE 11062, Digital Optical Technologies (2019)
70. Travis, A., Kollin, J.S., Georgiou, A.: Virtual Image Display with Curved Light Path. US 9759919B2 (2017)
71. Cheng, D., Wang, Y., Xu, C., Song, W., Jin, G.: Design of an ultra-thin near-eye display with geometrical waveguide and freeform optics. Opt. Express. **22**, 20705–20719 (2014)
72. Yang, J., Twardowski, P., Gérard, P., Fontaine, J.: Design of a large field-of-view see-through near to eye display with two geometrical waveguides. Opt. Lett. **41**, 5426–5429 (2016)
73. Pan, J., Hung, H.: Optical design of a compact see-through head-mounted display with light guide plate. J. Disp. Technol. **11**, 223–228 (2015)
74. Zhang, W., Wang, Z., Xu, J.: Research on a surface-relief optical waveguide augmented reality display device. Appl. Opt. **57**, 3720–3729 (2018)
75. Mukawa, H., Akutsu, K., Matsumura, I., Nakano, S., Yoshida, T., Kuwahara, M., Aiki, K., Ogawa, M.: A full color eyewear display using holographic planar waveguides. SID Symp. Digest Tech. Pap. **39**, 89–92 (2008)
76. Rolland, J., Vamivakas, N., Kitt, A., Bauer, A.: Freeform Nanostructured Surface for Virtual and Augmented Reality near Eye Display. US10371951B2 (2019)
77. Mukawa, H., Akutsu, K., Matsumura, I., Nakano, S., Yoshida, T., Kuwahara, M., Aiki, K.: A full-color eyewear display using planar waveguides with reflection volume holograms. J. Soc. Inf. Disp. **17**, 185–193 (2009)
78. Saarikko, P., Kostamo, P., Sainiemi, L., Vallius, T.: Display System. WO 2016/130343 A1 (2016)
79. Kress, B.C.: Optical waveguide combiners for AR headsets: features and limitations. In: Proc. SPIE 11062, Digital Optical Technologies (2019)
80. Rolland, J.P., Krueger, M.W., Goon, A.: Multifocal planes head-mounted displays. Appl. Opt. **39**, 3209–3215 (2000)
81. Schowengerdt, B.T., Seibel, E.J.: True 3-D scanned voxel displays using single or multiple light sources. J. Soc. Inf. Disp. **14**, 135–143 (2006)
82. Koulieris, G.-A., Bui, B., Banks, M.S., Drettakis, G.: Accommodation and comfort in head-mounted displays. ACM Trans. Graph. **36**, 1 (2017)
83. Akşit, K., Lopes, W., Kim, J., Shirley, P., Luebke, D.: Near-eye varifocal augmented reality display using see-through screens. ACM Trans. Graph. **36**, 1 (2017)
84. Matsuda, N., Fix, A., Lanman, D.: Focal surface displays. ACM Trans. Graph. **36**, 1 (2017)
85. Zhan, T., Xiong, J., Zou, J., Wu, S.-T.: Multifocal displays: review and prospect. PhotoniX. **1**, 10 (2020)
86. Chen, H., Weng, Y., Xu, D., Tabiryan, N.V., Wu, S.-T.: Beam steering for virtual/augmented reality displays with a cycloidal diffractive waveplate. Opt. Express. **24**, 7287–7298 (2016)
87. Hedili, M.K., Soner, B., Ulusoy, E., Urey, H.: Light-efficient augmented reality display with steerable eyebox. Opt. Express. **27**, 12572–12581 (2019)
88. Jang, C., Bang, K., Li, G., Lee, B.: Holographic near-eye display with expanded eye-box. ACM Trans. Graph. **37**, 1–14 (2018)
89. Travis, A.R.L., Chen, L., Georgiou, A., Chu, J., Kollin, J.: Wedge guides and pupil steering for mixed reality. J. Soc. Inf. Disp. **26**, 526–533 (2018)
90. Jang, C., Bang, K., Moon, S., Kim, J., Lee, S., Lee, B.: Retinal 3D: augmented reality near-eye display via pupil-tracked light field projection on retina. ACM Trans. Graph. **36**, 1–13 (2017)
91. Park, J.-H., Kim, S.-B.: Optical see-through holographic near-eye-display with eyebox steering and depth of field control. Opt. Express. **26**, 27076–27088 (2018)
92. Vaissie, L., Rolland, J.P.: Head Mounted Display with Eyetracking Capability. US 6,433,760 B1 (2002)
93. Carrasco-Zevallos, O., Nankivil, D., Keller, B., Viehland, C., Lujan, B.J., Izatt, J.A.: Pupil tracking optical coherence tomography for precise control of pupil entry position. Biomed. Opt. Express. **6**, 3405–3419 (2015)
94. Sheehy, C.K., Tiruveedhula, P., Sabesan, R., Roorda, A.: Active eye-tracking for an adaptive optics scanning laser ophthalmoscope. Biomed. Opt. Express. **6**, 2412–2423 (2015)
95. Privitera, C.M., Sabesan, R., Winter, S., Tiruveedhula, P., Roorda, A.: Eye-tracking technology for real-time monitoring of transverse chromatic aberration. Opt. Lett. **41**, 1728–1731 (2016)
96. Chi, J.N., Xing, Y.Y., Liu, L.N., Gou, W.W., Zhang, G.S.: Calibration method for 3D gaze tracking systems. Appl. Opt. **56**, 1536–1541 (2017)

97. Talukder, A., Morookian, J.-M., Monacos, S., Lam, R., LeBaw, C., Lambert, J.L.: Eye-tracking architecture for biometrics and remote monitoring. Appl. Opt. **44**, 693–700 (2005)
98. Barsingerhorn, A.D., Boonstra, F.N., Goossens, H.H.L.M.: Optics of the human cornea influence the accuracy of stereo eye-tracking methods: a simulation study. Biomed. Opt. Express. **8**, 712–725 (2017)
99. Fassi, A., Riboldi, M., Forlani, C.F., Baroni, G.: Optical eye tracking system for noninvasive and automatic monitoring of eye position and movements in radiotherapy treatments of ocular tumors. Appl. Opt. **51**, 2441–2450 (2012)
100. Karakaya, M., Bolme, D., Boehnen, C.: Eye gaze tracking using correlation filters. In: Proc. SPIE 9024, IS&T/SPIE Electronic Imaging (2014)
101. Mestre, C., Gautier, J., Pujol, J.: Robust eye tracking based on multiple corneal reflections for clinical applications. J. Biomed. Opt. **23**, 1–9 (2018)
102. Poletti, M., Rucci, M.: A compact field guide to the study of microsaccades: challenges and functions. Vis. Res. **118**, 83–97 (2016)
103. Sarkar, N.: Eye-tracking System and Method Therefor. US 10,213,105 (2019)
104. Rolland, J.P., Yoshida, A., Davis, L.D., Reif, J.H.: High-resolution inset head-mounted display. Appl. Opt. **37**, 4183–4193 (1998)
105. Rolland, J.P., Davies, M.A., Suleski, T.J., Evans, C., Bauer, A., Lambropoulos, J.C., Falaggis, K.: Freeform optics for imaging. Optica. **8**, 161–176 (2021)
106. Reimers, J., Bauer, A., Thompson, K.P., Rolland, J.P.: Freeform spectrometer enabling increased compactness. Light Sci. Appl. **6**, e17026 (2017)
107. Bauer, A., Pesch, M., Muschaweck, J., Leupelt, F., Rolland, J.P.: All-reflective electronic viewfinder enabled by freeform optics. Opt. Express. **27**, 30597–30605 (2019)
108. Schiesser, E.M., Bauer, A., Rolland, J.P.: Effect of freeform surfaces on the volume and performance of unobscured three mirror imagers in comparison with off-axis rotationally symmetric polynomials. Opt. Express. **27**, 21750–21765 (2019)
109. Takaki, N., Bauer, A., Rolland, J.P.: Degeneracy in freeform surfaces described with orthogonal polynomials. Appl. Opt. **57**, 10348–10354 (2018)
110. Papa, J.C., Howard, J.M., Rolland, J.P.: Starting point designs for freeform four-mirror systems. Opt. Eng. **57**, 101705 (2018)
111. Yang, T., Zhu, J., Wu, X., Jin, G.: Direct design of freeform surfaces and freeform imaging systems with a point-by-point three-dimensional construction-iteration method. Opt. Express. **23**, 10233–10246 (2015)
112. Yang, T., Jin, G.-F., Zhu, J.: Automated design of freeform imaging systems. Light: Sci. Appl. **6**, e17081 (2017)
113. Takaki, N., Papa, J.C., Bauer, A., Rolland, J.P.: Off-axis conics as base surfaces for freeform optics enable null testability. Opt. Express. **28**, 10859–10872 (2020)
114. Tohme, Y.: Trends in ultra-precision machining of freeform optical surfaces. In: Optical Fabrication and Testing 2008. Optical Society of America, Rochester (2008)
115. Horvath, N.W., Davies, M.A.: Concurrent engineering of a next-generation freeform telescope: mechanical design and manufacture. In: Proc. SPIE 10998, SPIE Defense + Commercial Sensing (2019)
116. Kim, S., Chang, S., Pak, S., Lee, K.J., Jeong, B., Lee, G.-j., Kim, G.H., Shin, S.K., Yoo, S.M.: Fabrication of electroless nickel plated aluminum freeform mirror for an infrared off-axis telescope. Appl. Opt. **54**, 10137–10144 (2015)
117. Brunelle, M., Blalock, T., Lynch, T., Ferralli, I., Nelson, J.D.: Larger format freeform fabrication and metrology. In: Proc. SPIE 10692, Optical Systems Design (2018)
118. Hong, Z., Liang, R.: IR-laser assisted additive freeform optics manufacturing. Sci. Rep. **7**, 7145 (2017)
119. Yan, G., Fang, F.: Fabrication of optical freeform molds using slow tool servo with wheel normal grinding. CIRP Ann. **68**, 341–344 (2019)
120. Li, Z., Liu, X., Fang, F., Zhang, X., Zeng, Z., Zhu, L., Yan, N.: Integrated manufacture of a freeform off-axis multi-reflective imaging system without optical alignment. Opt. Express. **26**, 7625–7637 (2018)
121. Li, Z., Fang, F., Chen, J., Zhang, X.: Machining approach of freeform optics on infrared materials via ultra-precision turning. Opt. Express. **25**, 2051–2062 (2017)
122. Fang, F., Zhang, N., Zhang, X.: Precision injection molding of freeform optics. Adv. Opt. Technol. **5**, 303 (2016)
123. Fuerschbach, K., Thompson, K.P., Rolland, J.P.: Interferometric measurement of a concave, phi-polynomial, Zernike mirror. Opt. Lett. **39**, 18–21 (2014)
124. Xu, D., Chaudhuri, R., Rolland, J.P.: Telecentric broadband objective lenses for optical coherence tomography (OCT) in the context of low uncertainty metrology of freeform optical components: from design to testing for wavefront and telecentricity. Opt. Express. **27**, 6184–6200 (2019)
125. Chaudhuri, R., Papa, J., Rolland, J.P.: System design of a single-shot reconfigurable null test using a spatial light modulator for freeform metrology. Opt. Lett. **44**, 2000–2003 (2019)
126. Michalko, A.M., Fienup, J.R.: Transverse translation diverse phase retrieval using soft-edged illumination. Opt. Lett. **43**, 1331–1334 (2018)
127. Noste, T., Evans, C.J., Miller, J.A., Hopper, L.E.: Concurrent engineering of a next-generation freeform telescope: metrology and test. In: Proc. SPIE 10998, SPIE Defense + Commercial Sensing (2019)
128. Gurganus, D., Owen, J.D., Dutterer, B.S., Novak, S., Symmons, A., Davies, M.A.: Precision glass molding of freeform optics. In: Proc. SPIE 10742, Optical Engineering + Applications (2018)
129. Chen, W.T., Zhu, A.Y., Sanjeev, V., Khorasaninejad, M., Shi, Z., Lee, E., Capasso, F.: A broadband achromatic metalens for focusing and imaging in the visible. Nat. Nanotechnol. **13**, 220–226 (2018)
130. Lin, D., Fan, P., Hasman, E., Brongersma, M.L.: Dielectric gradient metasurface optical elements. Science. **345**, 298–302 (2014)
131. Fattal, D., Li, J., Peng, Z., Fiorentino, M., Beausoleil, R.G.: Flat dielectric grating reflectors with focusing abilities. Nat. Photonics. **4**, 466–470 (2010)
132. Chen, W.T., Zhu, A.Y., Capasso, F.: Flat optics with dispersion-engineered metasurfaces. Nat. Rev. Mater. **5**, 604–620 (2020)
133. Park, J.-S., Zhang, S., She, A., Chen, W.-T., Yousef, K.M.A., Capasso, F.: Large-area, single material metalens in the visible: an approach for mass-production using conventional semiconductor manufacturing techniques. In: Conference on Lasers and Electro-Optics OSA Technical Digest. Optical Society of America (2019)., STh1O.5
134. She, A., Zhang, S., Shian, S., Clarke, D.R., Capasso, F.: Large area metalenses: design, characterization, and mass manufacturing. Opt. Express. **26**, 1573–1585 (2018)
135. Cheng, F., Ding, L., Qiu, L., Nikolov, D., Bauer, A., Rolland, J.P., Vamivakas, A.N.: Polarization-switchable holograms based on efficient, broadband multifunctional metasurfaces in the visible regime. Opt. Express. **26**, 30678–30688 (2018)
136. Nikolov, D.K., Cheng, F., Basaran, N., Bauer, A., Rolland, J.P., Vamivakas, A.N.: Long-term efficiency preservation for gradient phase metasurface diffraction gratings in the visible. Opt. Mater. Express. **8**, 2125–2130 (2018)
137. Cheng, F., Qiu, L., Nikolov, D., Bauer, A., Rolland, J.P., Vamivakas, A.N.: Mechanically tunable focusing metamirror in the visible. Opt. Express. **27**, 15194–15204 (2019)

138. Wang, S., Wu, P.C., Su, V.-C., Lai, Y.-C., Chen, M.-K., Kuo, H.Y., Chen, B.H., Chen, Y.H., Huang, T.-T., Wang, J.-H., Lin, R.-M., Kuan, C.-H., Li, T., Wang, Z., Zhu, S., Tsai, D.P.: A broadband achromatic metalens in the visible. Nat. Nanotechnol. **13**, 227–232 (2018)
139. Park, J., Kang, J.-H., Kim, S.J., Liu, X., Brongersma, M.L.: Dynamic reflection phase and polarization control in metasurfaces. Nano Lett. **17**, 407–413 (2017)
140. Lee, G.-Y., Yoon, G., Lee, S.-Y., Yun, H., Cho, J., Lee, K., Kim, H., Rho, J., Lee, B.: Complete amplitude and phase control of light using broadband holographic metasurfaces. Nanoscale. **10**, 4237–4245 (2018)
141. Balli, F., Sultan, M., Lami, S.K., Hastings, J.T.: A hybrid achromatic metalens. Nat. Commun. **11**, 3892 (2020)
142. Nikolov, D., Bauer, A., Cheng, F., Kato, H., Vamivakas, A.N., Rolland, J.P.: Metaform optics: bridging nanophotonics and freeform optics. Sci. Adv. **7**, eabe5112 (2021)

Aaron Bauer is a Research Assistant Professor at the Institute of Optics at the University of Rochester, where he received his Ph.D. in 2016. His research interests include applying novel optical surface types, such as freeform optics and metasurfaces, to real-world applications such as augmented reality. He is also involved in conventional optical system design for a variety of applications.

Jannick P. Rolland earned a Ph.D. in Optical Sciences from the University of Arizona. She worked in Computer Science at UNC-Chapel-Hill and at CREOL-UCF. She is the Brian J. Thompson Professor of Optical Engineering at the University of Rochester, the Director of CeFO, and the CTO of LighTopTech. Her research spans optical engineering innovations in imaging and visualization.

Tracking Systems: Calibration, Hardware, and Peripherals

9

Alexander Plopski, Naoto Ienaga, and Maki Sugimoto

Contents

9.1 Introduction . . . 211

9.2 Multisensor Integration . . . 212

9.3 Calibration Methods . . . 213
9.3.1 Notations . . . 213
9.3.2 Tip Tool Calibration . . . 213
9.3.3 Tracking the Same Target . . . 214
9.3.4 Hand–Eye Calibration . . . 214
9.3.5 Absolute Orientation . . . 215

9.4 Registration Methods . . . 216
9.4.1 Iterative Closest Point . . . 216
9.4.2 Point-Feature-Based Alignment . . . 217

9.5 Inertial Measurement Unit Calibration . . . 217
9.5.1 IMU Bias . . . 218
9.5.2 Sensor Fusion . . . 218
9.5.3 IMU–Camera Calibration . . . 219

9.6 Projector–Camera Calibration . . . 220
9.6.1 Pixel Mapping . . . 220
9.6.2 Spatial Calibration . . . 220

9.7 Optical See-Through Head-Mounted Display Calibration . . . 221
9.7.1 Interaction-Based Methods . . . 221
9.7.2 Interaction-Free Calibration . . . 222
9.7.3 Eye Tracking . . . 223

9.8 Evaluation Methods . . . 223
9.8.1 Objective Measurements . . . 223
9.8.2 Subjective Measurements . . . 224

9.9 Tracking Systems for Sensor Integration . . . 224
9.9.1 Mechanical Links . . . 225
9.9.2 Electromagnetic Sensors . . . 226
9.9.3 Inertial Measurement Units . . . 226
9.9.4 Flex Sensors . . . 226
9.9.5 Radio Signals . . . 226
9.9.6 Camera-Based Motion Capture Systems . . . 227
9.9.7 Markerless Tracking . . . 228
9.9.8 Projection-Based Sensing . . . 230

9.10 Applications . . . 232
9.10.1 Medical Applications . . . 232
9.10.2 Robotic Applications . . . 233
9.10.3 Entertainment Applications . . . 233

9.11 Conclusion . . . 233

References . . . 233

Abstract

A single sensor that tracks the user's pose is usually not sufficient for the creation of compelling augmented reality (AR) owing to system drift, noise, and tracking errors. Recent systems thus combine tracking information from multiple sensors, such as red green blue and infrared cameras, accelerometers, and depth sensors, to improve the illusion of an augmented world. In this chapter, we first introduce calibration methods, tracking sensors, and hardware that can track objects of interest, followed by a discussion of methods that can align the information gathered by these sensors.

Keywords

Tracking systems · Calibration methods · Registration methods · Hardware and tracking applications

9.1 Introduction

Whenever a new augmented reality (AR) application is announced and presented to the public, we are presented with overwhelmingly realistic demos that show virtual content perfectly embedded into the environment. Some examples of these are robots coming out of walls, virtual characters all around us, or users placing virtual furniture and posing with it.

A. Plopski (✉)
Graz University of Technology, Graz, Austria
e-mail: alexander.plopski@icg.tugraz.at

N. Ienaga
University of Tsukuba, Tsukuba, Japan
e-mail: ienaga@iit.tsukuba.ac.jp

M. Sugimoto (✉)
Keio University, Tokyo, Japan
e-mail: sugimoto@ics.keio.ac.jp

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_9

To realize these effects, it is necessary to continuously track the user (device) as they move through the environment. Furthermore, if virtual objects are to realistically interact with the environment – e.g., virtual balls bounce off the floor and walls or be occluded by and occlude real objects, or the environment be reflected in a virtual mirror – it is necessary to obtain information about the environment and align it with the user's view.

Such experiences commonly utilize a variety of sensors distributed throughout the environment and worn by the user that acquire information about the user's pose and view, as well as the environment model (Fig. 9.1). The system in Fig. 9.2 shows an example of AR environments integrating various sensors. The system can overlay a computed background model onto the robot arm based on information from the sensors on the mechanical link structure and the tracked pose of a head-mounted display (HMD).

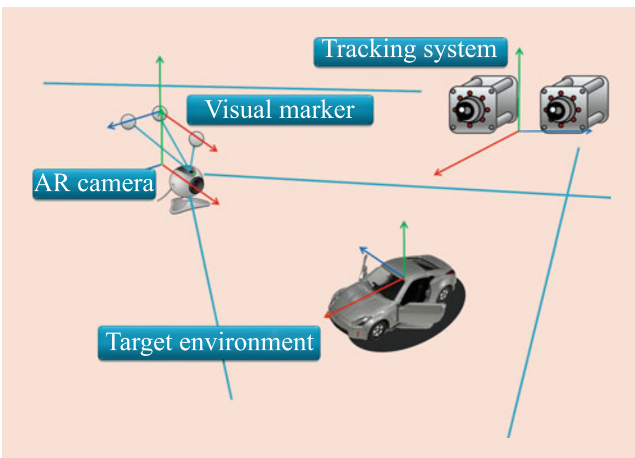


Fig. 9.1 Sensor integration in an AR environment must account for different elements that have their own coordinate systems. These elements could be IR tracking systems, AR cameras, markers, or the target environment

When creating AR experiences, we need to consider the available hardware and its capabilities, which can limit what is possible or enforce restrictions. Furthermore, when multiple sensors are incorporated into the target experience, the information captured by each sensor must be combined in a meaningful way. In this chapter, we review the basics of tracking modalities using a variety of sensors, such as cameras and accelerometers, and discuss common extrinsic calibration methods that can be used to align information obtained by various sensors distributed in the environment.

The previous chapter, "Principles of Object Tracking and Mapping," focused on the fundamental theories of cameras and tracking systems. In this chapter, we introduce methods and devices for sensor integration to make practical augmented reality systems. As a part of the integration methods, we explain the trade-off between modalities.

9.2 Multisensor Integration

Without any tracking, virtual content placed into the environment would move with the user and result in an unrealistic experience. To maintain the position and orientation of virtual objects, it is thus necessary to track the user's movement in the scene.

This can be achieved through basically two different approaches. One is inside-out tracking, tracking features in the world through sensors attached to the user. The other is outside-in tracking, placing features on the user and tracking them with sensors distributed in the environment. Both methods are viable, and the choice of one or the other may depend on the available technology, control over the environment,

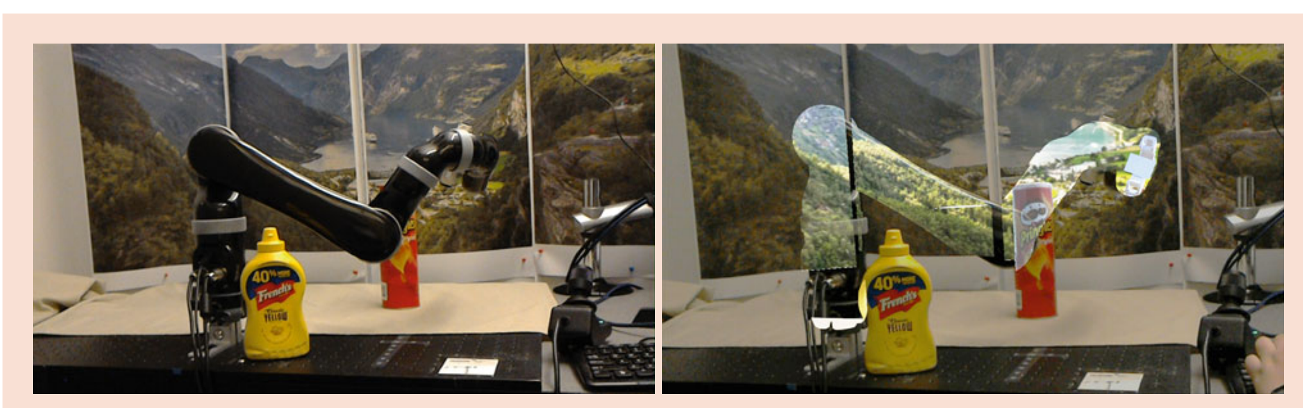


Fig. 9.2 Example of sensor integration in an AR environment. Computer graphics (CG) shown on an OST-HMD are overlaid over the robot, showing the occluded scene to the user. To correctly place the CG over the moving robot arm, the system needs to continuously track the pose of all robot joints and the user's pose as he moves in the environment, and it needs to estimate how to render the graphics so the user perceives them as correctly overlaid (Reprinted with permission from Taylor et al. [109])

or experience parameters. For example, the Oculus Rift CV1 uses outside-in tracking, where a set of cameras placed in the environment detects reflections of infrared (IR) markers on the HMD worn by the user, whereas Oculus Quest and Rift S utilize inside-out tracking, where cameras on the device detect natural features in the environment to determine the user's pose.

With advancements in visual tracking, more and more devices rely on inside-out visual tracking, in which the system in real time detects dominant environment features in images captured by one or more cameras [108]. As the user moves through the environment, the system tracks the location of these features to update the user's pose and adds further features to the environment map. With a monocular camera, the system requires an initial estimate of the user's motion to correctly estimate the scale of the generated map. When multiple cameras are available, each point can be detected by more than one camera and its position triangulated from rays back-projected from each camera. Should the cameras lose track of the user's pose, e.g., owing to rapid movement or a lack of features, these systems can recover the pose by matching (re)discovered features with a database of previously observed and reconstructed feature points.

Despite the improvements in visual tracking, relying on a single device imposes a series of limitations on the user experience, such as a lack of available information. This can not only limit the experience but also make it impossible in some environments. For example, optical tracking will instantly fail if the user points the camera at a uniformly colored surface. This limitation can be overcome by integrating information from a variety of sensors. Mobile phones complement visual tracking information with accelerometer, gyroscope, or GPS information; robots and cars can utilize odometry to approximate their location when other tracking methods fail, and HMDs can provide a limited subset of functions by utilizing gyroscopes in the device to approximate head orientation. Besides the redundancy of the available information and improved robustness, systems can also utilize the strength of each system to improve the overall experience. Although visual tracking is often very robust, it comes with a limited update rate of 30–120 Hz, which can result in a noticeable delay between the updated user pose and the content being rendered on the screen. On the other hand, inertial measurement units (IMUs) are less accurate and more prone to drift, but they can update the measurements at more than 1000 Hz. A system could thus rely on visual tracking for an initial high-quality estimate and update the content being displayed to the user from measurements obtained from the IMUs, resulting in a more consistent experience [48]. In the next section, we discuss commonly used methods to register observations from multiple tracking systems.

9.3 Calibration Methods

When utilizing multiple tracking systems, it is important to understand the spatial relationship of the measurements captured by each system. In the following, we discuss common approaches to estimating the transformations between sensors and common tools as shown in Table 9.1.

9.3.1 Notations

We refer to the coordinate system of an object O as O and to a 3D point P as $\mathbf{P}$. If we describe $\mathbf{P}$ in O's coordinate system, we will refer to it as ${}^O\mathbf{P}$. A transformation ${}^C_O\mathsf{T} = [{}^C_O\mathsf{R}\ {}^C_O\mathbf{t}]$ transforms a point from O to C as ${}^C\mathbf{P} = {}^C_O\mathsf{T}\,{}^O\mathbf{P} = {}^C_O\mathsf{R}\,{}^O\mathbf{P} + {}^C_O\mathbf{t}$. Thus, R is the rotational and $\mathbf{t}$ is the translational component of T. We have to note that the same notation would apply if $\mathbf{P}$ and T were in homogeneous coordinates. Finally, $()^{\mathrm{T}}$ refers to the transpose of a matrix or vector.

9.3.2 Tip Tool Calibration

Tip Tool Calibration is special in that it is not aimed at calibrating the transformation between different systems; rather, it calibrates the position of a rigidly attached point relative to a tracking target. Imagine a pen being held by the user or a mobile device (Fig. 9.3). Being able to accurately determine

Table 9.1 Comparison of different calibration methods

Method	Min. observations	Tracking systems
Tip tool	4	Single
Same target	1	Multiple
Hand–eye	3	Multiple
Absolute orientation	3	Multiple

Fig. 9.3 Example of tip tool being used as a stylus to draw in AR

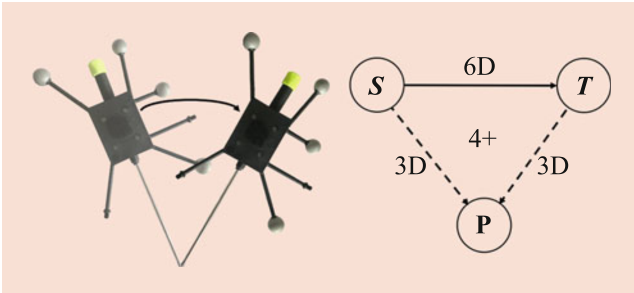


Fig. 9.4 The position of the tool's tip relative to its origin can be estimated by rotating the tip tool around a pivot

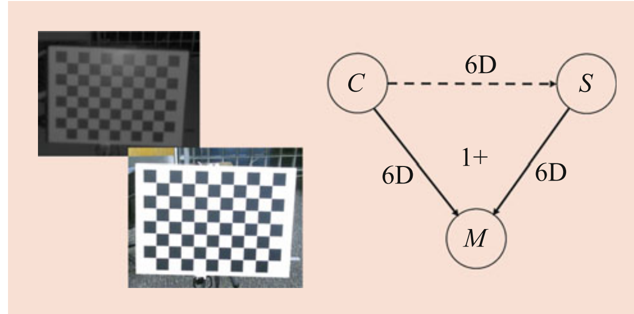


Fig. 9.5 A black and white checkerboard can be tracked by the RGB and IR cameras of the Microsoft Kinect, allowing one to calibrate the spatial relationship of the two sensors

the location of the tip and the axis of the pen enables a variety of applications. For example, if CG is overlaid over the pen, the user could be given the impression that it is a different tool. The tool could also be used to draw in space, interact with virtual content by touching it, or measure the distance between different points in the environment. In the example in Fig. 9.3, a target T, which can be tracked with an external system S, is attached to a screwdriver. To correctly overlay the pen onto the screwdriver, it is necessary to determine the location of the tip $\mathbf{P}$ relative to T. If there exists a point tracked by S that can be touched with the tip of the tool, described as ${}^{S}\mathbf{P}$, the location of ${}^{T}\mathbf{P}$ can be easily recovered as ${}^{T}\mathbf{P} = {}^{T}_{S}\mathsf{T}{}^{S}\mathbf{P}$.

If no such point ${}^{S}\mathbf{P}$ is available, a common calibration approach is the pivot calibration (Fig. 9.4). In this approach, the tip is placed onto an arbitrary point $\mathbf{P} = [x_0, y_0, z_0]^{\mathrm{T}}$, and the tool is rotated around the tip. The basic idea of this approach is that if the tool is rotated around the pivot at its tip, T transcribes a spherical surface around $\mathbf{P}$. Given $n \geq 4$ observations of the tip tool being rotated around the pivot, the goal is to minimize the error

$$e = \sum_{i=1}^{n} R^2 - \left({}^{S}x_i - {}^{S}x_0 + {}^{S}y_i - {}^{S}y_0 + {}^{S}z_i - {}^{S}z_0\right)^2, \quad (9.1)$$

where $[{}^{S}x_i, {}^{S}y_i, {}^{S}z_i]^{\mathrm{T}}$ is the tracked position of T for observation i and R is the radius of the sphere. The parameters of the sphere can be determined through a variety of algorithms, such as sphere fitting [2] or the algebraic one-step [64] or two-step approach [11]. Although the algebraic approaches were found to be preferable to sphere fitting [125], all of them are viable if one accounts for potential outliers.

Although many commercial tip tools place the tracking target T so that its center is on the tool's axis, making it easy to overlay CG onto the tool, custom-made tools may not provide such luxury. An easy way to determine the axis of the tip tool is to pivot the tool once more at $\mathbf{P}$ and rotate it around its axis [19]. When the tool has been pivoted, T transcribes a sphere around the pivot. This time, it transcribes a circle around a point $\mathbf{A}$ that lies on the axis of the tool. The tool can thus be described by the points ${}^{T}\mathbf{A}$ and ${}^{T}\mathbf{P}$.

9.3.3 Tracking the Same Target

The easiest way to determine the transformation between two sensors is to track the same target. For example, a marker could be tracked by multiple cameras, or retroreflective material could be detected either by an IR system and an RGB camera or by an RGB camera and a depth sensor (Fig. 9.5). If both systems (C and S) track a common target M, the transformation between them can be described as

$${}^{S}_{C}\mathsf{T} = {}^{S}_{M}\mathsf{T}{}^{M}_{C}\mathsf{T}. \quad (9.2)$$

Although this is the simplest and likely the most accurate way to determine the transformation between two tracking systems, it requires creative definition and accurate tracking of the common tracking target. Furthermore, as many components in a device may be tracking different modalities to ensure robustness and versatility, common tracking targets may be unavailable. In the following, we review common methods to calibrate rigidly connected sensors.

9.3.4 Hand–Eye Calibration

When two rigidly connected sensors cannot track a common target, the most common way to determine the transformation between them is through hand–eye calibration, which was originally developed for robots holding a camera [111] (Fig. 9.6). It assumes that both sensors can track their pose relative to their respective origin, which does not coincide for the two sensors. For example, the robot can track the pose of the end effector E relative to its base O, whereas a camera C that is attached to it tracks its pose relative to a marker M. This sensor arrangement gives this calibration approach its name as it estimates the transformation between the hand (the robot end effector) and the eye (the camera attached to it). Although initially designed for robot-camera systems, hand–eye calibration can be applied to a variety of tracking configurations, such as to estimate the transformation between a camera and a rigidly attached marker (Fig. 9.7).

The problem can be generally described by two types of equations. The first approach models all parameters of the scene by solving the equation of form

$$ {}^{M}_{C}\mathsf{T}\,{}^{C}_{E}\mathsf{T} = {}^{M}_{O}\mathsf{T}\,{}^{O}_{E}\mathsf{T}, \tag{9.3}$$

where $\mathsf{A} = {}^{M}_{C}\mathsf{T}$ is the pose of the camera relative to a tracked target, ${}^{C}_{E}\mathsf{T}$ is the transformation between the camera and the end effector, ${}^{M}_{O}\mathsf{T}$ is the transformation between the robot base and the target tracked by the camera, and ${}^{O}_{E}\mathsf{T}$ is the pose of the robot's end effector relative to the base. This formulation was originally proposed by Wang [115] and was later refined for use with quaternions and different error metrics [92, 131]. Here, the approach models the spatial relationship of the tracked target to the base, the robot end effector, and the camera at the same time.

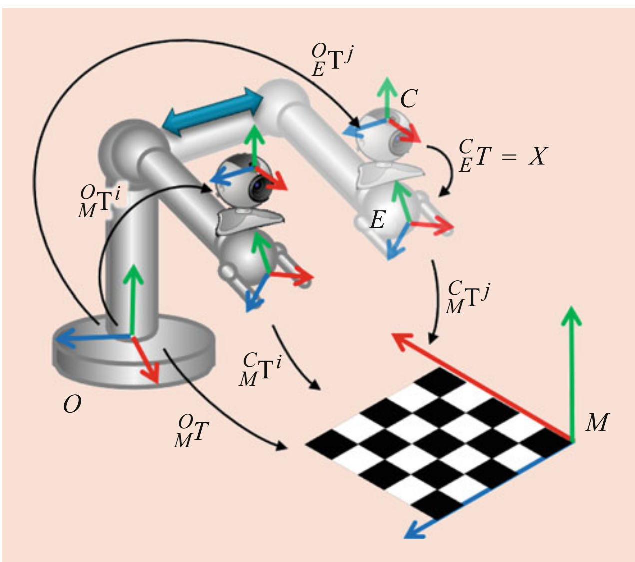


Fig. 9.6 Schematic of a robot holding a camera and tracking a target over multiple frames

In many applications, however, we are only interested in the transformation between the end effector and the camera. In such cases, the transformation can be estimated from a continuous motion as

$$\mathsf{AX} = \mathsf{XB}, \tag{9.4}$$

where $\mathsf{A} = {}^{C}_{M}\mathsf{T}^{i}\,{}^{M}_{C}\mathsf{T}^{j}$ is the transformation between the camera poses for observations i and j, $\mathsf{X} = {}^{C}_{E}\mathsf{T}$ is the transformation between the camera and the end effector, and $\mathsf{B} = {}^{E}_{O}\mathsf{T}^{i}\,{}^{O}_{E}\mathsf{T}^{j}$ is the transformation between the end effector poses at observations i and j. Although both systems can be used to estimate the alignment between the attached targets, the results ultimately depend on the system's error model [101].

A major limitation of this approach is that the similarity of the collected poses can have a significant impact on the reliability of the measurement. This can sometimes lead to highly inaccurate results, e.g., estimating that the sensor is more than a meter away from the robot's hand holding it. Since its inception, different variations of hand–eye calibration have been developed, including adaptations that work from continuous data collection.

9.3.5 Absolute Orientation

An alternative to hand–eye calibration is absolute orientation estimation. Absolute orientation determines the transforma-

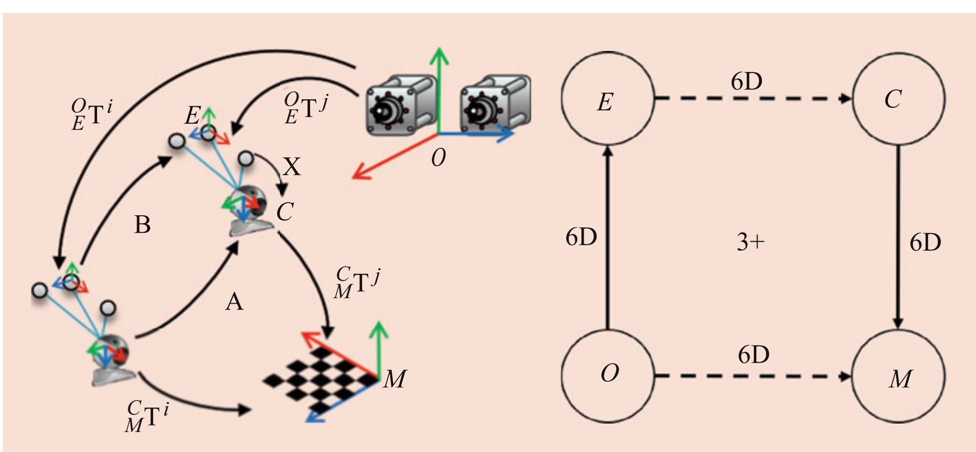


Fig. 9.7 Hand–eye calibration is applied to estimate the relationship between a camera and a tracking target attached to it

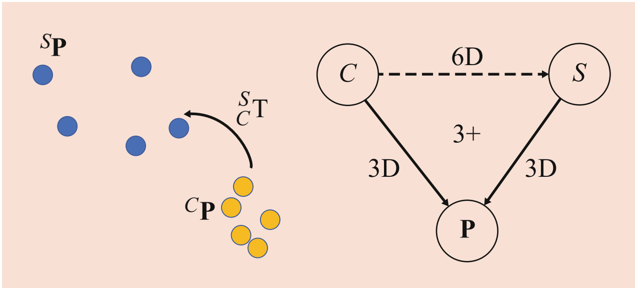


Fig. 9.8 Absolute orientation estimates the transformation between two point clouds from known correspondences. It can be used to estimate the transformation between two sensors S and C that track a point $\mathbf{P}$

tion between two systems that track the same point cloud with perfect point matches [36, 43, 97] (Fig. 9.8).

In the following, we describe a common scenario where the camera C is attached to a marker M tracked by an outside-in tracking system S. The same considerations can be applied to any rigidly attached system, such as a robot arm holding a tracking marker or two rigidly connected cameras or tracking targets. The camera C is tracking the position ${}^C\mathbf{P}$ of a fiducial marker T that is placed somewhere in the environment. Although the outside-in tracking system may not be able to estimate the pose of T, its position ${}^S\mathbf{P}$ can be pointed at with a tip tool. Moving the camera or the marker thus results in a series of observations ${}^S\mathbf{P}_i$ and ${}^C\mathbf{P}_i$. Absolute orientation determines the transformation ${}^S_C\mathsf{T} = \left[{}^S_C\mathsf{R}\ {}^S_C\mathbf{t}\right]$ by minimizing the error

$$e = \sum_{i=1}^{N} \|{}^S_C\mathsf{T}\,{}^C\mathbf{P}_i - {}^S\mathbf{P}_i\|. \tag{9.5}$$

The points ${}^S\mathbf{P}_i$ could be distributed over a large area, and small rotational errors will result in large errors for some points, potentially leading to suboptimal results. To address this issue, we can remap all points ${}^S\mathbf{P}$ onto ${}^{S'}\mathbf{P}$ and ${}^C\mathbf{P}$ onto ${}^{C'}\mathbf{P}$, where the average distance of the points from their respective mean is $\sqrt{2}$. ${}^{S'}\mathbf{P}$ is thus given as

$$ {}^S\mathbf{M} = \frac{1}{N}\sum_{i=1}^{N} {}^S\mathbf{P}_i \tag{9.6}$$

$$ s_S = \frac{1}{N}\sum_{i=1}^{N} \|{}^S\mathbf{P}_i - {}^S\mathbf{M}\| \tag{9.7}$$

$$ {}^{S'}\mathbf{P} = \frac{1}{s_S}\sum_{i=1}^{N} {}^S\mathbf{P}_i - {}^S\mathbf{M}. \tag{9.8}$$

${}^{C'}\mathbf{P}_i$, ${}^C\mathbf{M}$, and s_C can be computed in a similar manner. ${}^{C'}\mathbf{P}$ and ${}^{S'}\mathbf{P}$ are now distributed around $[0\ 0\ 0]^\mathrm{T}$, have the same scale, and differ only by rotation ${}^{S'}_{C'}\mathsf{R}$ around the origin. The rotation that aligns these point clouds can be recovered through the singular value decomposition (SVD) of a matrix

$$\mathsf{M} = \sum_{i=1}^{N} {}^{S'}\mathbf{P}_i ({}^{C'}\mathbf{P}_i)^\mathrm{T}, \text{ where} \tag{9.9}$$

$$\mathsf{M} = \mathsf{U}\Sigma\mathsf{V}^\mathrm{T}, \text{ and} \tag{9.10}$$

$${}^{S'}_{C'}\mathsf{R} = \mathsf{U}\mathsf{V}^\mathrm{T}. \tag{9.11}$$

Finally, the transformation ${}^S_C\mathsf{T}$ can be computed as

$${}^S_C\mathsf{T} = \left[\tfrac{1}{s_S}\mathsf{U}\mathsf{V}^\mathrm{T} s_C\ \left({}^S\mathbf{M} - \mathsf{U}\mathsf{V}^{\mathrm{T}\,C}\mathbf{M}\right)\right]. \tag{9.12}$$

Absolute orientation performs similarly to hand–eye calibration, but it can be applied in scenarios where the tracked target is not static, or the system can track only the location of the target, but not its pose, e.g., a single sphere.

9.4 Registration Methods

With the increasing variety of sensors and information they provide, it may be difficult to detect distinctive features that can be used to estimate their extrinsic parameters as described above. One example of such a situation is point clouds captured by multiple light detection and ranging devices. In this section we discuss methods for aligning point cloud data.

9.4.1 Iterative Closest Point

A common approach to aligning two point clouds, a point cloud with an existing model, or two models, is the iterative closest point (ICP) approach [9]. The basic idea of ICP is similar to that of absolute orientation: we assume that each point in point cloud C has a corresponding point, in this case unknown, in point cloud S as the closest point. Since it is not possible to estimate the transformation directly, ICP minimizes the alignment error by adjusting the transformation over multiple iterations (Fig. 9.9). Given the current estimation ${}^S_C\mathsf{T}_{i-1}$ from iteration $i-1$, ICP computes for each point ${}^C\mathbf{P}$ the closest point ${}^S\mathbf{P}$ and estimates the transform ${}^S_C\mathsf{T}_i$. If the offset between the two point clouds is relatively small, this approach can recover a good alignment. As the estimation for large point clouds – in particular, finding the closest match – has become increasingly computationally expensive, different optimization schemes have been developed over the years, such as subsampling of the point cloud [22], projecting the point onto the mesh [18],

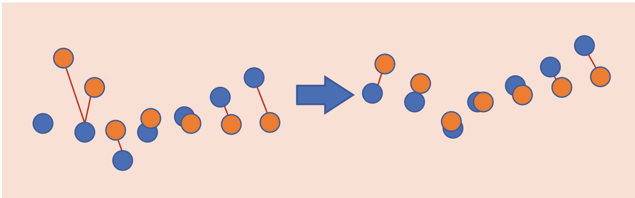


Fig. 9.9 After the closest points between the two point clouds (blue and orange) have been detected, an alignment is computed using absolute orientation, resulting in an improved alignment. The process is repeated until convergence

and intersecting the normal cast from a point with the mesh of an object [14]. Another major limitation of ICP is its dependence on the initial alignment and effects similar areas could have on the alignment quality. If the initial estimation is far from the correct alignment, the result may not converge onto the optimal solution. In ordinary cases, the calculation cost of ICP is high, but, owing to the computing power of general purpose graphics processing units), the ICP method can be used for real-time matching of point clouds to track camera pose and geometry in real time [76].

9.4.2 Point-Feature-Based Alignment

As the initial estimation can have a great impact on the quality of the alignment achieved by ICP, estimating similar features between captured point clouds could serve as an efficient initialization. Although features detected in images captured by cameras can be described by the gradient in their surroundings (e.g., [61]), this approach cannot be used as is for point clouds. However, if we consider the normal of the surface surrounding a point, we can generate unique descriptors for the point.

An example of such descriptors is point-feature histograms (PFHs) [94]. To describe the point feature p, PFHs consider k neighbor points within a specific radius r and calculate the variables for each pair of neighbor points p_i and p_j and estimated normals n_i and n_j (p_i is the point with a smaller angle). Here, the PFH defines a Darboux unit normal vector frame ($u = n_i$, $v = (p_j - p_i) \times u$, $w = u \times v$).

$$\alpha = v \cdot n_j \tag{9.13}$$

$$\phi = (u \cdot (p_j - p_i))/||p_j - p_i||) \tag{9.14}$$

$$\theta = tan^{-1}(w \cdot n_j, u \cdot nj) \tag{9.15}$$

The distribution of α, ϕ, θ represents the point feature of p. Extensions of PFHs in the form of improved 3D feature descriptors, e.g., fast point-feature histograms (FPFHs) [95] and scale invariant point features (SIPFs) [59], provide efficient computation by caching computed values and revising the formulations to match geometries with small computation costs. An example of feature matches using the SIPF descriptor is shown in Fig. 9.10.

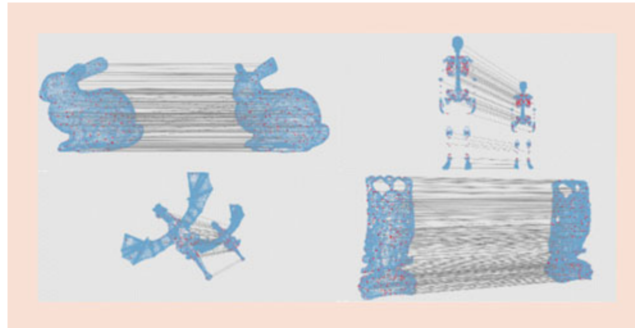


Fig. 9.10 Feature matches between different models with the SIPF descriptor (Adapted with permission from [59])

9.5 Inertial Measurement Unit Calibration

Inertial measurement units (IMUs) are commonly used in combination with cameras to increase the robustness of tracking information. An IMU usually incorporates an accelerometer and a gyroscope that provide measurements at a much higher rate than a camera (often within the range 100–1000 Hz). This allows tracking algorithms to track pose changes between the camera update frames, improve the robustness of the underlying tracking algorithms by providing a good prior on the camera's pose for the current frame, and allow late-stage rendering correction to reduce effects that the rendering latency could have on the placement of virtual content in AR and VR.

A typical design of IMUs has 6 DoF (3 DoF accelerometers + 3 DoF gyroscopes) and 9 DoF sensors (6 DoF accelerometers and gyroscopes + 3 DoF magnetic compasses). The sensor value of an 3DoF accelerometer $\mathbf{a}_s$ can be represented as follows:

$$\mathbf{a}_s = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \tag{9.16}$$

The acceleration due to the earth's gravity can be expressed as follows: the accelerometer is sensitive to gravity, so we can assume that the acceleration due to gravity ($\mathbf{a}_g$) appears on the sensor as follows:

$${}^{w}\mathbf{a}_g = \begin{pmatrix} 0 \\ 0 \\ -g \end{pmatrix} \tag{9.17}$$

Accelerometers are sensitive to acceleration by translation. Also, gravity can be measured by sensors with a rotation matrix as follows:

$$\mathbf{a}_s = {}^s_w\mathsf{R}^w\mathbf{a}_g. \tag{9.18}$$

Here, ${}^s_w\mathsf{R}$ can be described as a combination of rotation matrices around the individual Euler angles. ψ, θ, and ϕ are rotation parameters of the yaw, pitch, and roll angles respectively.

$${}^w_s\mathsf{R} = {}^s_w\mathsf{R}^{-1} = \mathsf{R}_z(\psi)\mathsf{R}_y(\theta)\mathsf{R}_x(\phi) \tag{9.19}$$

Thus, gravity can appear on the sensor values as follows:

$$\begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} = g \begin{pmatrix} sin(\theta) \\ -cos\theta sin(\phi) \\ -cos\theta cos(\phi) \end{pmatrix}. \tag{9.20}$$

It is possible to transform the equation to obtain the rotation parameters θ and ϕ. Accelerometers are not sensitive to rotation in the gravity axis, so ψ cannot be obtained. We need an additional sensor, such as a magnetic compass, to measure rotation.

$$\begin{pmatrix} \theta \\ \phi \end{pmatrix} = \begin{pmatrix} tan^{-1}\frac{a_y}{a_z} \\ -tan^{-1}\frac{a_x}{\sqrt{a_y^2+a_z^2}} \end{pmatrix} \tag{9.21}$$

In addition to the equations above, the rotation parameters can be obtained with the sensor values of gyro sensors.

$$\boldsymbol{\omega}_s = \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \tag{9.22}$$

The gyro sensor is sensitive to the rotation speed of the IMU sensor. The sensor values can be represented as follows:

$${}^w_s\mathsf{R}\boldsymbol{\omega}_s = \begin{pmatrix} 0 \\ 0 \\ \dot{\psi} \end{pmatrix} + \mathbf{R}_z(\dot{\psi}) \begin{pmatrix} 0 \\ \dot{\theta} \\ 0 \end{pmatrix} + \mathbf{R}_z(\dot{\psi})\mathbf{R}_y(\dot{\theta}) \begin{pmatrix} \dot{\phi} \\ 0 \\ 0 \end{pmatrix}. \tag{9.23}$$

The equation can be transformed as follows:

$$\begin{pmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{pmatrix} = \begin{bmatrix} 1 & sin\phi tan\theta & cos\phi tan\theta \\ 0 & cos\phi & -sin\phi \\ 0 & -\frac{sin\phi}{cos\theta} & \frac{cos\phi}{cos\theta} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}. \tag{9.24}$$

Also, the rotation angles of the IMU sensor can be described as follows:

$$\begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \phi_{t-1} \\ \theta_{t-1} \\ \psi_{t-1} \end{bmatrix} + \begin{bmatrix} \dot{\phi}\Delta t \\ \dot{\theta}\Delta t \\ \dot{\psi}\Delta t \end{bmatrix}. \tag{9.25}$$

9.5.1 IMU Bias

Inherently, all IMUs exhibit a bias and drift in the measured data. Although the drift needs to be accounted for through dedicated tracking algorithms, the bias needs to be estimated beforehand to account for sensor-dependent variability. The relationship between measured (m_0) and reported values (m) can be described as

$$\mathbf{m} = \mathsf{S}\mathbf{m}_0 + \mathbf{b}, \tag{9.26}$$

where S is a scaling matrix and $\mathbf{b}$ is a bias. Manufacturers of expensive IMUs often report these values in the corresponding data sheets, but for cheaper devices, they need to be estimated by the users. Some methods recover the parameters by exposing the IMU to motion at a known velocity, e.g., attaching it to a pendulum [6] or estimating them by observing values reported by a resting IMU in different poses [110].

9.5.2 Sensor Fusion

One of the simplest uses of sensor fusion is to fuse the accelerometer and gyro sensor information to obtain an accurate measurement for rotation parameters within a 6 DoF IMU sensor.

An extended Kalman filter (EKF) [28, 44] can be applied to merge multimodal sensor information. We try to predict the true state x_t at time t, and then we measure the state as an observation y_t. Here, f is a function to describe the state in consideration of the previous step. H is another function that indicates a projection of the state during the observation.

$$x_t = f_{t-1}(x_{t-1} + p_{t-1}) + q_{t-1} \tag{9.27}$$

$$y_t = H_t(x_t) + r^t \tag{9.28}$$

We assume system noise $q_t \sim N(0, Q_t)$, observation noise $r_t \sim N(0, R_t)$, and the average of the combination of noises $E[q_t r_t] = 0$. We express the error between an observation and an actual state as $p_{t-1} = y_{t-1} - x_{t-1}$.

Note: $N(\mu, \sigma)$ indicates a normal distribution, with the mean μ, the standard deviation σ, and the expected value E.

In the case of the fusion of 6 DoF IMU sensors, H_t can be considered an identity matrix I. It is possible to make equations with the sensor values as follows:

$$x_t = \begin{bmatrix} \phi_{t-1} + \dot{\phi}\Delta t \\ \theta_{t-1} + \dot{\theta}\Delta t \end{bmatrix} + q_{t-1} \tag{9.29}$$

$$y_t = \begin{bmatrix} \phi_t \\ \theta_t \end{bmatrix} + r^t = \begin{pmatrix} tan^{-1}\frac{a_y}{a_z} \\ -tan^{-1}\frac{a_x}{\sqrt{a_y^2+a_z^2}} \end{pmatrix}. \tag{9.30}$$

For the initialization process, the valuables can be set as follows:

$$\hat{x}_0 = E[x_0] \tag{9.31}$$

$$\hat{P}_0 = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T]. \tag{9.32}$$

During the tracking process, a predicted value $\bar{x}_t$ and prior posterior $\bar{P}_t$ can be described as in Eqs. (9.33) and (9.34). Here, a matrix that indicates the system noise can be expressed as $Q_{t-1} = E[q_{t-1}q_{t-1}^T]$.

$$\bar{x}_t = f_{t-1}(\hat{x}_{t-1}) = \begin{bmatrix} \phi_{t-1} \\ \theta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & sin\phi tan\theta & cos\phi tan\theta \\ 0 & cos\phi & -sin\phi \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \tag{9.33}$$

$$\bar{P}_t = \hat{F}_{t-1}\hat{P}_{t-1}\hat{F}_{t-1}^T + Q_{t-1} \tag{9.34}$$

In Eq. (9.34), we have $\hat{F}$ as a linearized approximate of f as it is made by nonlinear functions for an IMU sensor.

$$\hat{F}_{t-1} = \frac{\partial f_{t-1}(x)}{\partial x}|_{x=\hat{x}_{t+1}} \tag{9.35}$$

$$= \begin{bmatrix} \frac{\phi+\dot{\phi}\Delta t}{\partial\phi} & \frac{\phi+\dot{\phi}\Delta t}{\partial\theta} \\ \frac{\theta+\dot{\theta}\Delta t}{\partial\phi} & \frac{\theta+\dot{\theta}\Delta t}{\partial\theta} \end{bmatrix} \tag{9.36}$$

We can calculate it by assigning $\dot{\phi}$ and $\dot{\theta}$ from Eq. (9.24).

After having actual measurements of y_t, Kalman gain K_t, and estimated state $\hat{x}$, posterior $\hat{P}_t$ can be updated as follows:

$$K_t = \bar{P}_t H_t^T (H_t \bar{P}_t H_t^T + R_t)^{-1} \tag{9.37}$$

$$\hat{x}_t = \bar{x}_t + K_t(y_t - H_t\bar{x}_t) \tag{9.38}$$

$$\hat{P}_t = (I - K_t H_t)\bar{P}_t. \tag{9.39}$$

By iteration of the update process above, the fused angular values are obtained for each step.

9.5.3 IMU–Camera Calibration

Although IMUs can be deployed on their own to track the motion of the attached object, their most common application is to supplement other tracking systems that measure information at a slower rate. In the following, we discuss some common approaches to IMU–camera calibration as a common use case [47, 55, 65, 123] (Fig. 9.11). Generally, we can assume that the camera and IMU will be rigidly connected to each other and that the bias of the IMU does not vary once it has been estimated. In these systems, we need to account for two types of offsets, a temporal measurement offset and a spatial offset between the IMU and the camera. Although early research ignored the temporal offset, which may be acceptable if high accuracy is not necessary, it has been shown to negatively affect visual-inertial tracking [55, 65].

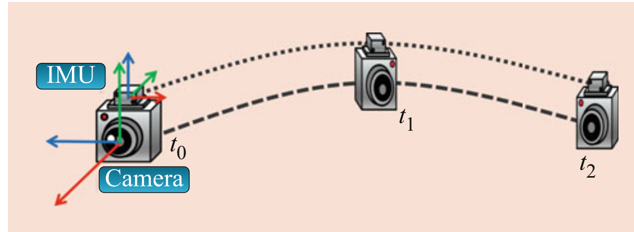


Fig. 9.11 As the camera and IMU are moved in the world, they update their pose independently of each other at different rates. Their spatial relationship can be obtained either by extracting multiple poses taken at the same time or through an EKF given the devices' trajectories

The temporal delay can be estimated either through a cross-correlation or a phase congruency [65]. When the assembly is rotated around a single axis, the angle of rotation θ can be obtained for both the camera and the IMU sensor as ${}^c\theta$ and ${}^{imu}\theta$. The cross-correlation approach is aimed at finding a temporal offset δt that results in the best alignment for all timestamps i as

$$\arg\max_{\delta t} \sum {}^c\theta_{i+\delta t} {}^{imu}\theta_i. \tag{9.40}$$

Although cross-correlation can perform well when δt is small, it can fail when the temporal offset is large. The phase congruency approach overcomes this limitation by searching for a phase shift that correctly aligns phase and amplitude of the recorded angles in the frequency domain. It is important to note that both approaches require the measurements to be obtained at the same sampling rate, as well as gap filling and interpolation, to compensate for differences between the sensors.

A simple approach to computing the spatial offset between the systems is through the previously described hand–eye calibration, utilizing several measurements extracted from a motion. If the calibration is performed beforehand, fiducial markers can be used to simplify the tracking of the assembly [21, 90]. When the calibration is performed online, visual features in the environment need to be tracked between consecutive frames to track the movement of the camera [47, 123].

In recent years, more sophisticated approaches that estimate the parameters of multiple elements of the assembly at the same time have received much attention [56, 78, 89]. Most commonly, these methods describe the system as an EKF that estimates different states (e.g., sensor offset and bias) over the course of a recording [56, 72]. The goal of these methods is to enable a plug-and-play approach that allows the system to estimate its parameters on the fly without a tedious preparation step.

9.6 Projector–Camera Calibration

Projector–camera systems present a special form of multisensory system in that the elements are used not only to track the location of the system but also to project information onto it, thus altering its appearance [10,31] (Fig. 9.12). Generally, we distinguish between two types of projector camera calibrations, mapping of projector pixels onto camera pixels, e.g., for adaptive augmentations of a static environment [119] or planar surfaces [4, 5], and estimating the spatial relationship of the projector and camera.

9.6.1 Pixel Mapping

Projector–camera systems are often deployed in static scenarios where the positions of the projector–camera system and the surface do not change, e.g., a building façade or a table with different sheets of paper on it. In these applications, the projection changes the appearance of the surface and can convert it into a canvas, creating a new experience and appearance, or it can highlight different elements by tracking their location with the camera and augmenting the corresponding pixel in the projector. The mapping is generally determined through the projection of a series of sinusoidal patterns that, in turn, project stripes of white or black pixels horizontally and vertically [25, 132]. After the displayed patterns are captured with the camera and the selection of a suitable threshold to differentiate between white and black pixels has been made, each pixel can be assigned a binary code, 1 if white (illuminated) and 0 otherwise. By matching the captured code with the projected code, we can determine the row and column of each pixel in the projector image, allowing mapping between camera and projector pixels. This information can also be used to estimate the spatial relationship [122] and produce a homography that aligns the captured and projected images.

9.6.2 Spatial Calibration

If the projector–camera system is shifted in the scene, the scene is not static, or the goal is to estimate a surface model of the scene, it is necessary to estimate the spatial relationship between the camera and the projector. Although different methods have been developed to automatically estimate the camera and projector parameters by determining a pixel correspondence [122] as described in the previous section, higher accuracy is generally achieved by separating this process into camera and projector calibration [8, 124]. Although there are different methods to determine this calibration, they are based on tracking the pose of a planar surface by the camera. When the projector overlays a pattern onto this tracked surface, we can estimate the 3D point of the corresponding pixel from the known pose of the planar surface. This allows the projector–camera system to be modeled as a stereo-camera system, where the projector is also calibrated as a pinhole camera and their spatial relationship is estimated from the 3D points of the projected overlay.

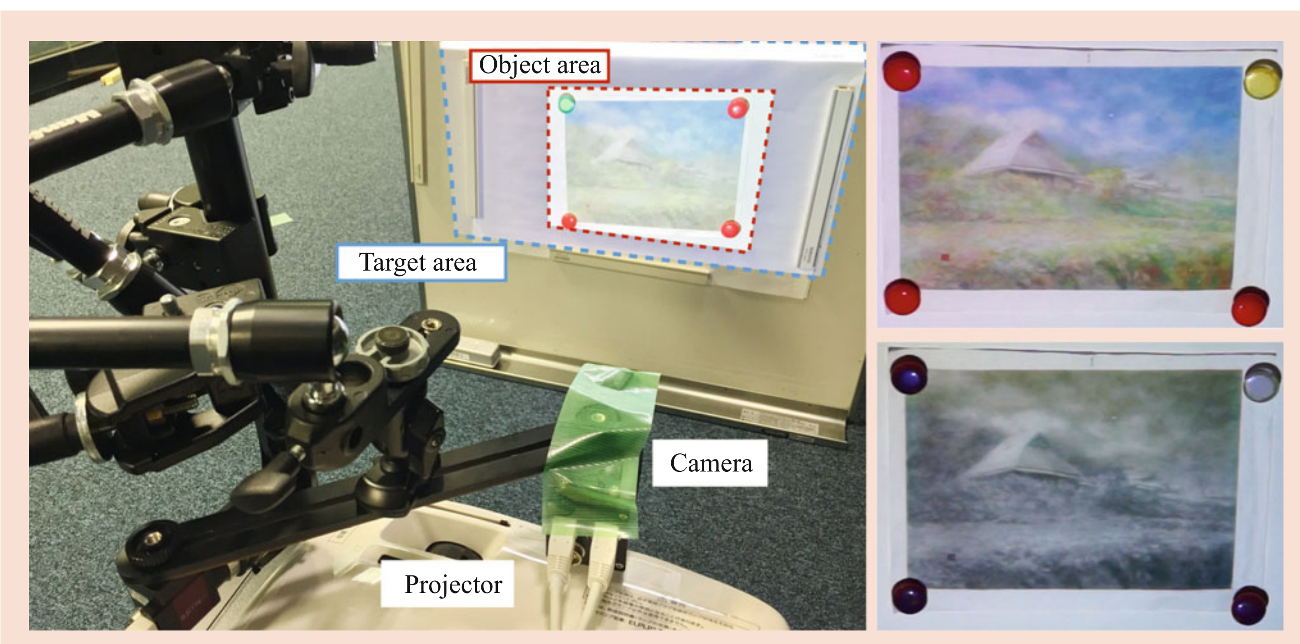


Fig. 9.12 Example of a projector–camera system overlaying content onto a planar surface, thus changing its appearance

9.7 Optical See-Through Head-Mounted Display Calibration

As more and more AR applications target optical see-through head-mounted displays (OST-HMD), we need to consider how the rendered CG will be presented to a user looking through the optics of the OST-HMD. The OST-HMD can be tracked either through an outside-in system with a marker attached to the device, e.g., the Oculus Rift DV1, or an inside-out system with sensors embedded into the headset, e.g., the Microsoft HoloLens. However, rendering CG while assuming that the tracked origin of the HMD coincides with the user's view would result in a misplaced augmentation (Fig. 9.13a) because the user's view through the OST-HMD would not match that of the camera. To correctly render the CG (Fig. 9.13b), it is thus necessary to estimate the user's view when the OST-HMD is worn. Generally, we assume that the user's view through the OST-HMD optics can be modeled as a camera whose center of projection and image plane match the position of the user's eye and the OST-HMD screen respectively. The goal of the calibration process is thus to estimate the parameters of this camera, specifically the intrinsic parameters of the camera and its pose relative to the tracked OST-HMD origin. Generally, we differentiate between estimation methods that require user interaction and those that are interaction-free [30].

9.7.1 Interaction-Based Methods

The most common approach to calibrating an OST-HMD through user interaction is the single-point active-alignment method (SPAAM) [45, 112]. The goal of SPAAM calibration is to estimate the projection matrix P of the user's eye E (modeled as a pinhole camera) relative to the HMD's origin H, so that

$$\mathbf{p} = {}^{E}_{H}\mathsf{P}^{H}\mathbf{P}, \tag{9.41}$$

where a 3D point $\mathbf{P}$ that is tracked relative to H is projected upon $\mathbf{p}$ (Fig. 9.14a). The corresponding pixel to be augmented on the screen is given as

$$u = \frac{\mathbf{p}_x}{\mathbf{p}_z} \text{ and } v = \frac{\mathbf{p}_y}{\mathbf{p}_z}. \tag{9.42}$$

From traditional computer vision [33], we know that the projection matrix of a camera can be estimated from n 2D-3D correspondences $\{\mathbf{p}, \mathbf{P}\}$ by solving the equation

$$\mathsf{B}\begin{bmatrix}\mathsf{P}_1^{\mathrm{T}}\\ \mathsf{P}_2^{\mathrm{T}}\\ \mathsf{P}_3^{\mathrm{T}}\end{bmatrix} = \begin{bmatrix}\mathsf{B}^1\\ \mathsf{B}^2\\ \vdots\\ \mathsf{B}^n\end{bmatrix}\begin{bmatrix}\mathsf{P}_1^{\mathrm{T}}\\ \mathsf{P}_2^{\mathrm{T}}\\ \mathsf{P}_3^{\mathrm{T}}\end{bmatrix} = 0, \tag{9.43}$$

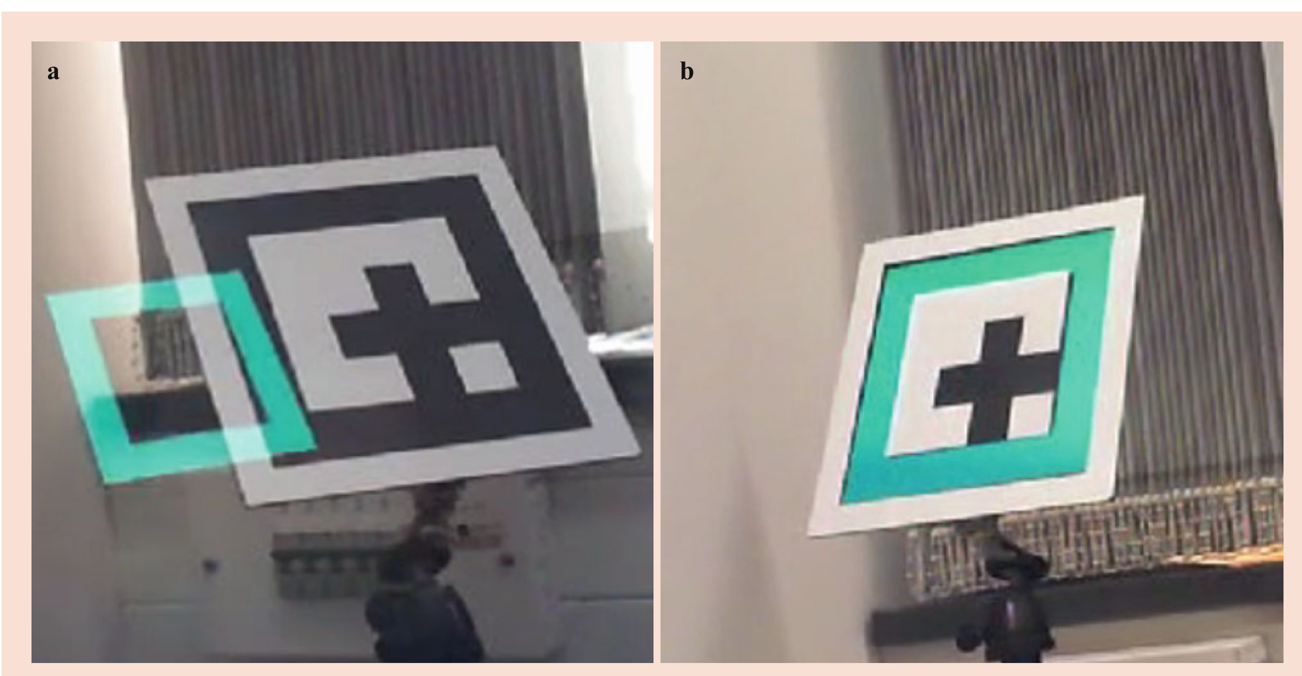


Fig. 9.13 A green rectangle overlaid onto the marker as seen through an OST-HMD (**a**) assuming the user's view matches that of the scene tracking camera and (**b**) after a calibration step

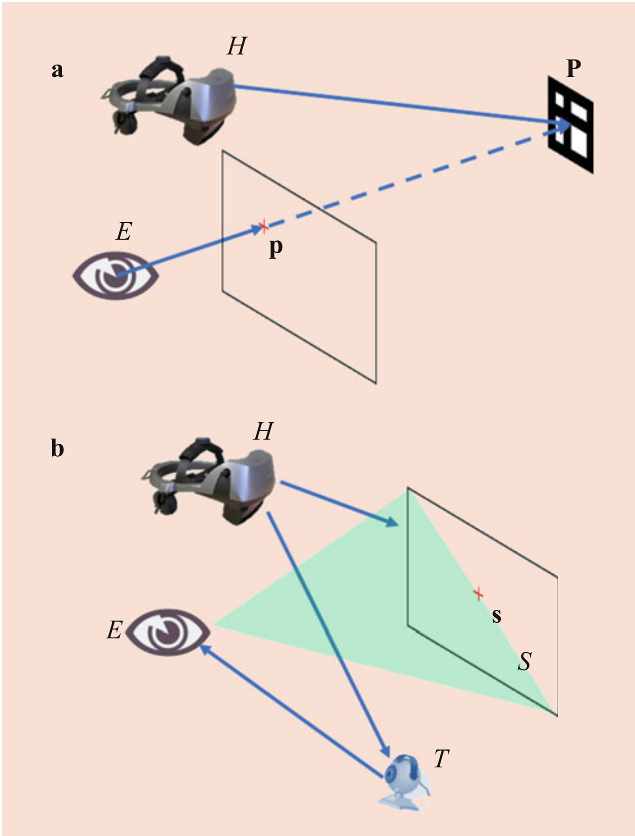


Fig. 9.14 Estimations of the projection models for the (**a**) interaction-based SPAAM and (**b**) interaction-free INDICA calibration approaches. The interaction-based calibration requires the user to align a point **P** that is tracked relative to the HMD frame *H* with multiple pixels **p**. The interaction-free calibration models the screen of the OST-HMD as a plane in space located at **s** and computes the projection matrix by tracking the position of the eye *E* with an eye-tracking camera *T* that is rigidly attached to *H*

where P_j is the j-th row of P and B^i is defined for the i-th pair as

$$\mathsf{B}^i = \begin{bmatrix} \mathbf{P}^i_x & \mathbf{P}^i_y & \mathbf{P}^i_z & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \mathbf{P}^i_x & \mathbf{P}^i_y & \\ \dots & 0 & 0 & -\mathbf{p}^i_x\mathbf{P}^i_x & -\mathbf{p}^i_x\mathbf{P}^i_y & -\mathbf{p}^i_x\mathbf{P}^i_z & -\mathbf{p}^i_x \\ & \mathbf{P}^i_z & 1 & -\mathbf{p}^i_y\mathbf{P}^i_x & -\mathbf{p}^i_y\mathbf{P}^i_y & -\mathbf{p}^i_y\mathbf{P}^i_z & -\mathbf{p}^i_y \end{bmatrix}. \tag{9.44}$$

Equation (9.43) can be solved through the SVD of the matrix B. As the matrix P has 11 degrees of freedom, because the projection is up-to-scale, theoretically it would be sufficient to collect six correspondence pairs; however, in practice, many more, closer to 15–20, are needed to achieve satisfactory accuracy and for robustness against outliers.

When calibrating a camera, the 2D-3D correspondences can be collected by detecting features with a known 3D location, e.g., a texture on a 3D printed model, in a camera image. However, this is not possible in an OST-HMD, as there is no camera to capture the user's view. SPAAM overcomes this problem by embedding the user into the correspondence collection process. As the user can determine what pixel on the screen matches a targeted 3D point, the process randomly highlights a pixel on the screen, e.g., with a cross, and asks the user to align it with a tracked location, e.g., a marker or an IR tracked sphere. After the user has aligned the two in their view and confirmed the completion of this step, the system records the tracked and highlighted locations and selects a new random location that the user needs to align again with the target. As one can imagine, this is a very tedious process, especially if it needs to be completed for each eye separately and every time the HMD is worn or shifts on the head, as this results in a shift of the center of the modeled user-perspective camera. Over the years, different variations of SPAAM have been developed to reduce the demand on the user, such as simultaneous stereo calibration for both eyes [23], reuse of an existing calibration [24], pre-calibration of the system with a camera [66], or utilization of a projection model with fewer degrees of freedom [42].

9.7.2 Interaction-Free Calibration

Despite the improvements and simplifications of interaction-based calibration, its main limitation still remains: users have to perform a calibration step whenever they use the system. One way to circumvent this limitation is to assume that a single calibration will suit all users, thus making the calibration step unnecessary. One example of this is the Microsoft HoloLens, which reuses an initial calibration unless it is triggered by the user. This can create a satisfactory experience as long as one does not require an accurate overlay of the CG over the environment. Another way is to make the alignment process interaction-free, thus removing the burden on the user. This concept was originally proposed as display-relative calibration (DRC) [80]. DRC separates the calibration process into an offline part that estimates a model of the HMD screen relative to its origin and an online part that estimates the position of the user's eyes to create a model similar to SPAAM (Fig. 9.14b). Recently, various interaction-free display calibration (INDICA) methods have investigated how the online part could be performed without user engagement through eye-tracking cameras [37,83]. This idea has also been adopted by the newest generation of OST-HMDs (e.g., HoloLens 2 and MagicLeap 1), which utilize an INDICA approach to automatically adjust rendering parameters to the users through eye tracking.

Interaction-free calibration assumes that the screen S can be modeled as a planar surface relative to H, the origin of the HMD frame. If we now consider a pinhole camera C, whose image plane coincides with the screen S, pointed at the screen so that it is perpendicular to it, i.e., ${}^{S}_{H}\mathsf{R} = {}^{C}_{H}\mathsf{R}$, we can describe the intrinsic parameters of this camera ${}^{C}\mathsf{K}$ as

$$ {}^{C}\mathsf{K} = \mathsf{AS}({}^{C}\mathbf{s}) = \begin{bmatrix} \alpha_x & & \\ & \alpha_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{s}_z & & -\mathbf{s}_x \\ & \mathbf{s}_z & -\mathbf{s}_y \\ & & 1 \end{bmatrix}. \quad (9.45) $$

In this case, ${}^{C}\mathbf{s} = [\mathbf{s}_x, \mathbf{s}_y, \mathbf{s}_z]^{\mathrm{T}}$ is the origin of S and is assumed without loss of generality to project at the center of the image plane. The parameters α_x and α_y convert the projected location into pixel values on the screen.

Correspondingly, the projection matrix ${}^{C}_{H}\mathsf{P}$ to project a 3D point to a pixel can be determined as follows:

$$ {}^{C}_{H}\mathsf{P} = {}^{C}\mathsf{K}\left[{}^{C}_{H}\mathsf{R}\ {}^{C}_{H}\mathbf{t}\right]. \quad (9.46) $$

As we see, the projection matrix can be calculated if we know the parameters of the HMD screen ${}^{H}\mathbf{s}$, α_x, α_y, ${}^{S}_{H}\mathsf{R}$, the orientation of the screen relative to the HMD, and ${}^{C}_{H}\mathbf{t}$, the location of C relative to H. The parameters of the screen are estimated during an offline calibration process, whereas ${}^{C}_{H}\mathbf{t}$ is estimated during runtime.

An alternative to actively constructing the projection matrix from the screen parameters is to reuse the estimated parameters for a known eye location. Assume that the parameters ${}^{C}\mathsf{K}$, ${}^{C}_{H}\mathsf{R}$, and ${}^{C}_{H}\mathbf{t}$ have been estimated by placing a user-perspective camera behind the OST-HMD and capturing a user's view through the HMD screen. When we consider the user's eye E that is tracked relative to the HMD, the offset between C and E can be computed as ${}^{E}_{C}\mathbf{t} = [\mathbf{t}_x, \mathbf{t}_y, \mathbf{t}_z]^{\mathrm{T}}$. The intrinsic parameters of E can be approximated following Eq. (9.45) as

$$ \begin{aligned} {}^{E}\mathsf{K} &= \mathsf{AS}\left({}^{H}\mathbf{s}\right) \\ &= \mathsf{AS}\left({}^{C}\mathbf{s} + {}^{E}_{C}\mathbf{t}\right) \\ &= {}^{C}\mathsf{KS}\left(\frac{\mathbf{t}_x}{\mathbf{s}_z}, \frac{\mathbf{t}_y}{\mathbf{s}_z}, 1 + \frac{\mathbf{t}_z}{\mathbf{s}_z}\right). \end{aligned} \quad (9.47) $$

As we can see, the intrinsic parameters of the user's eye through the HMD screen can be derived from the intrinsic parameters of the camera C. Furthermore, as we can assume that ${}^{E}_{H}\mathsf{R} = {}^{C}_{H}\mathsf{R}$, we can compute ${}^{E}_{H}\mathsf{P}$ by tracking the location of the eye relative to the HMD. It is important to note that all parameters of the OST-HMD screen are scaled depending on the choice of $\mathbf{s}_z$.

Although interaction-free calibration approaches have the potential to simplify the procedure and are increasingly being integrated into commercial HMDs, current models suffer from a series of assumptions that could prevent perfect alignment. For example, although the screen model is assumed to be planar, this is not true in most cases [49]. The model also does not account for distortions of the world by the optics of the OST-HMD [39]. Another drawback is the extensive effort needed to calibrate all components of the system during the offline stage [84] and the impact that errors in such estimations can have during runtime [38].

9.7.3 Eye Tracking

As described earlier, interaction-free calibration assumes that the position of the user's eye relative to the display can be tracked in real time. The most common approach is to detect features of the eye, such as the contour of the iris and pupil [37] or eyelids, or the reflections of LEDs on the cornea [83], and to fit these to a model. A common approach to estimating the pose of the eye is the pupil center corneal reflections (PCCR) method, which models the cornea as a sphere and estimates its location from the reflection of multiple LEDs whose position is known [32]. Although pupil position has been commonly used to estimate gaze, recent methods utilize multiple observations to create a 3D model of the eye [17, 107]. Another common approach is to match the captured appearance of the eye with a database of eye images to derive an estimation [102]. This approach has gained further popularity with the increasing interest in neuronal networks [3]. Although vision-based methods are predominantly used, other methods can be employed as well. Early research into eye tracking relied on coiled lenses attached to the human eye [93]. Recently, laser-based tracking has also been demonstrated as an alternative in HMDs [71].

9.8 Evaluation Methods

The user experience always depends on the quality of the tracking and alignment of the various elements. Depending on the situation, it may be important to achieve alignment that is as accurate as possible to ensure consistent and reliable tracking information. In other scenarios, it may be sufficient to approximate this information to a degree that is indistinguishable from the ground truth for users or satisfies the requirements for that particular application. For that reason, we can distinguish between objective and subjective measurements.

9.8.1 Objective Measurements

The most common evaluation of calibration results is the measurement of a known ground truth (Fig. 9.15a). The ground truth can be obtained from a variety of sources that range from existing CAD models of an observed target to a known distance between printed fiducial markers or observation of the same location from multiple viewpoints to

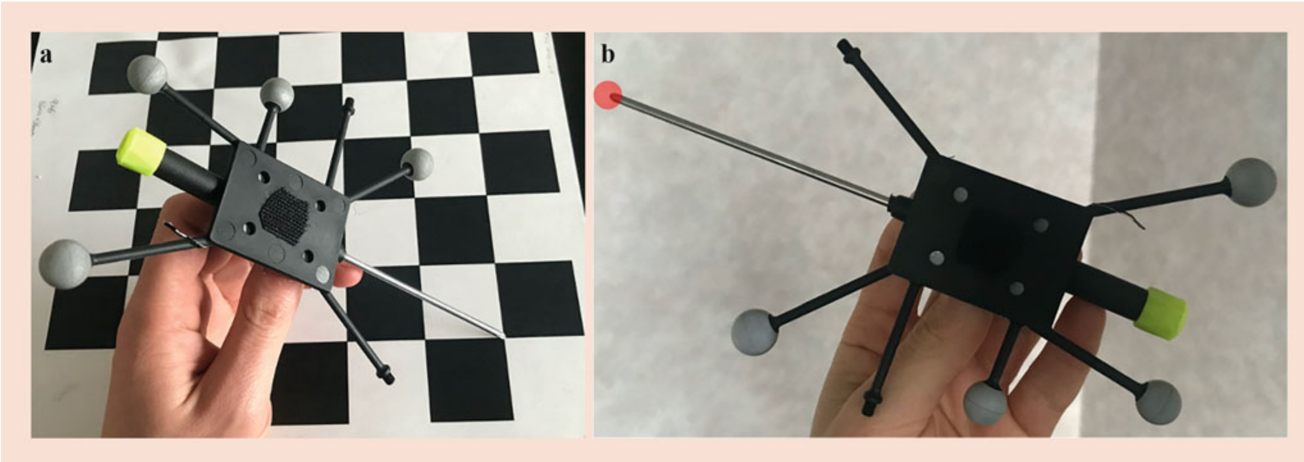


Fig. 9.15 Examples of an evaluation of tip tool calibration. (**a**) The accuracy can be verified objectively by touching known 3D points and measuring the error. (**b**) For subjective evaluation, the tip can be highlighted in an image captured by the camera

determine the stability of the measurements. In various cases, back-projection errors of feature points can be considered a metric for evaluating errors in a camera coordinate system. When designing an objective evaluation, it is imperative to consider the application requirements that must be satisfied. For example, for AR browsers where users will be observing objects that are tens to hundreds of meters away with a simple label rendered at the estimated target location, positional errors within the range of several meters may be acceptable. For that reason, one the GPS could be utilized to get an estimate of the user's location. At the same time, rotation becomes much more important, as a rotational error of a couple degrees may end up causing large errors over long distances.

9.8.2 Subjective Measurements

For applications where information is presented to users, it is often not necessary to measure with perfect accuracy. For users experiencing virtual content, it is sufficient if the misalignment is just below the noticeable level of difference from a perfectly aligned placement. The required accuracy may also depend on the target application. For example, when placing a virtual price label on a coat, it is sufficient to ensure that the label is placed over the coat; errors in the centimeter range do not affect the user experience.

Although objective measurements leave it to the developer to determine if the alignment and tracking accuracy are sufficient and provide a user-independent measurement of errors, subjective measurements focus on how these errors are perceived by users. An example of such an approach is the AR4AR concept [81]. AR4AR provides visual feedback to guide users during a calibration process and to showcase the quality of the calibration. Instead of representing the error as just a value, AR4AR presents the user with a visualization representation of the tracked information, e.g., the location of different cameras or a virtual model overlaid onto the tracked target, which allows users to decide if the accuracy meets their demands. An example is shown in Fig. 9.15b.

In some cases, subjective evaluation may lead to surprising results. For OST-HMDs, researchers found that even though SPAAM [113] resulted in an objectively smaller alignment error than the interaction-free display calibration (INDICA) [37] approach, participants could better estimate the location of virtual objects when INDICA was applied [74]. Furthermore, participants preferred the less tedious calibration and experience of the INDICA method overall.

9.9 Tracking Systems for Sensor Integration

For sensor integration with a video/optical see-through AR system, we can use external tracking systems such as camera-based motion capture systems and projection-based tracking systems. To track a target, we are able to use various sensing modalities. Cameras can be considered the most popular sensing devices in AR for capturing visual information. In addition to cameras, we are able to use other sensing modalities to track the translation and rotation of the target in many cases. Major modalities include optical sensing, mechanical sensing, magnetic sensing, inertial sensing, and so on. Each modality has different characteristics for various conditions. Table 9.2 shows the characteristics of these sensing modalities.

In video see-through configurations, visual images from RGB cameras are used for both visual presentation to users and tracking of visible targets. In such cases, there is no time lag between visual presentation and tracking information.

Table 9.2 Major tracking modalities and characteristics

	Optical occlusion	Magnetic interference	Physical restriction	Accuracy	Calculation cost
Mechanical link	N/A	N/A	High	Good	Low
Magnetic sensor	N/A	Sensitive	Fair	Fair	Low
Inertial measurement unit	N/A	Sensitive[a]	Less	Fair	Low
Camera-based	Sensitive	N/A	Less	Good	High
Projection-based	Sensitive	N/A	Fair	Good	Low

[a]With electromagnetic compass

This is the most important advantage of tracking systems with RGB cameras. The limitations of RGB camera-based tracking include optical occlusions caused by obstacles and limited sensitivity to optical wavelengths.

Ultimately, no single tracking method may be viable for all environments. For example, when users are in a feature-rich environment, the system can rely on the features detected in the environment to determine its location. When moved to an environment that provides outside-in tracking, it could incorporate this information into its estimations, and when in an environment that does not provide any easily available information, it could rely on sensors, such as its GPS or gyroscope, to track the user's movement through the scene [77]. A similar concept is also applied in the Microsoft HoloLens: to conserve power, it adjusts the number of active sensors used to track the environment based on the reliability of the tracking information. We have outlined many different systems that can be used to actively track user and object poses.

In this section, we provide an overview of those sensing modalities to show tracking techniques for advanced AR systems.

9.9.1 Mechanical Links

Utilizing mechanical links with rotation sensors is a long-standing method of measuring positions of target objects in robotics (Fig. 9.16), VR, and AR. Mechanical links have the advantage of faster calculation speed because there is no need to detect complicated features. Also, mechanical links provide accurate tracking information because of the stability of the structure. Given the length of a rigid link and the rotation angle from an origin, it is possible to calculate the end point of the link. Mechanical link structures can be found in various displays, haptic devices, and robots. For example, we can see a mechanical link structure named "the Sword of Damocles," with the original head-mounted display systems [106] invented by Dr. Ivan Sutherland. In the simplest case of measuring the rotation angles of n degree of freedom $\theta_1, \theta_2, \ldots, \theta_n$ with known link lengths $l_1, l_2, \ldots, l_n$, the transformation matrix to calculate the end point of the structure can be calculated by forward kinematics [16] as

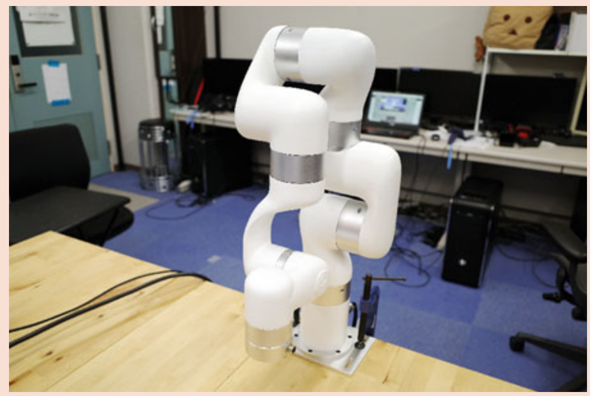


Fig. 9.16 A 7 DoF mechanical link structure of a robot arm. The angle of each joint can be obtained as rotation information

$$\mathsf{T} = \prod_{i=1}^{n} \mathsf{T}_{i-1}^{i}(\Theta_i). \tag{9.48}$$

Here, $\mathsf{T}_{i-1}^{i}(\theta_i)$ is defined with rotation and displacement between the joints i and $i+1$ parallel to the rotation axis, which can be represented by rotation $\alpha_{i,i+1}$ and distance $r_{i,i+1}$ as follows (s and c in this matrix indicate *sin* and *cos* function respectively):

$$\mathsf{T}_{i-1}^{i}(\Theta_i) = \begin{bmatrix} c\theta_i & -s\theta & 0 & 0 \\ s\theta_i & c\theta_i & 0 & 0 \\ 0 & 0 & 1 & l_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & r_{i,i+1} \\ 0 & c\alpha_{i,i+1} & -s\alpha_{i,i+1} & 0 \\ 0 & s\alpha_{i,i+1} & c\alpha_{i,i+1} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9.49}$$

One of the advantages of tracking using mechanical links is the small computation cost. Once we have the angle of a rotation sensor, we can calculate the position of the tracking target in real time at a very small computation cost.

One important limitation of tracking using mechanical links is the structure itself. The range of measurement is limited by physical restrictions. Furthermore, the weight of mechanical structures can increase as degrees of freedom increase.

9.9.2 Electromagnetic Sensors

Electromagnetic sensors can be used for various tracking situations where vision is occluded. For example, medical applications need to be able to track targets inside of patients' bodies. Tracking a nonrigid sensing target in an occluded region requires a method that does not rely on visual information. Electromagnetic sensors are one solution to such a situation. Electromagnetic induction is a way of tracking a sensor with induction coils. In a regular situation where electromagnetic sensing is needed, transmitter coils are activated to generate a magnetic field. Alternating current [86] and direct current [70] methods have been applied to modulate the electromagnetic field.

9.9.3 Inertial Measurement Units

Inertial measurement units (IMUs) are significant emerging sensing devices. IMUs allow us to measure acceleration and translation/rotation information with embedded systems. Owing to the rise of microelectromechanical systems, we can find cheap IMUs in smartphones and car navigation systems in our daily lives.

Besides the previously mentioned applications in visual-inertial odometry, there are various possibilities for using IMU sensors to track targets. One example is human body tracking. Figure 9.17 shows a body suit with embedded IMU sensors. As an advanced study, a method of performing body reconstruction from only six IMU sensors attached to the body has been proposed [114]. Shape reconstruction is a method of estimating the shape of the human body – more precisely, estimating the parameters of a model that can parametrically generate various human body shapes. In such methods, SMPL [60] has been widely used because it is capable of parametrically generating various human body shapes. STAR, which is an improved model of SMPL, has very recently been proposed [79]; it has achieved better generalization than SMPL and reduced the number of parameters by 20%. It will become the future standard in shape reconstruction. Furthermore, IMU sensors can also be used for real-time facial tracking when it is associated with a depth camera [130]. These sensors are able to help to predict high-level features with high frame rate-sensing information.

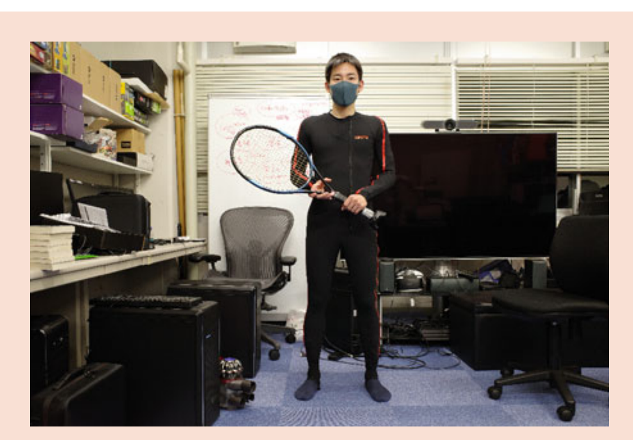


Fig. 9.17 An example of IMU-based motion capture systems. By having an accurate measurement of the length of each of the user's joints, IMU-based motion capture systems are able to predict the position and orientation of each body part

9.9.4 Flex Sensors

Flex sensors can be used to track the deformation properties of a tracking target. The major approaches are optical, conductive, and capacitive.

The optical approach is a traditional method that measures bending with optical fibers. Owing to the reflection and refraction inside of optical fibers, rays conducted to a fiber structure are attenuated according to the radius of the curvature. Data gloves [20] were implemented with the optical approach, and fingers of users were tracked by a plastic fiber structure.

Conductive and capacitive approaches measure electrical character changes due to the deformation of sensors.

The conductive approach uses resistance distributed to the sensor's conductive structure. The conductivity can be affected by the deformation when it is bent owing to the way in which the structure stretches.

The capacitive approach is another way of measuring the deformation by the amount of capacitance made on a flexible structure. The amount of capacitance printed on the structure can be increased or decreased by deformations.

These flexible sensors can be used as embedded sensors inside of specific measurement targets. For example, a set of flexible sensors is used to predict facial expressions under an HMD device associated with a camera-based sensing system [57].

9.9.5 Radio Signals

Wi-Fi and cellular phone signals can be used to track the position of devices. It is possible to predict the position of client devices based on the distribution of signals from stations [91]. There is also research that leverages Wi-Fi signals for human body pose tracking. With Wi-Fi signals, not only 2D pose [129] but also 3D pose [41] can be estimated. Wi-Fi signals make pose estimation possible, even in dark environments and in places with many obstacles, environments in which RGB cameras cannot work. In addition, Wi-Fi signals are superior to RGB cameras in terms of privacy protection.

9.9.6 Camera-Based Motion Capture Systems

Multicamera-based 3D position tracking systems are used in various situations to synchronize objects in physical and virtual space. In most common configurations, these external camera-based tracking systems work as outside-in tracking devices.

For AR applications, integration of such a 3D position tracking system provides several benefits for practical use. For example, when we integrate an outside-in tracking system with a camera-based video see-through AR system, which is an inside-out tracking system, even if there is an obstacle in front of the inside-out tracking camera, it is possible to track the coordinate system through the outside-in tracking.

As described in the previous chapter, marker-based tracking was utilized as the standard method of tracking objects for a long time. In recent years, markerless tracking and simultaneous localization and mapping (SLAM) approaches have gained traction in visual tracking, but marker-based tracking remains one of the most accurate and robust tracking methods. Marker-based tracking refers to tracking of a predefined target object.

In a camera-based motion capture system (Fig. 9.18), a calibration process of internal and external camera parameters is required before actual tracking can take place. In the calibration process, the intrinsic parameters (e.g., parameters of the center of the image plane and the focal length) and extrinsic parameters (6 DoF translation and rotation) of each camera device are obtained by capturing known marker patterns. After calibration, the tracking system is able to calculate 3D position of each tracking target (Fig. 9.19).

A motion capture system can be used to track the 3D position of a reflection marker made of retroreflective material. Also, human body pose tracking is possible when a reflective marker is attached to the body. Compared with other methods, more accurate tracking is possible, but a larger space is required, and the price is higher.

Passive Markers

Passive markers are often tracked by monocular cameras through the use of fiducial patterns. Instead of this kind of approach, with common multicamera motion capture systems, retroreflective materials are used as passive markers (Fig. 9.4). Retroreflective materials have a special reflection index, so that they reflect incoming rays in the same direction as other materials, but with less diffusion. If we have a camera associated with an illumination at the same location, the rays from the illumination reflected by the retroreflective materials come back to the camera with much higher intensity than other ambient lights.

By attaching markers made of retroreflective materials to targets, we are able to track the markers. For a tracking target to be tracked, markers need to be identified. For marker tracking with passive retroreflective markers, distances between markers can be used to identify each marker. The tracking performance of passive markers is reduced in the context of similar spatial structures.

Active Markers

Active markers such as LEDs can be used to track targets with high fidelity. Passive marker patterns can be misidentified owing to the similar structure of rigid bodies or texture in the

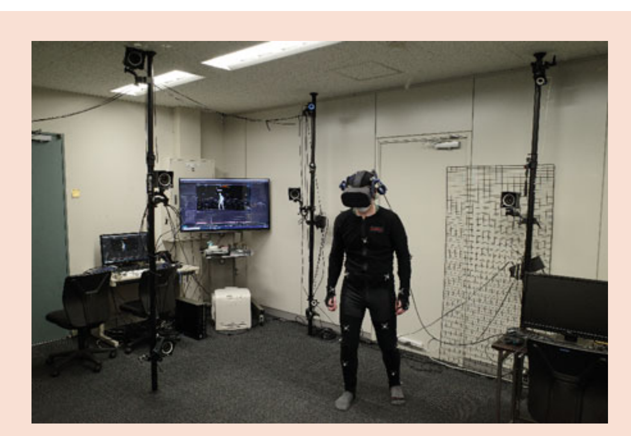


Fig. 9.18 An example of a camera-based motion capture space. IR cameras are mounted on poles. Human pose can be measured by attaching retroreflective markers to a suit

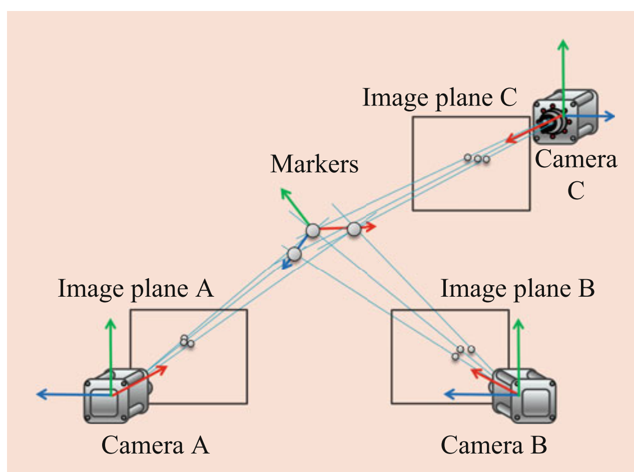


Fig. 9.19 The concept of a multicamera motion capture system. Intrinsic and extrinsic data from two or more cameras enable the system to detect the position of the markers

Fig. 9.20 A snapshot of an active marker set to track an HMD device. LEDs (red circles) and an electronic circuit to control light modulation are embedded in the 3D printed cover

measurement field, but this failure can be overcome through the use of active markers. Instead of illumination at the camera location, LEDs associated with the tracking targets take a role as active marker patterns (Fig. 9.20). For illumination synchronized with cameras to be provided for capturing motion, a circuit is required to control the LEDs. With the controller circuit, LEDs are able to produce higher intensity because of an intermittent illumination pattern. Furthermore, to increase the accuracy of the identification, the control unit is able to change its temporal illumination pattern to transmit an ID for each LED.

9.9.7 Markerless Tracking

Although marker-based tracking is an easy and robust method of tracking the camera in the environment, it requires users to actively prepare and manipulate the environment to ensure that the device can be tracked. Furthermore, when the camera is located very far away from the marker or looks at the marker at a very shallow angle, tracking accuracy and reliability drop significantly. Markerless tracking extracts natural features in the environment and tracks their position in consecutive frames. The 3D position of the features can then be recovered by either performing a predefined motion (i.e., shifting a device a known distance sideways), estimating the device's pose using other sensors (e.g., accelerometers or a gyroscope on a mobile phone), recovering the corresponding depth from a depth sensor (IR or ToF), or triangulating with an additional camera. After the initial position estimation, a variation of SLAM is utilized to continuously track the camera pose. For example, the human body can be tracked by markerless tracking, and facial and hand tracking can be useful techniques in various applications.

As an extension of ordinary point-based natural features, it is also possible to use other natural features such as line segments [29]. Such natural features can be observed in various artificial structures in our daily environment, as we are able to find line segments in ordinary doors, tables, and walls.

One remarkable sensor integration of markerless tracking with AR systems can be found in the Microsoft HoloLens. The tracking system of the original version of HoloLens relied on a multicamera-based SLAM technology with IMUs [26].

Human Pose Tracking

One of the advancements in markerless tracking in recent years is the ability to track natural features such as the human body, face, and fingers. Human body pose tracking provides essential functionality for applications such as interactive games that use AR and move a virtual avatar according to the movements of a real human, one example being the Virtual YouTuber. Here, we summarize the recent trend of markerless human body pose-tracking methods.

Recent developments in deep learning have made it possible to perform human pose tracking even from monocular RGB camera images. Here is a summary of methods using monocular RGB images as inputs for the convolutional neural network in terms of different targets.

Depending on computational resources, it is possible to estimate the 2D human pose for images of a large number of people in almost real time using either top-down approaches or bottom-up approaches. The top-down approaches first detect the respective person in the image, and then a keypoint (shoulders, elbows, wrists, necks, hips, etc.) estimation method for a single person is applied to each person. Bottom-up approaches group the keypoints for each person after the keypoints for all persons have been detected. The bottom-up approaches have the advantage over the top-down that the computation time does not increase greatly, even when the number of persons in the image increases.

OpenPose [12] is a remarkable 2D pose estimation method that uses RGB images as inputs. It was first presented in 2017 and was still being improved as recently as 2020. Also, the well-organized code is available (https://github.com/CMU-Perceptual-Computing-Lab/openpose). It estimates keypoint positions with high accuracy by introducing part affinity fields, an expression of how human keypoints are connected. OpenPose can detect not only body keypoints but also face and hand keypoints at the same time. A sample result of OpenPose is shown in Fig. 9.21b. The face and hands are particularly important parts of the human body, and many estimation methods for the face and hands have been proposed separately from body pose estimation.

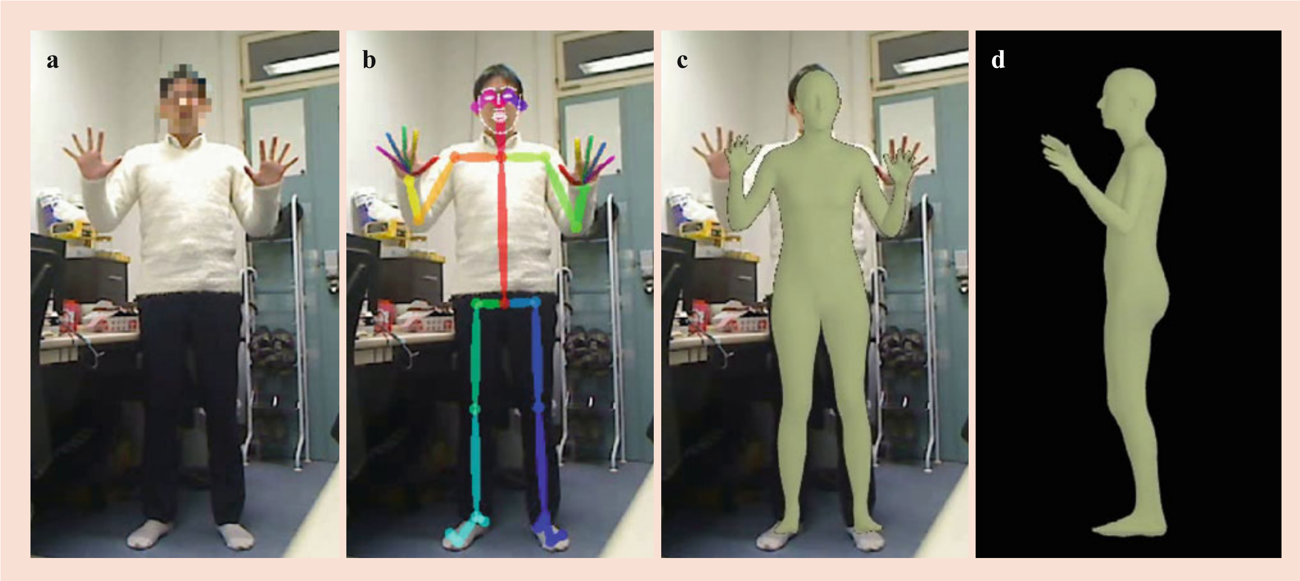


Fig. 9.21 Examples of human pose (**a**) tracking. OpenPose [12] is capable of pose tracking with high accuracy (**b**) (**c**), it can be used to render a side view by reconstructing the body shape (**d**)

On the other hand, AlphaPose, a top-down pose estimation library that can track (give the same person the same ID between video frames), has been published (https://github.com/MVIG-SJTU/AlphaPose). Even though AlphaPose adopts the top-down approach, accuracy does not decrease greatly, even if the number of people in the image increases, and it can run at high speed [58].

Methods of estimating a 3D human pose from only RGB images have also been proposed, most of them targeting a single person. A method of performing 3D pose estimation of multiple people by estimating the position of each person in absolute coordinates has also been proposed [73].

When estimating a 3D pose from an RGB image, there is an ambiguity because 2D and 3D poses do not always correspond exactly. As the movement of a person can be observed continuously in the video, it is possible to eliminate the ambiguity by utilizing temporal consistency [82]. This method can estimate 3D pose with high accuracy by using temporal information, and the estimated pose has temporal coherence.

A method of performing shape reconstruction for a video has also been proposed [50]. This method has realized high-precision shape reconstruction in real time for videos of multiple people. A sample result of the method is shown in Fig. 9.21c, d. Also, there are several remarkable approaches to recognizing multiple elements (body, face, hand, etc.) [121].

In recent years, it has become common to organize code for ease of use and publish it to increase the impact of the research that produced it. Most of the studies listed in this section publish their code. Readers who are interested in it are able to try to use it.

There are several survey papers dedicated to human pose tracking [15], and readers who would like more detail can read them. Also, we can consider the difference between marker-based and markerless tracking [13] when we select a human body tracking method.

The first-person camera is also an RGB camera, but it differs in that it is fixed to the human first-person view. Of course, the body is largely invisible in a first-person view, but there has been an attempt to estimate human body pose from a first-person video [40]. The ubiquity of such action cameras as the GoPro is in the background of this research. As the body is hardly visible, the pose is estimated from the fluctuation pattern of the image and the change in the surrounding landscape.

To capture geometric information, we can use depth sensors as described in the previous chapter. Depth cameras are advantageous for human pose tracking as geometrical features can be captured as depth information. Depth sensor information can be represented as a point cloud, and matching algorithms can be applied to track geometrical information in various pose-tracking systems and SLAM-based systems.

Furthermore, it is possible to estimate the 3D positions of human keypoints by a depth camera. Kinect includes a function for keypoint position estimation in its official library. However, the ToF depth cameras are unstable outside (in sunlight). There is a possibility that this shortcoming of ToF will be overcome in the near future [1].

These human pose-tracking methods can be integrated with AR systems. 2D pose-tracking systems do not require detailed camera parameters for detecting human pose. It is possible to overlay 2D graphics according to human pose with 2D tracking results. However, the ability to overlay a 3D CG model onto a camera image is limited because it is hard to estimate the camera parameters of the scene from 2D tracking results. The capability of 2D-based tracking systems for AR is thus limited. When we utilize a 3D pose-tracking system in AR systems, we need to have camera parameters before we render the CG models.

As examples of AR systems with human pose tracking, one entertainment system provides visual augmentation for physical sports [7]; also, human pose tracking can be applied to education, as is done by an application that provides anatomical knowledge of the human skeleton [98].

Facial Tracking

Tracking facial geometry by camera devices is an emerging field that can be used to change the appearance of a person through AR applications. Depth information is a useful feature for tracking facial information. There are facial tracking systems with structured light [117] and depth cameras [118]. Apple provides a functionality to track facial pose and the geometry of users with the TrueDepth cameras that are integrated with their smartphones and tablet PCs as part of ARKit.

Also, monocular cameras can be applied for facial tracking (Fig. 9.22) with machine learning-based landmark detection methods [46] [127]. It is a common technique for tracking parts of faces. Predicting probability maps of landmark points [69] is another recent trend. A deep learning-based incremental learning approach can be found as a part of recent trials in the facial tracking field [120].

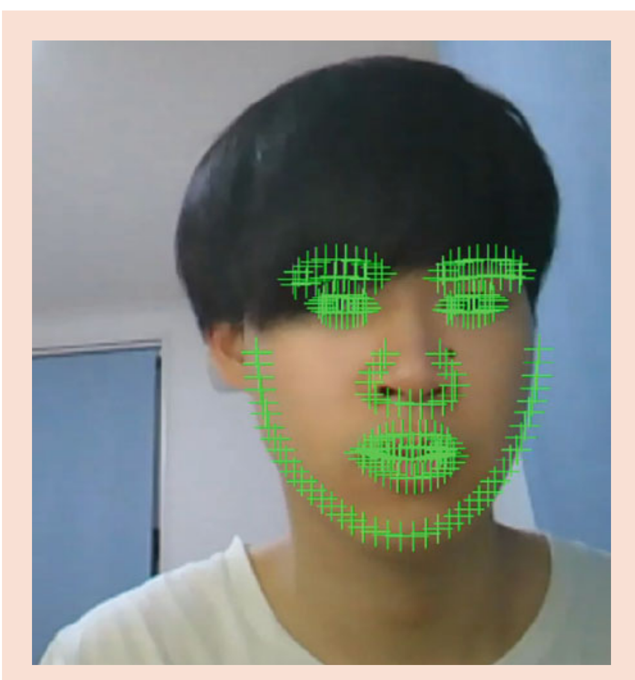


Fig. 9.22 Result of facial landmark tracking using the machine learning-based approach of Kazemi and Sullivan [46]

Hand Tracking

Hand parts provide very important cues for rendering CG models that relate to people's physical activities. In a previous study, a color-based approach tracked a hand with markers in real time remarkably well [116].

With the advancement in techniques for tracking natural features, camera-based markerless motion capture systems are available for hand tracking. A probability-based approach has introduced the possibility of tracking hand pose with camera images [99]. Also, machine learning-based approaches are used for tracking hands for pixel-level detection [54] and pose tracking [128] with monocular camera devices. Depth camera devices can be used to predict hand pose [85, 96].

As a consumer device, leap motion is a well-known hand-tracking device associated with IR imaging modules. It is used for various hand-tracking [67] scenarios.

These hand-tracking results can be used to interact with virtual objects. By detecting users' 3D hand pose, it is possible to make virtual physical interaction (e.g., grabbing, pinching, and pushing) possible in AR environments.

Thermal Tracking

IR image sensors are able to capture invisible rays; they can be used in AR to extend the capability to track various features. An infrared camera is an imaging device with sensitivity to wavelengths longer than red. Infrared can be categorized as near infrared or far infrared. Near infrared is used for tracking markers in camera-based motion capture systems, and far infrared cameras can capture thermal information. By calibrating a thermographic camera with an RGB camera, it is possible to visualize heating information on natural visual features as an extension of human perception [68]. Also, thermal tracking is useful for sensing physical interaction between users and surrounding space owing to thermal conductance. Thermal Touch [52] introduced the possibility of detecting touch interaction in an AR environment.

9.9.8 Projection-Based Sensing

Projection-based sensing is a projection system with simple photo sensors on it (Fig. 9.23) that can be used instead of camera-based tracking systems. For example, if we emit a light for tracking space at a specific time and we receive the light with a simple light sensor, we can find the location of the sensor. The major advantage of projection-based tracking is a lower computation cost than general camera-based tracking systems.

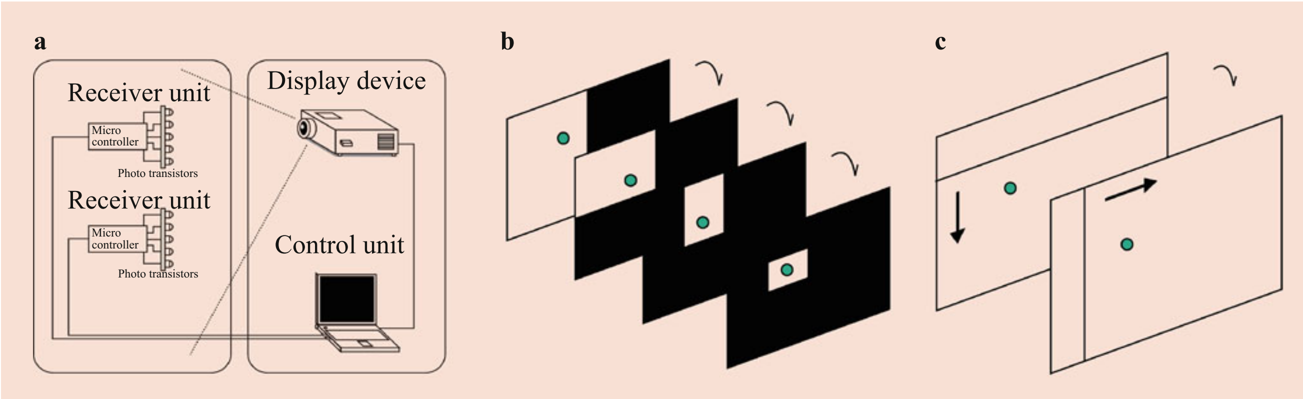


Fig. 9.23 (**a**) A system configuration to track photo sensors by using a projection device [104]. (**b**) An example of a spatial division coding pattern to detect a photo sensor by projecting a binary code. (**c**) An example of spatial scanning pattern to detect the sensor position

In this section, several projection-based tracking systems are introduced. By using time and space modulation, it is possible to track multiple sensors with the projection-based tracking approach. One remarkable projection system for position sensing is radio frequency identification technology (RFID) [87]. Display-based computing systems [104] are another attempt to measure 2D position and rotation simultaneously by using projection patterns from a display device. Prakash [88] is an interesting tracking system that digitizes the sensor position with a binary projection pattern. Visible light communication uses embedded light projections of an ordinary visible light pattern such as a room light or projector for computer displays [35].

Spatial Division Code

Spatial division detects sensors in the measurement field by projecting coded patterns. Put simply, if we project a black and white pattern, we can check the signal received by a photo sensor and detect the absolute position of that sensor. Advanced coding techniques, such as Gray Code [27], a reflected binary code, were applied [87] to improve the accuracy of measurement. Spatial division code reduces the average number of decoding errors by minimizing the code distance between spatial regions next to each other.

Spatial Scanning

One of the simplest ideas for measuring the position of a photo sensor is to scan the measurement space using a line pattern projected at a constant speed. Such a scanning-based approach can be used as a tracking system for a handheld display device [105]. Given the starting timing t_s,the duration d of the scan, and the peak timing t_p of the signal received by the photo sensor, the normalized position of the sensor p can be calculated as follows:

$$p = \frac{t_p - t_s}{d} \tag{9.50}$$

Fig. 9.24 A base station for Valve's Lighthouse tracking. There are two laser scanning modules for orthogonal axes inside the base station

A two-axis scanning system makes it possible to measure the 2D position of the tracking target. The 3D position of the tracking target can be obtained by having scanning modules in two locations at a baseline distance. If we can track the 3D position of at least three sensors on a tracking target, we can predict the target within 6 DoF.

Valve's Lighthouse technology [126] is an actual tracking device with such an idea. A Lighthouse base station (Fig. 9.24) is equipped with two line laser modules actuated by mechanical motors that are able to scan the measurement space in orthogonal two axes. For the initialization process, the external parameters of the base stations can be estimated by having at least three sensors with a known 3D layout. After obtaining the external parameters of two base stations in a tracking space, we can track the 3D position of photo sensors in the tracking space with a very small computation cost. The photo sensors can be embedded with head-mounted displays, and hand controllers, and so on.

9.10 Applications

Tracking systems can be used for various AR content if proper calibration methods exist. By utilizing the results from the tracking, we are able to make the relationship between physical and cyberspaces spatially consistent.

9.10.1 Medical Applications

Medicine is one of the practical fields in which AR navigation and visualization with tracking systems can be used. In ordinary situations, doctors need to see surgical targets inside of patients' bodies. However, human vision can only see light in a limited range of wavelengths. To see through a patient's body, medical imaging systems, such as ultrasound, X-rays, and magnetic resonance, are needed [75]. AR systems allow visualization of captured images and models reconstructed by medical imaging systems.

To track the position and orientation of the patient's body, the reconstructed models and actual surgical targets need to be spatially consistent. In minimally invasive surgeries using medical AR, doctors insert instruments (e.g., endoscopes and forceps) into the body, where there are optical occlusions. Tracking systems are often used to track the extrinsic parameters of endoscopes. To track a rigid instrument, it is possible to attach markers around the grip so they can be tracked by outside-in tracking systems [53]. For nonrigid instruments, there should be a sensor around the effective point (e.g., tool tip or a configuration). Magnetic sensors and SLAM-based tracking systems can be used in such situations. Figure 9.25 shows an endoscope with retroreflective markers. Hand–eye calibration can be applied to the system.

AR applications are able to aid in the understanding of various complicated structures related to the medical field. Visualizing bone structures for educational purposes [98] is one possible application for human body tracking.

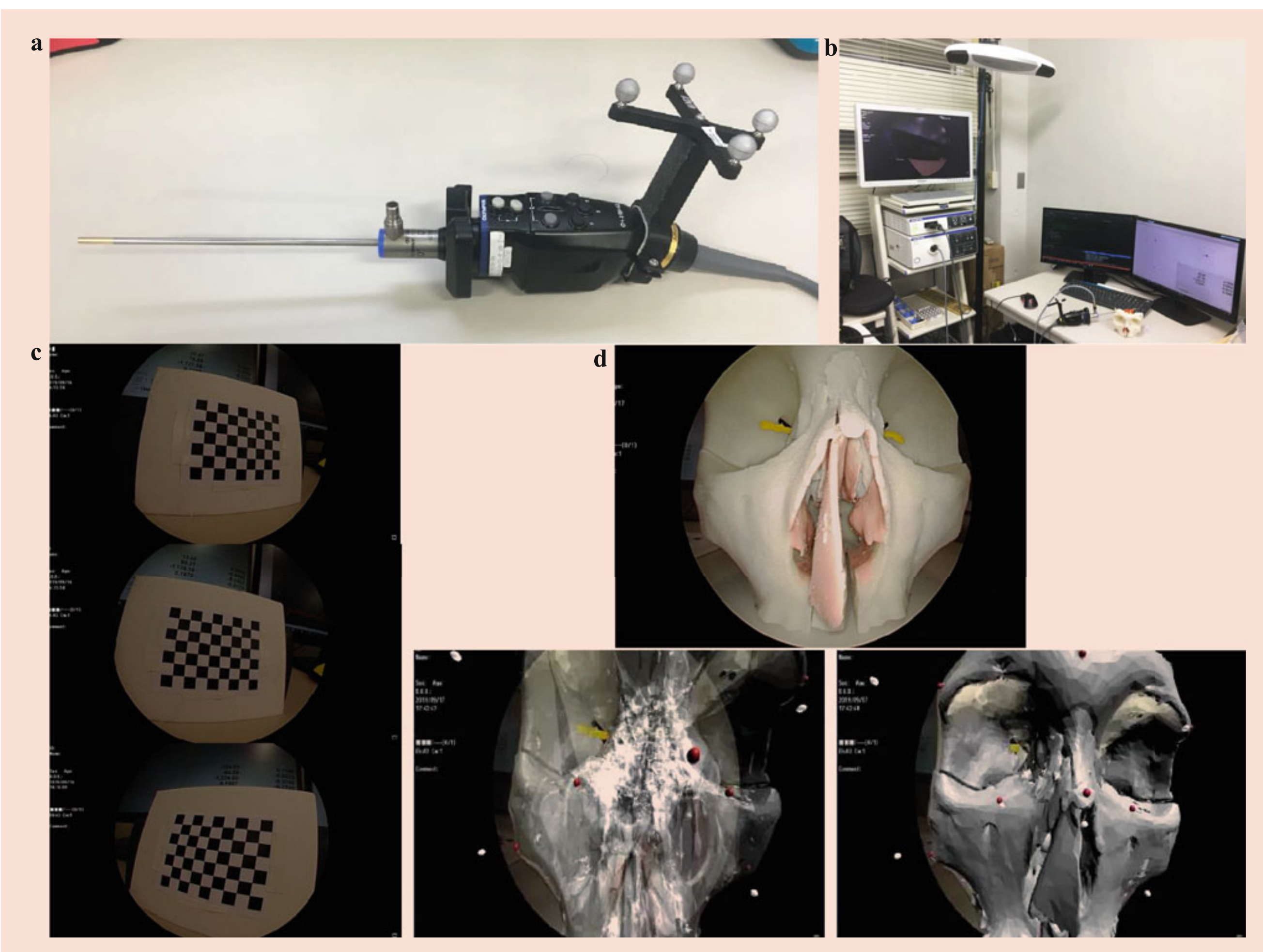


Fig. 9.25 An example of endoscopic AR. The camera inside the endoscope (**a**) is calibrated against the external tracking system (**b**) by capturing images of a checkerboard (**c**). The skeleton (**d**) can then be overlaid with correctly registered CG

9.10.2 Robotic Applications

Robotics is another practical field for the use of sensor integration. Ordinary industrial robots have mechanical links with motors and rotary encoders. By using the tracking results from the mechanical link structures, we are able to draw CG models on robots in real time. Figure 9.2 shows an example of an application using mechanical sensing results with an optical see-through display to overlay a model of the background onto the robot, creating the illusion of transparency [109].

The position and orientation of a mobile robot can be tracked by tracking systems. Also, internal sensors such as odometers, IMU sensors, and SLAM-based inside-out tracking can be used with first-person view camera images on a robot to make video see-through AR applications. Outside-in tracking systems provide better results for the drawing of CG models in absolute-coordinate systems.

AR can be used to generate a virtual third-person view of a mobile robot [103]. It can also visualize the results from sensors and plan the course of a humanoid robot [100]. Furthermore, AR applications [34, 63] allow operators to make predictions on the basis of visual representations of the targets of their operations in context.

Fig. 9.26 An example of tracking applications: AR shooting game with tiny robots tracked by a display-based tracking system (Reprinted with permission from [51])

9.10.3 Entertainment Applications

Various AR environments with tracking systems are also used in entertainment. The results of human body and facial tracking can be used to augment people's appearances [62]. Visual augmentation can also be applied to physical objects such as toy vehicles. Figure 9.26 shows a projection-based AR system with gradient fiducials for tracking mobile targets [51]. Dynamic virtual functions can be provided by applying visual augmentation to the tracking targets.

9.11 Conclusion

In this chapter, we introduced tracking systems with calibration methods, sensing modalities, and sensor integration that enable the development of robust AR systems. As described in the introduction, tracking technology is a core technique for making AR systems that consider the spatial relationship of physical and virtual spaces. Traditional sensing modality provides reliable tracking results, and machine learning techniques allow us to consider advanced sensing approaches. By selecting and integrating efficient sensing modalities and calibration methods in accordance with the context of each application, we are able to develop feasible AR systems.

References

1. Achar, S., Bartels, J.R., Whittaker, W.L., Kutulakos, K.N., Narasimhan, S.G.: Epipolar time-of-flight imaging. ACM Trans. Graph. **36**(4), 1–8 (2017)
2. Ahn, S.J., Rauh, W., Warnecke, H.-J.: Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. Pattern Recogn. **34**(12), 2283–2303 (2001)
3. Akinyelu, A.A., Blignaut, P.: Convolutional neural network-based methods for eye gaze estimation: a survey. IEEE Access **8**, 142581–142605 (2020)
4. Akiyama, R., Yamamoto, G., Amano, T., Taketomi, T., Plopski, A., Fujimoto, Y., Kanbara, M., Sandor, C., Kato, H.: Illusory light: perceptual appearance control using a projection-induced illusion. Comput. Graph. **91**, 129–140 (2020)
5. Akiyama, R., Yamamoto, G., Amano, T., Taketomi, T., Plopski, A., Sandor, C., Kato, H.: Robust reflectance estimation for projection-based appearance control in a dynamic light environment. IEEE Trans. Vis. Comput. Graph. **27**(3), 2041–2055 (2021)
6. Alves, J., Lobo, J., Dias, J.: Camera-inertial sensor modelling and alignment for visual navigation. Mach. Intell. Robot. Control **5**(3), 103–112 (2003)
7. Araki, H., Fukuda, H., Motoki, T., Takeuchi, T., Ohta, N., Adachi, R., Masuda, H., Kado, Y., Mita, Y., Mizukami, D., Kakeya, N.: "HADO" as techno sports was born by the fusion of it technology and sports. In: Shirai, A., Chretien, L., Clayer, A.-S., Richir, S., Hasegawa, S. (eds.) ReVo 2017: Laval Virtual ReVolution 2017 "Transhumanism++", vol. 1. EPiC Series in Engineering, pp. 36–40 (2018)
8. Audet, S., Okutomi, M.: A user-friendly method to geometrically calibrate projector-camera systems. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern

Recognition Workshops, pp. 47–54. IEEE Computer Society, Washington (2009)
9. Besl, P.J., McKay, N.D.: Method for registration of 3-D shapes. In: Sensor Fusion IV: Control Paradigms and Data Structures, vol. 1611, pp. 586–606. International Society for Optics and Photonics, Bellingham (1992)
10. Bimber, O., Raskar, R.: Spatial augmented reality: Merging real and virtual worlds. A K Peters, Natick/CRC Press, Boca Raton (2005)
11. Birkfellner, W., Watzinger, F., Wanschitz, F., Ewers, R., Bergmann, H.: Calibration of tracking systems in a surgical environment. IEEE Trans. Med. Imaging **17**(5), 737–742 (1998)
12. Cao, Z., Martinez, G.H., Simon, T., Wei, S., Sheikh, Y.A.: Openpose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. **43**(1), 172–186 (2019)
13. Ceseracciu, E., Sawacha, Z., Cobelli, C.: Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: proof of concept. PLoS One **9**(3), e87640 (2014)
14. Chen, Y., Medioni, G.: Object modelling by registration of multiple range images. Image Vis. Comput. **10**(3), 145–155 (1992)
15. Chen, Y., Tian, Y., He, M.: Monocular human pose estimation: a survey of deep learning-based methods. Comput. Vis. Image Underst. **192**, 102897 (2020)
16. Denavit, R.S., Hartenberg, J.: A kinematic notation for lower-pair mechanisms based on matrices. Trans. ASME J. Appl. Mech. **22**, 215–221 (1955)
17. Dierkes, K., Kassner, M., Bulling, A.: A novel approach to single camera, glint-free 3D eye model fitting including corneal refraction. In: Proceedings of the ACM Symposium on Eye Tracking Research & Applications, pp. 1–9. Association for Computing Machinery, New York (2018)
18. Dorai, C., Wang, G., Jain, A.K., Mercer, C.: Registration and integration of multiple object views for 3D model construction. IEEE Trans. Pattern Anal. Mach. Intell. **20**(1), 83–89 (1998)
19. Eck, U., Pankratz, F., Sandor, C., Klinker, G., Laga, H.: Precise haptic device co-location for visuo-haptic augmented reality. IEEE Trans. Vis. Comput. Graph. **21**(12), 1427–1441 (2015)
20. Foley, J.D.: Interfaces for advanced computing. Sci. Am. **257**(4), 126–135 (1987)
21. Furgale, P., Rehder, J., Siegwart, R.: Unified temporal and spatial calibration for multi-sensor systems. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1280–1286. IEEE Computer Society, Washington (2013)
22. Gelfand, N., Ikemoto, L., Rusinkiewicz, S., Levoy, M.: Geometrically stable sampling for the ICP algorithm. In: Proceedings of the International Conference on 3-D Digital Imaging and Modeling, pp. 260–267. IEEE Computer Society, Washington (2003)
23. Genc, Y., Sauer, F., Wenzel, F., Tuceryan, M., Navab, N.: Optical see-through hmd calibration: a stereo method validated with a video see-through system. In: Proceedings IEEE and ACM International Symposium on Augmented Reality, pp. 165–174. IEEE Computer Society, Washington (2000)
24. Genc, Y., Tuceryan, M., Navab, N.: Practical solutions for calibration of optical see-through devices. In: Proceedings of the International Symposium on Mixed and Augmented Reality, pp. 169–175. IEEE Computer Society, Washington (2002)
25. Geng, J.: Structured-light 3D surface imaging: a tutorial. Adv. Opt. Photon. **3**, 128–160 (2011)
26. Gourlay, M., Olszta, P., Mathe, Z., Sinha, S., Pollefeys, M., Klein, G.: Computer vision on Microsoft Hololens. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
27. Gray, F.: Pulse code communication. U.S. patent (1947)
28. Grewal, M.S., Andrews, A.P.: Kalman filtering: Theory and practice with MATLAB. John Wiley & Sons, Hoboken (2014)
29. Grompone, R., Gioi, V., Jakubowicz, J., Morel, J.-M., Randall, G.: LSD: a fast line segment detector with a false detection control. IEEE Trans. Pattern Anal. Mach. Intell. **32**(4), 722–732 (2008)
30. Grubert, J., Itoh, Y., Moser, K., Swan, J.E.: A survey of calibration methods for optical see-through head-mounted displays. IEEE Trans. Vis. Comput. Graph. **24**(9), 2649–2662 (2017)
31. Grundhöfer, A., Iwai, D.: Recent advances in projection mapping algorithms, hardware and applications. Comput. Graphics Forum **37**(2), 653–675 (2018)
32. Guestrin, E.D., Eizenman, M.: General theory of remote gaze estimation using the pupil center and corneal reflections. IEEE Trans. Biomed. Eng. **53**(6), 1124–1133 (2006)
33. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press, Cambridge (2003)
34. Hashimoto, S., Ishida, A., Inami, M., Igarashi, T.: Touchme: An augmented reality based remote robot manipulation. In: Kiyokawa, K., Kuhlen, T., Reiners, D. (eds.) Proceedings of the International Conference on Artificial Reality and Telexistence, pp. 61–66 (2011)
35. Hiraki, T., Fukushima, S., Watase, H., Naemura, T.: [Paper] Dynamic PVLC: pixel-level visible light communication projector with interactive update of images and data. ITE Trans. Media Tech. Appl. **7**(4), 160–168 (2019)
36. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. **4**(4), 629–642 (1987)
37. Itoh, Y., Klinker, G.: Interaction-free calibration for optical see-through head-mounted displays based on 3D eye localization. In: Proceedings of the IEEE Symposium on 3D User Interfaces, pp. 75–82. IEEE Computer Society, Washington (2014)
38. Itoh, Y., Klinker, G.: Performance and sensitivity analysis of INDICA: interaction-free display calibration for optical see-through head-mounted displays. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 171–176. IEEE Computer Society, Washington (2014)
39. Itoh, Y., Klinker, G.: Light-field correction for spatial calibration of optical see-through head-mounted displays. IEEE Trans. Vis. Comput. Graph. **21**(4), 471–480 (2015)
40. Jiang, H., Grauman, K.: Seeing invisible poses: estimating 3D body pose from egocentric video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3501–3509. IEEE Computer Society, Washington (2017)
41. Jiang, W., Xue, H., Miao, C., Wang, S., Lin, S., Tian, C., Murali, S., Hu, H., Sun, Z., Su, L.: Towards 3D human pose construction using wifi. In: Proceedings of the Annual International Conference on Mobile Computing and Networking, pp. 1–14 (2020)
42. Jun, H., Kim, G.: A calibration method for optical see-through head-mounted displays with a depth camera. In: Proceedings of the IEEE Virtual Reality, pp. 103–111. IEEE Computer Society, Washington (2016)
43. Kabsch, W.: A solution for the best rotation to relate two sets of vectors. Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **32**(5), 922–923 (1976)
44. Kalman, R.E., et al.: A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960)
45. Kato, H., Billinghurst, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: Proceedings of the IEEE and ACM International Workshop on Augmented Reality, pp. 85–94. IEEE Computer Society, Washington (1999)
46. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874. IEEE Computer Society, Washington (2014)
47. Kelly, J., Sukhatme, G.S.: Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. Int. J. Robot. Res. **30**(1), 56–79 (2011)

48. Kijima, R., Ojika, T.: Reflex HMD to compensate lag and correction of derivative deformation. In: Proceedings of the IEEE Virtual Reality, pp. 172–179. IEEE Computer Society, Washington (2002)
49. Klemm, M., Seebacher, F., Hoppe, H.: High accuracy pixel-wise spatial calibration of optical see-through glasses. Comput. Graph. **64**, 51–61 (2017)
50. Kocabas, M., Athanasiou, N., Black, M.J.: Vibe: Video inference for human body pose and shape estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5253–5263. IEEE Computer Society, Washington (2020)
51. Kojima, M., Sugimoto, M., Nakamura, A., Tomita, M., Inami, M., Nii, H.: Augmented coliseum: An augmented game environment with small vehicles. In: Proceedings of the IEEE International Workshop on Horizontal Interactive Human-Computer Systems, pp. 3–8. IEEE Computer Society, Washington (2006)
52. Kurz, D.: Thermal touch: thermography-enabled everywhere touch interfaces for mobile augmented reality applications. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 9–16. IEEE Computer Society, Washington (2014)
53. Lai, M., Skyrman, S., Shan, C., Babic, D., Homan, R., Edström, E., Persson, O., Burström, G., Elmi-Terander, A., Hendriks, B.H.W., de With, P.H.N.: Fusion of augmented reality imaging with the endoscopic view for endonasal skull base surgery; a novel application for surgical navigation based on intraoperative cone beam computed tomography and optical tracking. PLoS One **15**(1),1–13 (2020)
54. Li, C., Kitani, K.M.: Pixel-level hand detection in ego-centric videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3570–3577. IEEE Computer Society, Washington (2013)
55. Li, M., Mourikis, A.I.: Online temporal calibration for camera–imu systems: theory and algorithms. Int. J. Robot. Res. **33**(7), 947–964 (2014)
56. Li, M., Yu, H., Zheng, X., Mourikis, A.I.: High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 409–416. IEEE Computer Society, Washington (2014)
57. Li, H., Trutoiu, L., Olszewski, K., Wei, L., Trutna, T., Hsieh, P.-L., Nicholls, A., Ma, C.: Facial performance sensing head-mounted display. ACM Trans. Graph. **34**(4) (2015)
58. Li, J., Wang, C., Zhu, H., Mao, Y., Fang, H.-S., Lu, C.: Crowdpose: efficient crowded scenes pose estimation and a new benchmark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 10863–10872. IEEE Computer Society, Washington (2019)
59. Lin, B., Wang, F., Zhao, F., Sun, Y.: Scale invariant point feature (SIPF) for 3D point clouds and 3D multi-scale object detection. Neural Comput. Applic. **29**(5), 1209–1224 (2018)
60. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. ACM Trans. Graph. **34**(6), 1–16 (2015)
61. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 24–32 (2004)
62. Lugrin, J.-L., Latt, J., Latoschik, M.E.: Anthropomorphism and illusion of virtual body ownership. In: Proceedings of the International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, pp. 1–8. The Eurographics Association, Geneve (2015)
63. Maeda, N., Morita, J., Sugimoto, M.: Pathfinder vision: teleoperation robot interface in consideration of geometry for supporting future prediction. In: Proceedings of the International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, pp. 29–35. The Eurographics Association, Geneve (2014)
64. Magee, D., Zhu, Y., Ratnalingam, R., Gardner, P., Kessel, D.: An augmented reality simulator for ultrasound guided needle placement training. Med. Biol. Eng. Comput. **45**(10), 957–967 (2007)
65. Mair, E., Fleps, M., Suppa, M., Burschka, D.: Spatio-temporal initialization for imu to camera registration. In: Proceedings of the IEEE International Conference on Robotics and Biomimetics, pp. 557–564. IEEE Computer Society, Washington (2011)
66. Makibuchi, N., Kato, H., Yoneyama, A.: Vision-based robust calibration for optical see-through head-mounted displays. In: Proceedings of the IEEE International Conference on Image Processing, pp. 2177–2181. IEEE Computer Society, Washington (2013)
67. Marin, G., Dominio, F., Zanuttigh, P.: Hand gesture recognition with leap motion and kinect devices. In: Proceedings of the IEEE International Conference on Image Processing, pp. 1565–1569. IEEE Computer Society, Washington (2014)
68. Matsumoto, K., Nakagawa, W., Saito, H., Sugimoto, M., Shibata, T., Yachida, S.: AR visualization of thermal 3D model by hand-held cameras, vol. 3, pp. 480–487. SciTePress, Setúbal (2015)
69. Merget, D., Rock, M., Rigoll, G.: Robust facial landmark detection via a fully-convolutional local-global context network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 781–790. IEEE Computer Society, Washington (2018)
70. Meskers, C.G.M., Fraterman, H., van der Helm, F.C.T., Vermeulen, H.M., Rozing, P.M.: Calibration of the "flock of birds" electromagnetic tracking device and its application in shoulder motion studies. J. Biomech. **32**(6), 629–633 (1999)
71. Meyer J., Schlebusch, T., Fuhl, W., Kasneci, E.: A novel camera-free eye tracking sensor for augmented reality based on laser scanning. IEEE Sensors J. **20**(24), 15204–15212 (2020)
72. Mirzaei, F.M., Roumeliotis, S.I.: A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. IEEE Trans. Robot. **24**(5), 1143–1156 (2008)
73. Moon, G., Chang, J.Y., Lee, K.M.: Camera distance-aware top-down approach for 3D multi-person pose estimation from a single rgb image. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10133–10142. IEEE Computer Society, Washington (2019)
74. Moser, K., Itoh, Y., Oshima, K., Swan, J.E., Klinker, G., Sandor, C.: Subjective evaluation of a semi-automatic optical see-through head-mounted display calibration technique. IEEE Trans. Vis. Comput. Graph. **21**(4), 491–500 (2015)
75. Navab, N., Heining, S., Traub, J.: Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications. IEEE Trans. Med. Imaging **29**(7), 1412–1423 (2010)
76. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136. IEEE Computer Society, Washington (2011)
77. Newman, J., Wagner, M., Bauer, M., MacWilliams, A., Pintaric, T., Beyer, D., Pustka, D., Strasser, F., Schmalstieg, D., Klinker, G.: Ubiquitous tracking for augmented reality. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 192–201. IEEE Computer Society, Washington (2004)
78. Nikolic, J., Burri, M., Gilitschenski, I., Nieto, J., Siegwart, R.: Non-parametric extrinsic and intrinsic calibration of visual-inertial sensor systems. IEEE Sensors J. **16**(13), 5433–5443 (2016)
79. Osman, A.A.A., Bolkart, T., Black, M.J.: STAR: a spare trained articulated human body regressor. In: Proceedings of the European Conference on Computer Vision, pp. 598–613. Springer International Publishing, New York City (2020)

80. Owen, C.B., Zhou, J., Tang, A., Xiao, F.: Display-relative calibration for optical see-through head-mounted displays. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 70–78. IEEE Computer Society, Washington (2004)
81. Pankratz, F., Klinker, G.: [Poster] AR4AR: using augmented reality for guidance in augmented reality systems setup. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 140–143. IEEE Computer Society, Washington (2015)
82. Pavllo, D., Feichtenhofer, C., Grangier, D., Auli, M.: 3D human pose estimation in video with temporal convolutions and semi-supervised training. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7753–7762. IEEE Computer Society, Washington (2019)
83. Plopski, A., Itoh, Y., Nitschke, C., Kiyokawa, K., Klinker, G., Takemura, H.: Corneal-imaging calibration for optical see-through head-mounted displays. IEEE Trans. Vis. Comput. Graph. **21**(4), 481–490 (2015)
84. Plopski, A., Orlosky, J., Itoh, Y., Nitschke, C., Kiyokawa, K., Klinker, G.: Automated spatial calibration of HMD systems with unconstrained eye-cameras. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 94–99. IEEE Computer Society, Washington (2016)
85. Qian, C., Sun, X., Wei, Y., Tang, X., Sun, J.: Realtime and robust hand tracking from depth. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1106–1113. IEEE Computer Society, Washington (2014)
86. Raab, F.H., Blood, E.B., Steiner, T.O., Jones, H.R.: Magnetic position and orientation tracking system. IEEE Trans. Aerosp. Electron. Syst. **AES-15**(5), 709–718 (1979)
87. Raskar, R., Beardsley, P., van Baar, J., Wang, Y., Dietz, P., Lee, J., Leigh, D., Willwacher, T.: RFIG lamps: Interacting with a self-describing world via photosensing wireless tags and projectors. ACM Trans. Graph. **23**(3), 406–415 (2004)
88. Raskar, R., Nii, H., de Decker, B., Hashimoto, Y., Summet, J., Moore, D., Zhao, Y., Westhues, J., Dietz, P.H., Barnwell, J., Nayar, S.K., Inami, M., Bekaert, P., Noland, M., Branzoi, V., Bruns, E.: Prakash: Lighting aware motion capture using photosensing markers and multiplexed illuminators. ACM Trans. Graph. **26**(3), 36 (2007)
89. Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., Siegwart, R.: Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 4304–4311. IEEE Computer Society, Washington (2016)
90. Rehder, J., Siegwart, R.: Camera/IMU calibration revisited. IEEE Sensors J. **17**(11), 3257–3268 (2017)
91. Rekimoto, J., Miyaki, T., Ishizawa, T.: Lifetag: wifi-based continuous location logging for life pattern analysis. In: Hightower, J., Schiele, B., Strang, T. (eds.) Location- and Context-Awareness, pp. 35–49. Berlin, Heidelberg (2007)
92. Rémy, S., Dhome, M., Lavest, J.-M., Daucher, N.: Hand-eye calibration. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications, vol. 2, pp. 1057–1065. IEEE Computer Society, Washington (1997)
93. Robinson, D.A.: A method of measuring eye movemnent using a scieral search coil in a magnetic field. IEEE Trans. Bio-med. Electron. **10**(4), 137–145 (1963)
94. Rusu, R.B., Marton, Z.C., Blodow, N., Dolha, M., Beetz, M.: Towards 3D point cloud based object maps for household environments. Robot. Auton. Syst. **56**(11), 927–941 (2008)
95. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 3212–3217. IEEE Computer Society, Washington (2009)
96. Sharp, T., Keskin, C., Robertson, D., Taylor, J., Shotton, J., Kim, D., Rhemann, C., Leichter, I., Vinnikov, A., Wei, Y., Freedman, D., Kohli, P., Krupka, E., Fitzgibbon, A., Izadi, S.: Accurate, robust, and flexible real-time hand tracking. In: Proceedings of the Annual ACM Conference on Human Factors in Computing Systems, pp. 3633–3642. Association for Computing Machinery, New York (2015)
97. Sheng, Q.H., Zhang, B.: Absolute orientation based on line coordinates. Photogramm. Rec. **32**(157), 12–32 (2017)
98. Stefan, P., Wucherer, P., Oyamada, Y., Ma, M., Schoch, A., Kanegae, M., Shimizu, N., Kodera, T., Cahier, S., Weigl, M., Sugimoto, M., Fallavollita, P., Saito, H., Navab, N.: An AR edutainment system supporting bone anatomy learning. In: Proceedings of the IEEE Virtual Reality, pp. 113–114. IEEE Computer Society, Washington (2014)
99. Stenger, B., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Model-based hand tracking using a hierarchical bayesian filter. IEEE Trans. Pattern Anal. Mach. Intell. **28**(9), 1372–1384 (2006)
100. Stilman, M., Michel, P., Chestnutt, J., Nishiwaki, K., Kagami, S., Kuffner, J.: Augmented reality for robot development and experimentation. Technical Report CMU-RI-TR-05-55, Carnegie Mellon University, Pittsburgh (2005)
101. Strobl, K.H., Hirzinger, G.: Optimal hand-eye calibration. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4647–4653. IEEE Computer Society, Washington (2006)
102. Sugano, Y., Matsushita, Y., Sato, Y.: Learning-by-synthesis for appearance-based 3D gaze estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1821–1828. IEEE Computer Society, Washington (2014)
103. Sugimoto, M., Kagotani, G., Nii, H., Shiroma, N., Matsuno, F., Inami, M.: Time follower's vision: a teleoperation interface with past images. IEEE Comput. Graph. Appl. **25**(1), 54–63 (2005)
104. Sugimoto, M., Kodama, K., Nakamura, A., Kojima, M., Inami, M.: A display-based tracking system: display-based computing for measurement systems. In: Proceedings of the 17th International Conference on Artificial Reality and Telexistence (ICAT 2007), pp. 31–38. IEEE, Piscataway (2007)
105. Summet, J., Sukthankar, R.: Tracking locations of moving hand-held displays using projected light. In: Hans Gellersen, W., Want, R., Schmidt, A. (eds.) Pervasive computing, pp. 37–46. Berlin, Heidelberg (2005)
106. Sutherland, I.E.: A head-mounted three dimensional display. In: Proceedings of the Fall Joint Computer Conference, Part I, pp. 757–764. Association for Computing Machinery, New York (1968)
107. Swirski, L., Dodgson, N.: A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting. In: Proceedings of the International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction, pp. 1–11 (2013)
108. Taketomi, T., Uchiyama, H., Ikeda, S.: Visual slam algorithms: a survey from 2010 to 2016. IPSJ Trans. Comput. Vis. Appl. **9**(1), 16 (2017)
109. Taylor, A., Matsumoto, A., Carter, E.J., Plopski, A., Admoni, H.: Diminished reality for close quarters robotic telemanipulation. In: Proceedings of the IEEE/RSJ International Converence on Intelligent Robots and Systems, pp. 111531–111538. IEEE Computer Society, Washington (2020)
110. Tedaldi, D., Pretto, A., Menegatti, E.: A robust and easy to implement method for IMU calibration without external equipments. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 3042–3049. IEEE Computer Society, Washington (2014)

111. Tsai, R.Y., Lenz, R.K.: A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. IEEE Trans. Robot. Autom. **5**(3), 345–358 (1989)
112. Tuceryan, M., Navab, N.: Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR. In: Proceedings of the IEEE and ACM International Symposium on Augmented Reality, pp. 149–158. IEEE Computer Society, Washington (2000)
113. Tuceryan, M., Genc, Y., Navab, N.: Single-point active alignment method (SPAAM) for optical see-through HMD calibration for augmented reality. Presence Teleop. Virt. **11**(3), 259–276 (2002)
114. von Marcard, T., Rosenhahn, B., Black, M.J., Pons-Moll, G.: Sparse inertial poser: automatic 3D human pose estimation from sparse IMUs. Comput. Graphics Forum **36**(2), 349–360 (2017)
115. Wang, C.-C.: Extrinsic calibration of a vision sensor mounted on a robot. IEEE Trans. Robot. Autom. **8**(2), 161–175 (1992)
116. Wang, R.Y., Popović, J.: Real-time hand-tracking with a color glove. ACM Trans. Graph. **28**(3) (2009)
117. Weise, T., Li, H., Van Gool, L., Pauly, M.: Face/off: live facial puppetry. In: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 7–16. Association for Computing Machinery, New York (2009)
118. Weise, T., Bouaziz, S., Li, H., Pauly, M.: Realtime performance-based facial animation. ACM Trans. Graph. **30**(4) (2011)
119. Willi, S., Grundhöfer, A.: Robust geometric self-calibration of generic multi-projector camera systems. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 42–51. IEEE Computer Society, Washington (2017)
120. Wu, C., Shiratori, T., Sheikh, Y.: Deep incremental learning for efficient high-fidelity face tracking. ACM Trans. Graph. **37**(6) (2018)
121. Xiang, D., Joo, H., Sheikh, Y.: Monocular total capture: posing face, body, and hands in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10957–10966. IEEE Computer Society, Washington (2019)
122. Yamazaki, S., Mochimaru, M., Kanade, T.: Simultaneous self-calibration of a projector and a camera using structured light. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 60–67. IEEE Computer Society, Washington (2011)
123. Yang, Z., Shen, S.: Monocular visual–inertial state estimation with online initialization and camera–IMU extrinsic calibration. IEEE Trans. Autom. Sci. Eng. **14**(1), 39–51 (2016)
124. Yang, L., Normand, J.-M., Moreau, G.: Practical and precise projector-camera calibration. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 63–70. IEEE Computer Society, Washington (2016)
125. Yaniv, Z.: Which pivot calibration? In: Medical imaging 2015: Image-guided procedures, robotic interventions, and modeling, vol. 9415 (2015)
126. Yates, A., Selan, J.: Positional tracking systems and methods. U.S. patent, US20160131761A1 (2016)
127. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multi-task learning. In: Proceedings of the European Conference on Computer Vision, pp. 94–108. Springer International Publishing, New York City (2014)
128. Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., Grundmann, M.: Mediapipe hands: on-device real-time hand tracking. In: Proceedings of the CVPR Workshop on Computer Vision for Augmented and Virtual Reality, pp. 4321–4325. IEEE Computer Society, Washington (2020)
129. Zhao, M., Li, T., Alsheikh, M.A., Tian, Y., Zhao, H., Torralba, A., Katabi, D.: Through-wall human pose estimation using radio signals. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7356–7365. IEEE Computer Society, Washington (2018)
130. Zheng, Z., Yu, T., Li, H., Guo, K., Dai, Q., Fang, L., Liu, Y.: Hybridfusion: real-time performance capture using a single depth sensor and sparse IMUs. In: Proceedings of the European Conference on Computer Vision, pp. 389–406. Springer International Publishing, New York City (2018)
131. Zhuang, H., Roth, Z.S., Sudhakar, R.: Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form ax=yb. IEEE Trans. Robot. Autom. **10**(4), 549–554 (1994)
132. Zuo, C., Feng, S., Huang, L., Tao, T., Yin, W., Chen, Q.: Phase shifting algorithms for fringe projection profilometry: a review. Opt. Lasers Eng. **109**, 23–59 (2018)

Alexander Plopski is a Postdoctoral Fellow at the Institute of Computer Graphics and Vision at the Graz University of Technology. He completed his PhD in Information Science and Technology at Osaka University in 2016. From 2016 to 2019, he worked as an Assistant Professor at the Nara Institute of Science and Technology. From 2019 to 2022, he was a Postdoctoral Fellow at the HCI Group at the University of Otago. His research interests are computer vision, eye tracking, and mixed reality.

Naoto Ienaga received his PhD from the Graduate School of Science and Technology, Keio University in 2020. He has been an Assistant Professor at the Faculty of Engineering, Information and Systems, University of Tsukuba since 2021. He now works on designing and developing software by appropriately using machine learning and computer vision techniques for solving practical problems in fisheries and occupational therapy.

Maki Sugimoto received his PhD from the University of Electro-Communications, Tokyo, in 2006. Currently, he is a Professor at the Department of Information and Computer Science, Faculty of Science and Technology, Keio University, Japan. His research interests include embedded optical sensing systems for AR/VR and collective visual sensing systems with machine learning.

10 Embodied Interaction on Constrained Interfaces for Augmented Reality

Lik-Hang Lee, Tristan Braud, and Pan Hui

Contents

10.1 Resurgence of Wearable Computers . . . 240
10.1.1 Interaction with Today's Wearable AR Headsets . . . 240
10.1.2 Drawing a Parallel to Desktops and Smartphones . . . 240
10.1.3 The Constrained Interfaces on Wearable AR Headsets . . . 240
10.1.4 Rethinking on the Constrained AR Interfaces . . . 242
10.1.5 Spotlights of the Chapter . . . 242
10.1.6 Structure . . . 243

10.2 Related Work . . . 243
10.2.1 Freehand Pointing on Constrained Hardware . . . 244
10.2.2 Text Entry on Constrained Screen Real Estate . . . 246
10.2.3 Optimized Text Entry Layout Design . . . 246
10.2.4 Summary . . . 247

10.3 TiPoint . . . 248
10.3.1 System Requirements . . . 248
10.3.2 Interaction Overview . . . 248
10.3.3 Interaction Approaches . . . 248
10.3.4 Mid-air Interaction Strategy for Small-Screen Display . . . 249
10.3.5 Implementation . . . 249
10.3.6 System-Wide Implementation . . . 252
10.3.7 Application Scenarios . . . 252

10.4 HIBEY . . . 253
10.4.1 System Design . . . 253
10.4.2 Uncertainty on Keyboard-Less Environment . . . 254
10.4.3 Implementation and User Performance . . . 256
10.4.4 Application Scenarios . . . 258

10.5 TOFI . . . 258
10.5.1 System Implementation . . . 259
10.5.2 Optimization of the Keyboard Layout . . . 260
10.5.3 User Performance . . . 264
10.5.4 Application Scenarios . . . 265

10.6 Take-Home Message for AR Interaction Techniques . . . 265
10.6.1 Conclusions . . . 265
10.6.2 Future Outlook: Toward the Miniature and Subtle Interfaces . . . 267

References . . . 267

L.-H. Lee (✉)
Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
e-mail: likhang.lee@kaist.ac.kr

T. Braud (✉)
Division of Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong, China
e-mail: braudt@ust.hk

P. Hui
Computational Media and Arts Thrust, Hong Kong University of Science and Technology, Hong Kong, China

Division of Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong, China
e-mail: panhui@ust.hk

Abstract

Wearable computers have seen a recent resurgence in interest and popularity in which augmented reality (AR) smartglasses are poised to influence the way we complete our work and daily tasks. Nowadays, industrial applications of these smartglasses are focused on interior designs, remote collaborations, as well as e-commerce. Under five key constraints on AR smartglasses such as miniature touch interface, small-screen real estate, user mobility, limited computational resource, and short battery life, existing user interaction paradigms designed for desktop computers and smartphones are obsolete and incompatible with the scenarios of AR smartglasses. The cumbersome and difficult interaction with the AR smartglasses becomes a hurdle to their wider industrial applications. Thus, an unmet demand for designing interaction techniques on AR smartglasses is undoubtedly critical.

In this chapter, we present three interaction techniques, namely, TiPoint, HIBEY, and TOFI, in order to enhance object manipulation and text entry in the constrained environment of AR smartglasses. These techniques are devised in a way that leverage on advantageous features of human body and experiences such as the dexterity of fingertip, lexicographical order ingrained

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_10

in our memory, proprioception, as well as opposable thumbs. We thoroughly address the key constraints on AR smartglasses and explore different modalities with various hardware and peripherals.

Keywords

Wearable computing · Human-computer interaction · Interface design · Augmented reality · Target acquisition · Keyboard-less text entry · Constrained interfaces · Thumb-to-finger space interaction · Mid-air interaction · Embodied interaction

10.1 Resurgence of Wearable Computers

Wearable computers, such as smart watches and smartglasses, are on the way of enabling a future in which human users will be accompanied by computer-based personal assistants anytime and anywhere. The number of smart watches and smartglasses being sold is rising in recent years at the yearly growth rates of no less than 160%. Back to the year of 2014, Apple Inc. has first entered the smart watch market, and the number of smart watches sales increases drastically by 28.2 times from 5 million units in 2014 to 141 million units in 2018. Even though the smartglasses are in its infancy stage, the market research (e.g., STATISTA) has projected that the number of smartglasses will follow the blooming trend of smart watches and becomes another significant wearable computer in the routine usage.

While the sales number of wearable computers indicate a wider acceptance to wearable computing, these wearable computers mainly focus on micro-activities for small tasks, such as reading message or receiving email notification on smartglasses [2], as well as overlaying digital information on the top of physical working environment [3]. These micro-activities are regarded as passive – if the users intend to reply a message and type text, switching the operations on the smartphones is unavoidable.

We believe that wearable computers have the potential for more. Instead of only micro-activities, we envision a future in which the users can do the same functionality on nowadays smartphones. While there are many open challenges for human-computer interaction on mobile devices, we focus on improving the interaction techniques on the current smartglasses. The philosophical view is that the wearable computers and human users will adapt to each other through the well-designed human-wearable symbiosis. Wearable computers leverage the resources of human users to resolve the constraint issues such as limited computational resource and small-size screen real estate. On the other hand, human users utilize wearable computers as the attached interfaces of human augmentation.

10.1.1 Interaction with Today's Wearable AR Headsets

Smartglasses are not equipped with touchscreen and direct manipulation on the digital interfaces is not applicable. Touch controllers (Fig. 10.1a) or trackball (Fig. 10.1b) is commonly applied to operate the pointing cursor in the interfaces. Figure 10.1c illustrates the target acquisition on Microsoft HoloLens, where the users apply head movements to control the cursor to the targets and hand gesture to confirm the target selection. Our prior work [1] studies the interaction approaches of 12 smartglasses available in the market, and most of them rely on the external controllers, trackpad, and buttons, as well as voice commands. Even though voice commands are regarded as natural and intuitive, issues such as social acceptance [4] and security vulnerabilities [5] have limited its applicability on smartglasses. Alternative interaction approaches should be considered accordingly.

10.1.2 Drawing a Parallel to Desktops and Smartphones

Before the mouse is invented for desktop computers, interacting with objects on the screen without a pointing device is clumsy and indirect [6]. The mouse makes the interaction with personal computers user-friendly than ever before and leads to the popularity of personal computers. Touchscreen technology leads to the popular use of smartphones due to its user-friendly design,which enables the users to directly manipulate on the objects in the screen. Similar to the wearable computers, early desktop computers and smartphones have limited usage because of its bottleneck in user-friendliness, where the mass population of consumers have expected straightforward interfaces and easy-to-use wearable computers.

10.1.3 The Constrained Interfaces on Wearable AR Headsets

Today we are at a point at which the numbers and types of smartglasses are fast growing and proliferating. However, one of the key differences between the desktop computers/smartphones and the wearable augmented reality headsets is the more **constrained interfaces** on wearable computers. The definition of constrained interfaces in this book chapter is *the limited resources and form factors on wearable augmented reality headsets make challenging constraints on interfaces design, and the constraints include hardware configurations, screen real estate, tiny interface, and user mobility*. The detailed explanation is as follows.

Fig. 10.1 Two AR headsets representing the low-end and high-end commercial products and their input methods. (**a**) Mad Gaze ARES – trackball. (**b**) Microsoft HoloLens – cursor drove by head movement

1. Constrained hardware configuration: Smartglasses and other smart wearables have shared two key limitations: limited computation power and short battery life, which are not favorable for computationally intensive tasks, for instance, the computer-vision hand gestures for interacting with the icons and menus on smartglasses. Google Glass (a low-end smartglasses in the current market) contains a CPU of ARM Cortex-A9 MPCore SMP at 1 GHz clock speed and 1–3 h usage of battery life. This configuration is similar to the desktop computers at the age of 2000, where Intel Inc. made the claim that it was the first to market the CPU of 1 GHz clock speed.In other words, the smartglasses users have worn a decade-old computer on their heads. Even though the semiconductor manufacturers are able to produce smaller yet powerful chipset for smartglasses, the battery will be used up quickly if intensive computation tasks are running. Thus, the interaction techniques under the limited hardware configuration of wearable computers should be designed carefully.
2. Constrained screen real estate (on smartglasses): Field of view (FOV) decides the size of the digital overlays popping up in front of the user sight. For example, Microsoft HoloLens (first generation) provides 30 $\times$ 17 degree FOV, implying 34.5 degree FOV diagonally in the screen ratio of 16:9. The users can see a screen size similar to a 16:9 monitor with 15 in. diagonal that is 2 ft. away from the user sight. The experience of FOVs among the high-end smartglasses in the market, such as Microsoft HoloLens, DAQRI Smart Glasses, and Meta 2, is roughly estimated. The users experience the screen at the size of 1x–3x to a 9.7 in. iPad Pro (240 mm (9.4 in.) (h) 169.5 mm (6.67 in.) (w)). In contrast, the low-end smartglasses with smaller FOV, for instance, Google Glass owning a FOV equivalent to 25 in. screen from 8 ft. away, can barely accommodate one-sentence message and notifications, which implies most apps should keep the information simple and design the interfaces wisely. Considering the limited FOV, the design paradigm for the smartglasses should be reconsidered, where the users have to interact with the physical world through the digital overlays shown on the screen real estate.
3. Indirect touch on miniature interfaces: Without doubt smartphones are the most ubiquitous computing device with spacious touch interfaces, and we assumed that the size of touchscreen devices on smartglasses can be made smaller. But paradoxically, the users encounter indirect manipulation through the miniature-sized interfaces – i.e., it is more difficult to accurately hit the small touch targets on the miniature-sized touch interface as the target acquisition procedures highly rely on visual demands. Due to this constraint, the smartglasses only serve as an extension to the smartphones, and their functions limits to message notification, biometric information collection, user health status monitoring, as well as location positioning and city navigation. In addition, text entry for message input on smartglasses is usually restricted to predefined texts and emojis for one-click replies, because of the limit-sized interfaces on the tiny keypads of full QWERTY soft keyboard.
4. Physical constraint from mobility: The small interfaces for text entry on nowadays smartglasses have not thoroughly considered the issues of mobility and social acceptance, on the top of the input easiness. For instance, the touch interfaces on the frame of Google Glass and the mid-air hand gestures of Microsoft HoloLens require lifting up the hand to the eye level, which is not only tedious but also draws unwanted attention from the surroundings. Thumb-to-finger interaction [4] can serve as an alternative of unnoticeable and subtle text entry; however, the small-sized finger space can barely accommodate the full QWERTY keyboard with two hands [7]. Two-handed text

entry is not suitable for the mobile situation, for instance, holding a shopping bag. Considering a hand should be reserved for the mobile situation, the redesign of thumb-to-finger text entry within a single-handed area becomes necessary.

10.1.4 Rethinking on the Constrained AR Interfaces

The keyboard and mouse we are interacting with desktop computers in the 1990s are designed for sedentary inputs of textual and graphical contents. O'Sullivan and Igoe's [8] depict human users with the desktop computers appearing as an alien with only one finger, an eye, and two ears, which means other capabilities of human being are not yet utilized. Since the past decade, smartphones have demonstrated an evident change in the interaction between human and computer: the smartphone users apply their fingers as a stylus to interact with the icons, menus, and windows on touchscreens that serve as both input and output devices. However, the current design of interaction techniques for wearable computers is principled from the desktops, where rigorous distinction exists between the tangible and spacious interfaces and human users. The body-worn wearable computers serve as an extension of our body [9], but the small-sized wearable computers attached on human bodies poses various constraints, as discussed in Sect. 10.1.3. The direct adoption of these interfaces on wearable computers is considered fatal and makes the user interaction difficult.

Perhaps a closer analog of the touchscreens on smartphones sheds a light on the same interfaces doing both input and output; the body-worn devices and the users' physical body and cognitive ability can bind together as an integrated entity of input and output, and thus resolve the aforementioned constraints as well as establish appropriate input capability. Knowing that human users are very skillful at the physical world including our body in particular [10], the users become a part of the windows, icons, menus, and pointers (WIMP) [11] on the wearable computers. Originated from the theories of embodiment [12] that focus on how our bodies and active experiences influence how we perceive, feel, and think, embodied interaction [13] advocates that the habits, skills, experiences, and abilities of human being that we already have should be at the core of designing the interaction interfaces and techniques. In other words, the users' physical body and cognitive sense act as a key driver in the experience of user interaction itself. The body-worn wearable computers and embodied interaction can view holistically as the coincidence of input and output interfaces. The philosophical view is that the wearable computers will adapt to humans through the well-designed embodied interaction. Considering that the wearable computers serve as an extension of the human body, this book chapter advocates the symbiosis of human users and wearable computers, and pushes the emerging landscape of body-worn computers toward more usable interfaces as well as bidirectional devices. By examining the aforementioned four constrained scenarios, the optimal interaction techniques between human users and wearable computers are identified. These interaction techniques, as listed below, leverage the advantages of the humans' physical body, experiences, and skills.

1. Dexterous fingertips in TiPoint can serve as a pointing device of spacious mid-air interaction;
2. HIBEY applies the human knowledge of alphabetic order to achieve a minimized text entry interface;
3. Human thumbs can naturally locate the keypad of TOFI keyboard within the finger space.

Therefore, this book chapter is concerned with *the embodied interaction that serves as a design strategy to improve the input capabilities of the constrained interfaces on AR headsets*.

10.1.5 Spotlights of the Chapter

In this chapter, we focus our discussion on ***designing interaction techniques for constrained augmented reality interfaces****, enabling users to perform text entry on wearable AR headsets.* In particular, two input systems and corresponding interaction techniques for various constrained scenarios (poor hardware configuration, limited screen real estate, and miniature touchscreen) on smartglasses are designed and implemented, named HIBEY and TOFI.

- **TiPoint**: We implement pointing gestures on low-end smartglasses with constrained computational resources for the direct manipulation of digital overlays. We leverage the embedded monocular camera to capture the video frames of user hand from the egocentric view of the user and identify the fingertip position through the graph-based algorithm. TiPoint serves as an alternative input interface other than the external touchpad or trackballs on the spectacular frame of smartglasses, which achieves a more intuitive interaction enabling the users to the targets through swiping a finger in mid-air. Figure 10.2a demonstrates the usage of TiPoint in 2D interfaces including target acquisition and text entry.
- **HIBEY**: The design of soft keyboard from smartphones is less considerable for augmented reality, as the soft keyboard occupies significant areas of screen real estate and hence hinders the user interaction with the physical world. Therefore, we intend to hide the soft keyboard commonly located at the screen center of smartglasses. Through investigating the user perception to the keyboard-less text

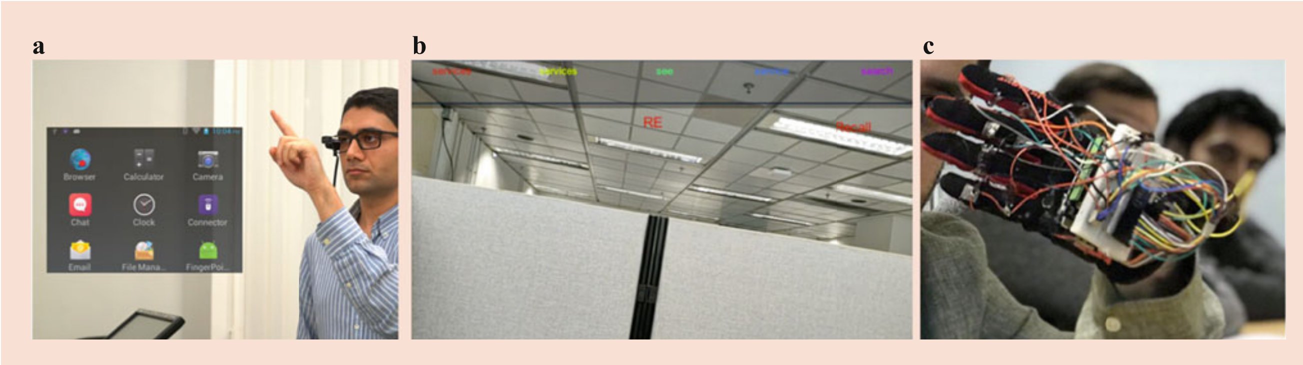


Fig. 10.2 A pictorial description of three select works: (**a**) Target acquisition with TiPoint. (**b**) Keyboard-less interfaces of HIBEY on an augmented reality headset. (**c**) A prototypical text entry wearable named TOFI

entry interface on smartglasses, we propose probabilistic models to handle uncertainty in character selection and word suggestion. Figure 10.2b shows that HIBEY reserves most of the screen real estate. As the proposed text entry interface minimizes to the upper edge area, the central areas are released enabling the user to interact with the physical world without in-between visual blockages.

- **TOFI**: Even though the text entry through thumb-to-finger interaction has appealing properties such as discreet and subtle movements, and provides comfortable yet accurate input with no visual demand on the keyboard layout, existing approaches of two-handed thumb-to-finger interaction limit mobility as both hands are occupied for text entry. Therefore, we shrink the two-handed interfaces to one-handed interfaces for higher mobility and design an optimized layout with quadmetric considerations on a prototypical glove. Thus, we design an ambiguous keyboard that fits into the constrain-sized palm, and four metrics are considered for the miniature keyboard – (i) the goodness of character pair configuration, (ii) the familiarity with the QWERTY-like keyboard, (iii) the easiness of force interaction on the ambiguous keys, and (iv) the level of comfort for thumb reach inside the finger space. Figure 10.2c gives an overview of the 12-keypad configuration of TOFI prototypical glove. Users can subtly choose the characters within the finger space without noticeable movements, and maintain his forearm in a relaxed position and input text through a series of thumb-to-finger touches.

10.1.6 Structure

We begin this book chapter by reviewing the related researches that propose interaction techniques for augmented reality smartglasses (Sect. 10.2). Afterward, we dedicate separate sections to each of the constrained scenarios on the smartglasses: TiPoint (Sect. 10.3), HIBEY (Sect. 10.4), and TOFI (Sect. 10.5). The interaction techniques in three sections share the incremental design approach. That is, the design clues gathered from the human users impact the final decision of the system design. Thus, we primarily discuss the system design in depth, while we describe the final results briefly in the chapter. The full details of the evaluation results are available at [14–16]. We concluded this book chapter through a discussion of the benefits and further usage scenarios of our interaction techniques in a connected overview (Sect. 10.6).

10.2 Related Work

In this section, we first give a brief introduction to the emerging head-worn computers, defined as smartglasses. A review of the related work on interaction techniques for target acquisition and text entry on smartglasses follows.

Smartglasses are defined as head-worn wearable computers. The smartglasses own a chipset for computational capabilities and optical head-mounted displays (OHMDs) for showing digital overlays to the users. Additionally, multiple sensors are available to capture the user's input. The users with the smartglasses can get the display of digital overlays on the top of physical environment, which facilitates the two-way interaction between the augmented reality and the smartglasses users [1].

Touchscreen becomes the most popular modality, and hence taps and touches are the primary gestures for today's commercial mobile devices such as smartphones, smart watches, and digital camera. The nowadays smartglasses have small form factor, and the touch interfaces are shrunken down to miniature size or even unavailable. Smartglasses currently rely on speech recognition as the primary input method accordingly. Therefore, the research community has proposed numerous alternative interaction approaches for the augmented reality smartglasses. It is important to improve the smartglasses with better input capabilities as well as satisfied user experience. The interaction experience is significant as the current input capabilities on smartglasses cause the limited usage of micro-interactions.

Our survey paper has classified the interaction approach for smartglasses into three categories [1] . First, the interaction through handheld device [17] makes use of handheld tangible devices (e.g., smartphones) or a touch-sensitive controller wired with smartglasses (e.g., Sony's SmartEyeglass). Next, the touch-based interaction includes the gestural inputs on devices and on-body touch interaction. The existing works establish touch-sensitive hardware interfaces. Touch-sensitive surface is attached on smart rings, wristband, watches, spectacle frame of glasses, as well as the user's body as electronic skins. Finally, touchless interaction employs non-handheld and non-touch approaches, which are usually regarded as natural interfaces, for instance, hand gestures, head and body movements, gaze interaction, and voice recognition. The remainder of the classes (touch and touchless) are discussed in this section. Figure 10.3 depicts the classification of interaction approaches. The four subcategories under the touch and touchless input are briefly discussed, as follows. The full details are available at [1].

1. **On-device interaction**: The users with on-device interaction can work on a sensible surface of tangible devices including smartglasses [18] and peripheral sensors as an augmented touch surface. Various form sizes can be applied on the tangible devices including but not limited to rings [19–22], wristbands [23, 24], sleeves [25], and belts [26].
2. **On-body interaction**: The human skin can serve as the interaction surface with tactile cue as an additional feedback mechanism, which results in higher performance than touchless input especially mid-air input [27]. The existing studies on on-body interaction focus on the interaction on various parts of human body. The palms [27–32], forearms [33–35], fingers [30, 36, 37], faces [38], and ears [39] are usual interaction surfaces because of the properties of easy-to-reach and social acceptance.
3. **Hands-free interaction**: The users with hands-free interaction involve no hand control, including voice recognition, head gestures [40], eye tracking [41, 42], as well as tongue gestures [43].
4. **Freehand interaction**: The users with freehand interaction can employ hand gestural inputs with the advantages of intuitiveness and convenience in the environment [44]. Hand gestures can be classified into eight types [45]: pointing, semaphoric-static, semaphoric-dynamic, semaphoric-stroke, pantomimic, iconic-static, iconic-dynamic, and manipulation. Wearable sensors or embedded cameras are commonly employed to capture the image frames of the user's hands such as glove [46–49], camera (RGB camera [50–54], and depth camera [23, 55, 56]), which is highly relevant to Sects. 10.3 and 10.4.

Among the above subcategories, this book chapter focuses on the pointing techniques and text entry on smartglasses employing various hardware and peripheral sensors. The approaches discussed in Sects. 10.3 and 10.4 apply the embedded camera available on the augmented reality smartglasses. Also, we discuss the thumb-to-finger interaction, a type of on-body interaction with the attached sensors inside the finger space, designated for target acquisition and text entry tasks, which is highly relevant to Sect. 10.5.

10.2.1 Freehand Pointing on Constrained Hardware

Direct pointing and manipulation are analog to the touch interface on touchscreen devices, but the mouse pointing device is not available on smartglasses and the visual content is no longer touchable. Through employing freehand interaction, a pointing hand [54] acts as a mouse cursor [57], enabling user interactions indirectly with 2D interfaces (e.g., multi-button menus [55]) projected on the optical display of augmented reality smartglasses [54, 55]. Hand gestures also serve other purposes such as fiducial marker [52] to show up 3D augmented reality objects in the physical environment, and users accordingly can manipulate with these 3D objects with bare hands [56] in office environment [51]. The hand gestural interaction is regarded as a natural and intuitive method. However, it shows slower completion time than the interaction on touch-sensitive interfaces due to the long dwell time [58]. In other words, touch-based interaction achieves better user performance. Thus, gesture type toward barehanded direct pointing [59] is a promising research direction for interface interaction on the augmented reality smartglasses.

Unfortunately, there have been very few existing works exploring the direction using fingertip as a feature for direct pointing and manipulative interaction. To the best of our understanding, Fingermouse [60] is a very early gesture-mouse system for wearable computers, which applies stereoscopic vision to recognize the hand and finger position. Their work merely provides a method of hand image processing but does not specify the smartglasses application scenarios and does not consider the usability of the hardware interface for nowadays smartglasses.

We find two highly relevant works but they are far from practical uses on smartglasses. A recent work [61] from the computer vision community applies deep learning to improve the accuracy of hand image recognition in a cluttered environment but does not show the implementation for resource-constrained AR headsets. Another work [62] focuses on the algorithmic enhancement in terms of accuracy of recognizing the finger counting. However, their work relies on the power-hungry system. Also, it counts finger and interprets the meaning of finger counts. Therefore the

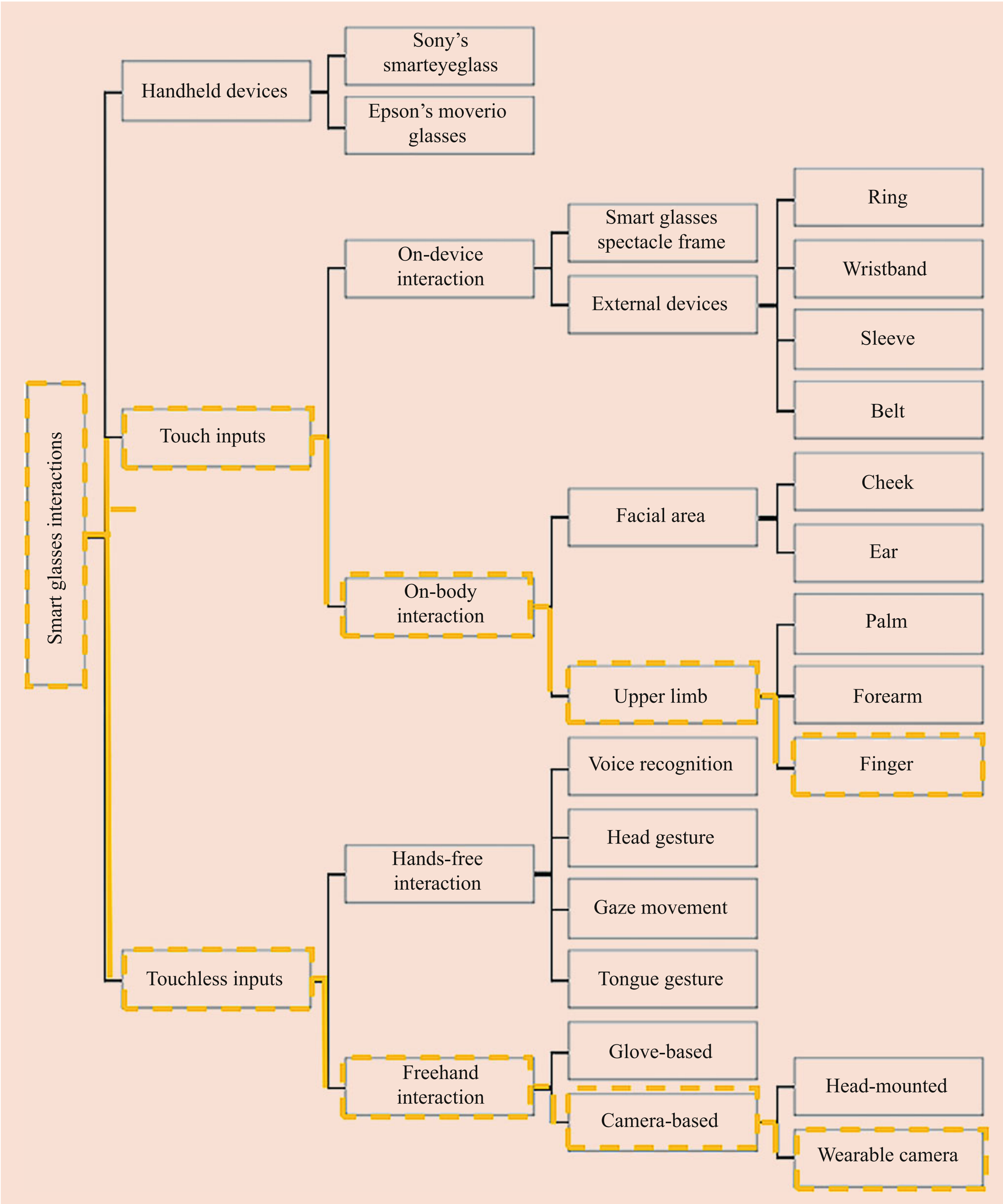


Fig. 10.3 Classification of interaction approaches for smartglasses as defined in [1], with the focuses of the interaction techniques in this chapter, leveraging finger-based interaction and camera-based freehand interaction

finger counting is not ideal for tasks of fast repetition (e.g., text input) due to dwelling time between two gestures [63]. In addition, most of them do not apply to the monocular camera available in smartglass as well as smartphones [64, 65]. A prototype system with a binocular camera mounted on the smartglasses combines both the 2D color features and 3D depth information to detect the fingertip position. However, 3D depth camera is usually unavailable in smartphones and smartglasses, except some self-contained head-mounted display (e.g., Microsoft HoloLens). Although the Mixed Reality Toolkit of Microsoft HoloLens 2 offers a more comprehensive hand gesture set, the key enabler of depth cameras, again, is not always available on the majority of smartglasses. Furthermore, the additional cost of depth cameras makes the smartglasses less attractive to consumers as well as manufacturers. As a result, a research opportunity is left for further vision-based fingertip interaction research with the premise of constrained hardware configuration.

To the best of our knowledge, most direct pointing algorithms based on aforementioned body part detection such as facial area, palm, and entire hand are designed for powerful systems, where resources are abundant, and do not apply to the constrained hardware of smartglasses. To address the aforementioned constraints on AR smartglasses (Sect. 10.1.3, new solutions are needed to resolve most of the problematics raised by other solutions by providing a software-based direct pointing method affordable and practical to run smoothly on present smart glasses and smartphone headsets.

10.2.2 Text Entry on Constrained Screen Real Estate

On Google Glass and Microsoft HoloLens, speech recognition is the key text input method for inquiring various contents. Nevertheless, it subjects to some less appropriate scenarios especially in shared or noisy environments, for example, making disturbance and obtrusion to the people in the surrounding or being accidentally activated by environmental noise [66]. In addition, the mute individuals cannot employ speech to interact with the smartglasses, and thus it is less acceptable than body gestures or handheld devices input approaches [67]. Even though the research community of human-computer interaction research has conducted a number of studies on text input [68], the text entry problem on smartglasses has recently been solved by the on-device and on-body interaction approaches, where the researchers design the text entry interfaces leveraging the spectacle frame of smartglasses and peripheral sensors on external devices.

For example, SwipeZone [18] divides the touchable spectacle frame into three zones. Users can swipe on one of these zones and select the character mapped to the zone. In another example, a ring-form input device [69] allows character selection on a virtual QWERTY keyboard through a two-stage method, as follows. First, the user targets a character zone by performing hand movement vertically and horizontally inside the virtual keyboard. In the virtual keyboard, an ambiguous keypad contains three consecutive keys. The user thus confirms the target key through moving a finger (i.e., middle, index, or ring fingers). On-device interaction approaches demonstrate precise and responsive interaction. However, the major drawbacks are cost of the additional devices and the time for putting on the device [70].

Alternatively, the optimal use of an external controller wired with smartglasses enables users to perform off-hand text entry, which allows users to operate a cursor and select keys on a virtual on-screen keyboard such as Dasher input system [71]. Other text entry systems employ an imaginary QWERTY keyboard on the user's palm [31], as well as writing graffiti words on the user's palm [32].

Freehand interactions have exhibited their outstanding capabilities in 3D interfaces. Nevertheless, the current works of vision-based freehand interactions are primarily interested in the manipulation of 3D objects [52]. Most of the users prefer interacting with 3D objects through hand gesture over the touch input approaches. Indeed, performing hand gesture in front of the facial area is natural and straightforward [17]. However, the freehand interactions have drawbacks including the low entry rates and long dwelling times of gesture recognition [63]. A prior work on mid-air text input on a QWERTY keyboard achieves up to 29.2 WPM [72]. However, the screen space is mostly consumed by the virtual keyboard, and the sensor of LEAP Motion is not available on most of the augmented reality smartglasses [1].

However, the above discussed works have neglected the key constraint of small-sized screen real estate on augmented reality smartglasses. On the augmented reality smartglasses(Fig. 10.4), all the above on-screen keyboards or other forms of text entry interfaces located at the center of the small display (Sect. 10.1.3), causing space-consuming and inconvenient user interaction, thus violates the original purpose of allocating digital overlays in parallel to the physical objects.

10.2.3 Optimized Text Entry Layout Design

In the existing literature, numerous studies propose multi-metric optimization for keyboard layouts on smartphones. Dunlop et al. [73] evaluate the optimization of semi-ambiguous keyboards on mobile phones. They strive to balance the keypad size with the prediction problems, using a keyboard layout similar to the QWERTY configuration. In

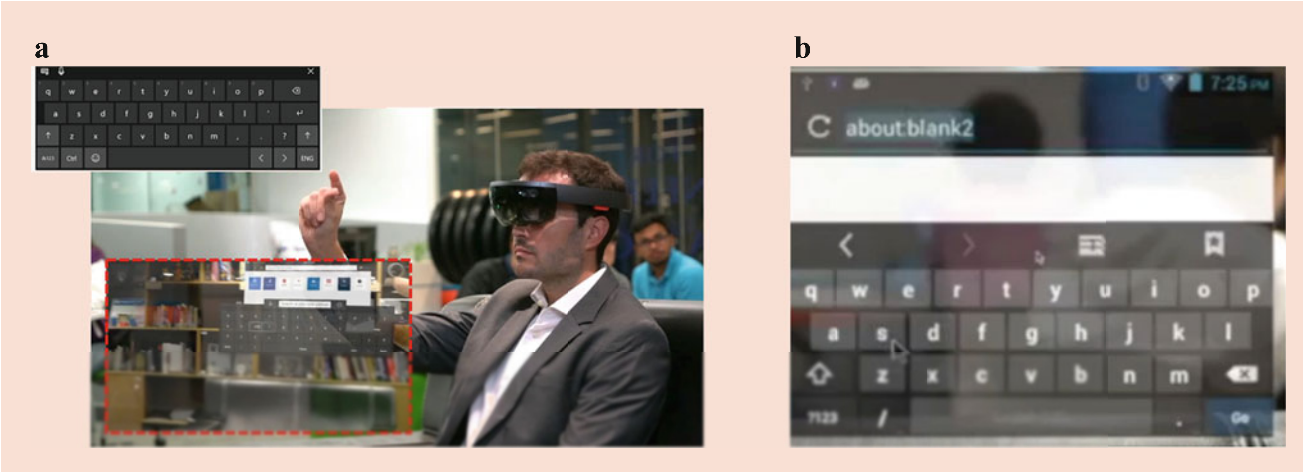


Fig. 10.4 Existing keyboards on AR smartglasses: Microsoft HoloLens (**a**) and Mad Gaze (**b**)

another work [74], they present a triple metric optimization for keyboard layouts considering speed, familiarity, and spell-checking ambiguity. Smith et al. [75] find a Pareto frontier between three metrics of gesture clarity, gesture speed, and the similarity to the QWERTY keyboard. Gong et al. [76] formulate the optimization problem with four metrics including comfort, accuracy, word disambiguation, and layout learnability.

Furthermore, the arrival of force-sensitive touchscreen interfaces on commercial products like smartphones (e.g., iPhone sixth gen. or later version) and smart watches (e.g., Apple Watch) has risen the research interests in text input. One promising direction is to leverage the force input to augment the keyboard layout. Prior works prove the human's ability to exert force in subtle levels (six levels [77] and up to ten levels [78] with sufficient feedback). The force intensity and division impact the error rate of force exertion [79], rising drastically from 4.9% for one layer to 35% for six layers.

The literature on force-assisted target acquisition and text input for tiny touchscreen interfaces is very limited, as discussed below. To the best of our knowledge, our prior works serves (e.g., TOFI) as the first groundwork of force-sensitive interfaces on augmented reality smartglasses. Hsiu et al. [80] propose an ambiguous QWERTY keyboard on smart watches in which every key contains two characters. The two characters in the same key are distinguished by two discrete levels of force exerted on the screen. Their two-level QWERTY layout achieves 12.4 WPM due to simpler force level design but occupies most of the screen estate on the smart watches. Another study [81] proposes a force-assisted scanning ambiguous keyboard (SAK) on smartphones. It requires only a thumb-sized interface to exert force and select characters. Their one-key layout achieves an average 4.2 WPM for character-level input solely by user judgment and 11 WPM for word-level input supported by probabilistic word auto-completion.

With the arrival of force-sensitive touchscreen, it is worthwhile to study the disambiguation methods and the optimized layout considering the dimension of force input with the aforementioned metrics in the existing works. In contrast, we are interested in the feasibility and optimization of an ambiguous keyboard with multiple character keys. We aim at reducing the number of layers to maintain a low error rate, in order to achieve a highly mobile AR text entry solution within the finger space (a type of on-body interaction). In addition, the optimized keyboard layout has the advantages of single-gesture inputs (e.g., one tap maps to one character), guaranteeing reasonable text entry performance in the constrained AR interfaces.

10.2.4 Summary

Although the target acquisition and text entry have been widely studied in the field of computer-human interaction, existing design of interaction techniques for personal computers and smartphones becomes obsolete in the smartglasses due to the additional constraints on these devices. On smartglasses, the digital overlays are no longer touchable due to the missing touchscreens. Freehand interaction provides an alternative to realize the direct manipulation of the digital overlay in mid-air. Also, the small-sized touch interfaces on smartglasses restricted the volume of information and the available contents. There are a lot of research opportunities in redesigning interaction techniques for target acquisition and text entry on these devices. Even though the prediction-based or machine learning approaches are emerging in the field of computer-human interaction, our works of target acquisition and text entry in the later sections focus on the interaction approaches and the layout designs, in which the artificial intelligence techniques such as hand image recognition, probabilistic model, dictionary model, and optimization models serve as the tools to facilitate the interface and layout designs.

10.3 TiPoint

As of now, the low-end smartglasses market is still experimenting with interaction techniques and many different methods are in use, none providing full satisfaction to the users. Most of these headsets rely on hardware solutions such as touchpad and trackball. However, such devices either are embedded on the glass frame, which forces the user to raise his arm for interaction, or consist in cumbersome external apparatus. Even though smartglasses manufacturers start to adapt the interface to simplify the interactions, it often results in difficult user inputs and hence stripping functionalities. For instance, Google Glass relies exclusively on voice text input and does not allow manual additions to the embedded dictionary. Microsoft HoloLens takes a radically different approach, by combining head movement and hand gesture detection. However, due to the weight of the headset, long-time use leads to fatigue and physical strain. These examples are the most representative of the current situation regarding interaction methods. More recent smartglasses follow the same trend. Vuzix Blade relies on a touchpad and voice control, while Sony's SmartEyeglass uses an external controller.

TiPoint [14] is a barehanded pointing technique for direct manipulation with digital overlays of smartglasses. TiPoint is implemented as a lightweight computer vision algorithm to detect the fingertip of the user, providing an intuitive and efficient interface for AR smartglasses. Additionally, the user's fingertip is detected by the embedded monocular camera, instead of the expensive depth cameras. Users with TiPoint can directly control the movement of the cursor. Once the fingertip-driven pointer reaches the target location, the user either can tap on the hardware interface or perform a specific gesture to select the object.

As our solution relies on barehanded mid-air interaction, it solves most of the aforementioned issues. First, TiPoint does not depend on voice or head movement; therefore, its utilization is more concealed while respecting the user's privacy. Second, it does not rely on bulky external hardware to work, simplifying the user experience. Third, TiPoint has been thought as a software-only solution, which can be embedded as an interaction layer in specific applications, or enabled system-wise, to handle every user interaction on AR smartglasses.

10.3.1 System Requirements

TiPoint, as a nonintrusive interaction technique, for easy-to-use interaction with AR smartglasses considers the following five principles:

1. A nonintrusive interaction technique enables users to perform interaction with no additional hardware or fiducial markers required. Freehand interaction in mid-air is one of the options.
2. The repetitive point-and-click operation for target acquisition should be swift and simple. Semaphoric gestures and sign language interactions have drawbacks such as the gesture complexity and dwelling time. Thus, an intuitive pointing technique similar to the mouse cursor is preferred.
3. The subtle movements are preferred because large body movements are prone to fatigue.
4. An affordable and efficient algorithm supports the fingertip-based interaction technique to be used offline. It is worthwhile to mention that the smartglasses present limited hardware specifications (e.g., computation resources and battery life).
5. Finally, the interaction technique should enhance the performance of doing various tasks, compared to other hardware solutions such as touchpads or trackballs.

10.3.2 Interaction Overview

The user's fingertip serves as a barehanded mouse cursor in mobile scenarios. Users with TiPoint can interact with icons and menus, and the corresponding swipe-based gestures in the interface, detected and tracked by the monocular camera embedded in the smartglasses. The user's fingertip becomes a distinctive feature in the interaction, and its mid-air position is mapped to the cursor's position, in a similar way a mouse drives the pointer for a personal computer. Our technique allows the user to move one's hand in the three-dimensional space freely, which enables mid-air interaction with comfortable and natural postures in front of the smartglasses camera. As users keep a lifting hand in front of camera, TiPoint reduces the user's arm fatigue by keeping the user's hand movements to a minimum area of (24–32 × 16–21 cm) [82, 83].

10.3.3 Interaction Approaches

TiPoint offers two interaction approaches, namely, freehand mode and fast-repetitive mode. The freehand mode is able to handle scenarios requiring multiple input options (e.g., browser application), including scrolling, selecting icons, and zooming in and out, while the fast-repetitive mode enables users to focus on a number of select-and-click operations in a fast and responsive manner (e.g., character input). The combination of freehand and fast-repetitive modes allows full coverage of interaction missions in the situation of indirect manipulation, satisfying the five principles as mentioned in Sect. 10.3.1.

Freehand Mode

The first interaction approach involves only mid-air pointing gestures, which encourages freehand interaction with AR and VR headsets. In this mode, the cursor directly follows the user's fingertip, and actions can be triggered after hovering the cursor on top of a point of interest for a short duration. The cursor's color then changes to signal the user that a gesture is expected.

We define three main categories of hand gestures:

1. *Mid-air click*. Users can click on icons in the headset interface, which is analogous to finger-tapping on a touchscreen interface. After hovering the cursor for half a second on top of the object, the user can slightly lower his fingertip to perform a click at the designated location.
2. *Center-to-edge swipe*. In this scenario, we define an area in the middle of the screen. After hovering the cursor for half a second in the screen area, the gesture is triggered by moving the fingertip toward any of the edges. The swipe action to a particular edge can be mapped to the operations such as scrolling within app menus or flipping to previous or next pages on a web browser.
3. *Corner-to-corner swipe*. Similar to center-to-edge swipe, the user can execute the gestures by performing a swipe from one corner to another corner. This mode is activated when the cursor hovers in a corner of the screen for half a second. Corner-to-corner offers 12 additional input options, which can be used to handle objects with up to six attributes. However, the corner-to-corner swipe should be kept to a minimum to reduce arm movements and hence user fatigue.

Fast-Repetitive Mode

This interaction approach mixes pointing gestures with the hardware interface of the smartglasses. The miniature touch-sensitive interface on the spectacle frame of smartglasses enables users to perform confirmation of the pointing targets, while the user can maintain the intuitive pointing hand to locate the target in the digital overlay.

10.3.4 Mid-air Interaction Strategy for Small-Screen Display

On the touchscreen interfaces, the termination of gestural interaction is defined by the user's finger leaving the touchscreen. In mid-air with TiPoint, the system needs to distinguish deliberate from unintended hand movements as termination signals in order to avoid the Midas touch problem [63]. TiPoint takes corner and edges as end points of swipe gestures, which allows a quick and unidirectional swipe, with the following advantages. It helps to maintain the simplicity of the small-screen interface. The number of functional icons can be reduced with the assistance of gestural input, and the saved space can be used for content display. It has a sufficient number of gestural options. The proposed interaction approach provides a total of 17 input options to handle various tasks in the headset interfaces, ranging from selecting icons and objects, controlling 2D interfaces such as scrolling page up and down, to manipulating objects with multiple attributes such as 3D objects. The two interaction approaches can be an example of bimodal inputs, where they can be switched depending on the input scenario [84]. We summarize all the states of freehand and fast-repetitive modes in Fig. 10.5.

10.3.5 Implementation

TiPoint is implemented as a standalone demonstration application. The hand detection algorithm itself and swipe-based gestures are implemented as a library written in Java, in order to be easily included in other applications.

We use a camera resolution of 320×240 pixels as it is more suitable for our image segmentation process. This resolution reduces the artifacts while providing acceptable (>15 *FPS*) performance. Moreover, it corresponds exactly to the screen resolution, making the mapping of the cursor to the fingertip easier. First, the *OpenCV* library converts the image from YUV420sp to HSV. We then split the image into 10×10 pixel squares to create our data structure of the extended disjoint union representation. Finally, we disable the Auto Exposure and Auto White Balance and set Focus to Infinity to keep a consistent image with lighting condition changes. The fingertip detection consists of three phases: image segmentation, disjoint union arborescence construction, and fingertip location estimation. These steps are represented in Fig. 10.6.

The image segmentation phase starts with converting the image from Android standard color space YUV420sp to HSV. We then apply a threshold to extract the skin tone color and return a binary image. We denote the output of

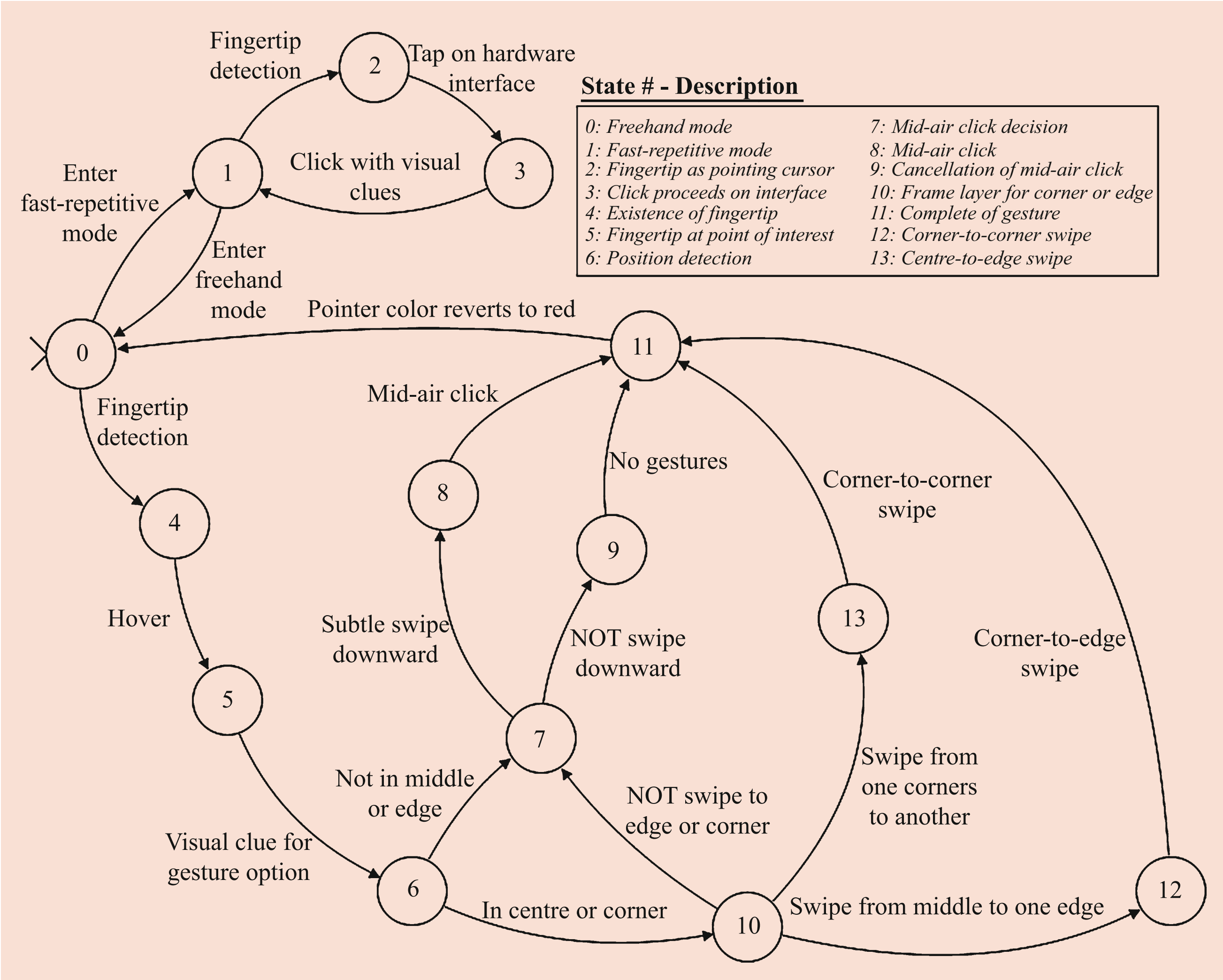


Fig. 10.5 State diagram for pointing gesture and swipe-based gestures

this phase with the binary function $I(x, y) \in \{0, 1\}$ so that $0 \leq x < W$, $0 \leq y < H$, where W and H represent the width and height of the image. The binary function $I(x, y) = 1$ if and only if the pixel at location (x, y) belongs to the skin tone and $I(x, y) = 0$ otherwise. For simplicity reasons, we employ the HSV threshold values defined by Sobottka et al. [85] for Asian and Caucasian skin ($0 < H < 50, 0.23 < S < 0.68, 0 < V < 1$).

A number of works employ morphological transformations to remove the artifacts from the resulting threshold image; however, morphological operations (specifically opening and closing) are computationally demanding, which is not practical for the smartglasses of limited computational power and short battery life. Therefore, we propose an alternative filter method to remove artifacts at the graph construction stage of a disjoint union arborescence. Therefore, we apply a graph-based technique named disjoint union arborescence construction for fingertip detection. In graph theory, an arborescence graph [86] is a directed acyclic graph where there is only one path from the root node to every other node in the graph.

Let $A\langle V, E\rangle$ represent the arborescence graph with the set of vertices V and the set of edges E. We let $DA = \{A_1\langle V, E\rangle, A_2\langle V, E\rangle, \ldots, A_m\langle V, E\rangle\}$ denote a set of m arborescence graphs where the set of vertices for any two arborescence graph in DA is disjoint. We call the set DA the disjoint union arborescence set. Our TiPoint method creates an efficient data structure for the set DA to be constructed from $I(x, y)$. A node $v_{(x,y)}$ belongs to an arborescence graph $A_i\langle V, E\rangle$ according to the following criteria:

$$v_{(x,y)} \in V \Leftrightarrow \quad \forall (i, j) \in F, \quad I(x + i, y + j) = 1 \qquad (10.1)$$

where F is set of coordinates that defines a filter of size S:

$$F = \{(i, j) \mid i = 0 \wedge 0 \leq j < S\} \cup \{(i, j) \mid j = 0 \wedge 0 \leq i < S\} \qquad (10.2)$$

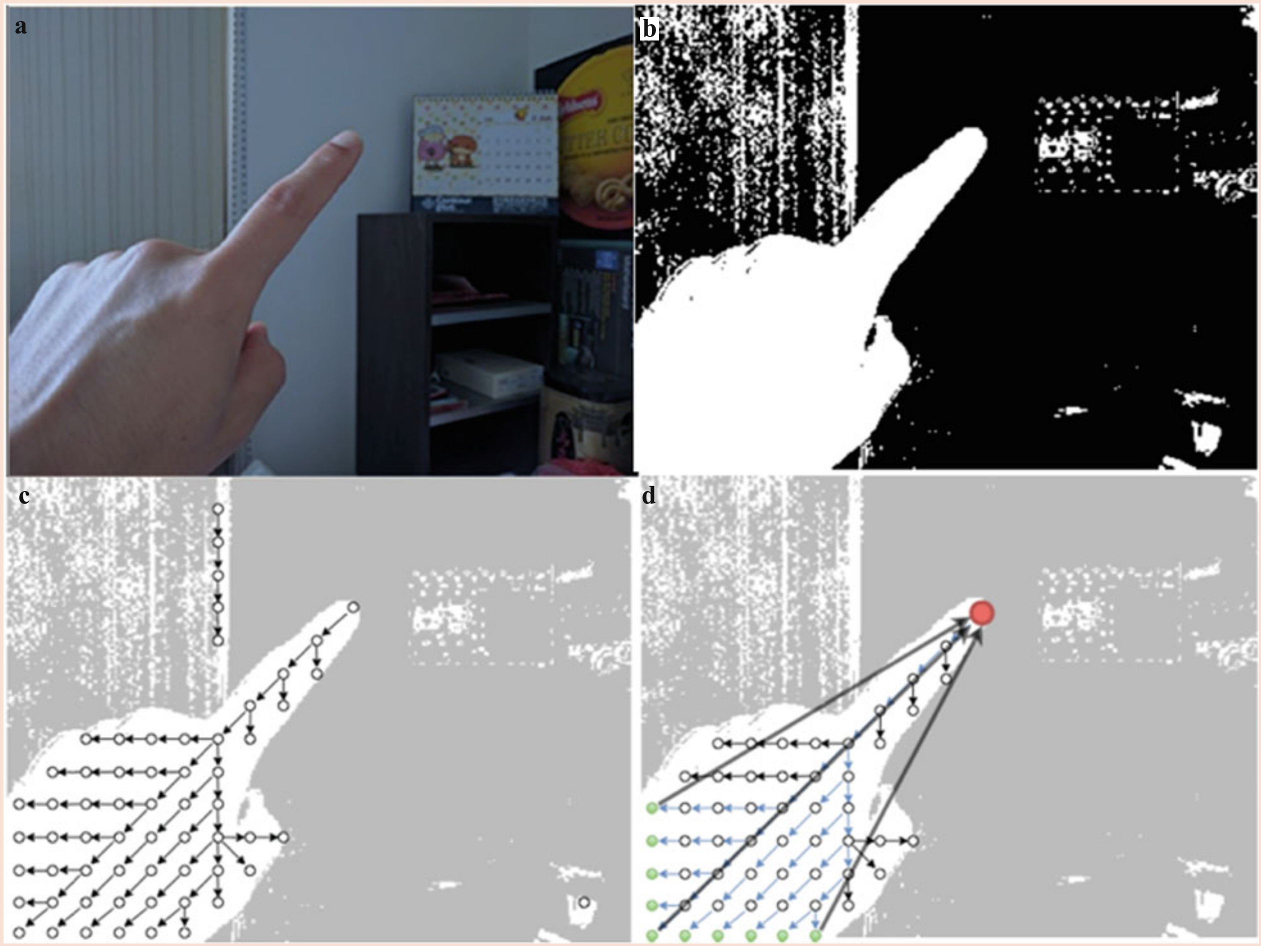


Fig. 10.6 The three phases used in our TiPoint algorithm for detecting the location of the fingertip: (**a**) original image, (**b**) segmented hand region, and (**c**) arborescence graph of segmented hand region

The image $I(x, y)$ is scanned by a sliding window of size S, and for each sliding window, the condition in Eq. (10.1) is applied to the pixel at the center of the window. If the vertex $v_{x,y}$ is chosen, then a new arborescence graph is added to the set DA and a recursive breadth-first search is initiated to direct the sliding window on the neighboring windows where there is a potential new node. The breadth-first search only moves the sliding window on the same scan line or the line below because the image is always scanned from the top row to the bottom row. During the breadth-first search operation, our algorithm incrementally marks the depth of each node and updates the number of nodes belonging to each depth level. The algorithm also marks the visited sliding windows so that they will never be visited again in future scans. Therefore, the image will be scanned in linear time in the size of the number of sliding windows (e.g., for an image of size 320×240 and a filter size $S = 10$, there are $32 \times 24 = 768$ sliding windows).

Regarding the fingertip location estimation, the TiPoint algorithm returns the data structure representing the set DA including the depth level of each node and the number of nodes at any given depth. In this phase, the algorithm selects the arborescence graph with the maximum number of nodes from DA, and the fingertip is located at the root node of the graph. The hand orientation can also be computed by choosing the nodes on the longest path from the root node in the graph and finding the vector that connects the root node to them as shown in Fig. 10.6.

These three steps are computationally affordable to accommodate for the low processing power of any kind of smartglasses, from the mass market (e.g., Mad Gaze, Google Glass) to higher end headsets such as Microsoft HoloLens.

As TiPoint detects the fingertip of the user thanks to his skin color, several problems may occur:

- Darker skin tones cannot be recognized by the system due to the threshold function. This problem can be solved by prior calibration. The user may take several pictures of his bare hand for TiPoint to take color samples.
- Other individuals in the background may cause interference in the fingertip detection. We apply two mechanisms

to solve this issue. First of all, we only perform our computations on the largest detected object. As the user's hand is the closest to the camera, it is the most likely to be detected. Second, low-end glasses such as Google Glass or Mad Gaze feature cameras with very low sensitivity. As such, colors tend to fade away after a couple of meters. An individual located further than 3 meters will not be detected as his skin color on the picture will not be within the threshold.

- Some large objects may be skin-colored. In our experiments, we encountered the case of a beige sofa that would be detected as a finger. For some objects, we may be able to increase the performance by calibrating the system to the exact skin color of the user.
- In order to construct our disjoint union arborescence, the system needs pre-configuration regarding which hand is used. Furthermore, it assumes a closed fist with only a single finger pointing out.

TiPoint aims at providing an interaction method for low-end smartglasses. As such, it is limited by the hardware specifications. TiPoint works well in simple scenarios and can withstand more complex situations, thanks to our implementation choices. With the evolution of smartglasses hardware, we can foresee further techniques to improve TiPoint's performances (stereoscopic vision, image recognition).

10.3.6 System-Wide Implementation

As the last part of our testing, we extend the implementation of TiPoint to a system-wide pointing technique for Mad Gaze ARES. We focus on these smartglasses as their interface follows the WIMP paradigm while providing a full Android environment, facilitating the portage of the application-level library. Moreover, as we explained in the introduction, the pointing device of Mad Gaze relies on a trackball for which interaction relies on a combination of two simple actions, which allows us to easily emulate such actions.

The Mad Gaze trackball operates as a mouse in the operating system. As such, we emulate this behavior using the uinput module. This module provides a file descriptor in `/dev/uinput` in which we can write to send mouse and keyboard events. As this file is located in a protected location, we need root access for read and write access. For the core of our implementation, we reuse the Java library developed for the test application. The read and write operations using the uinput driver are performed in a separate C code using the NDK, necessary to perform such low-level operations.

For facility of distribution, we implement TiPoint as an Android application requiring minimal intervention from the user. Although the application requires root access to emulate the trackball (which can be solved by giving the appropriate permissions to the file during the OS building process), TiPoint can be installed through any usual mechanism without requiring heavy modification to the kernel or the operating system. Due to this development model, we cannot directly access the hardware of the smartglasses, especially the raw video frames from the camera. We therefore have to rely on the Java Camera API for processing the images. For security and privacy reasons, the Camera API requires displaying the camera view somewhere on the screen. We skirt this limitation by reducing the camera view to a 1×1 pixel box at the top left corner of the screen, where it remains unnoticeable. We proceed to access the rest of the system from the application by closing the main activity as soon as the application launches and run the image processing in a background process. To display the pointer, we use an overlay view that allows us to display a red dot at the fingertip location on top of the system. We represent the camera and the pointer overlays.

We focus on the freehand interaction principles described in Sect. 10.3.3. We implement the following features:

1. **Moving pointer**: The red dot follows the fingertip in real time.
2. **Click**: When the system detects the pointer to remain in the same location for more than 500 ms, TiPoint turns into action mode. 500 ms of hovering time is designed based on the rule of thumb. We recruited a total of five participants to check their adaptability to the hovering time. All of them considered that the 500 ms of hovering time is an appropriate waiting time. The pointer changes to green, and the user confirms the action by slightly tilting its fingertip downward.
3. **Scroll**: This serves as an example of **center-to-edge swipe**. Stabilizing the pointer for 500 ms in the center of the screen activates the scroll mode. The pointer turns to cyan and can be used to scroll through menus. The user can quit the scroll mode either by pushing the pointer "outside" of the screen or by stabilizing the pointer for another 500 ms.

10.3.7 Application Scenarios

To demonstrate that TiPoint is applicable to smartglasses with various hardware specifications, we implement it on Google Glass v.1.0 and Mad Gaze ARES Glass. Google Glass runs on Android 4.4.0 with a 1.2 GHz Dual Core CPU, 1.0 GB RAM, 16 GB storage capacity, 5-megapixel camera, and 570 mAH battery life. Mad Gaze has the following specifications: CPU with 1.2 GHz Quad Core, 512 MB RAM, 4 GB storage capacity, 5-megapixel camera, and 370 mAH battery life, Android 4.2.2. The Mad Gaze is chosen because the device has a slightly more challenging hardware specification to run our technique than Google Glass. Both have similar characteristics with current mass market glasses and are representative of what can be achieved with present hardware.

We present the capabilities of a system-wide implementation of TiPoint in a video (https://www.youtube.com/watch?v=fjWs31o6ffc) The video demonstrates the character input for a website URL with freehand mode. The user can use their finger to tap on the virtual keyboard in mid-air. Also, the user can navigate the application main page and select the target application in mid-air. The swipe gestures toward edge and corner are applied in the scrolling of application main page.

One of the extended applications is the control of large screen display. Nowadays, a number of exhibition centers have been equipped with large-screen display. Instead of standing at a fixed position adjacent to the personal computer, the user with smartglasses can manipulate with the digital contents on large-screen display at any position, for instance, standing together with the audience and facing to the screen. We also create a game called "Wakkamole," in which the user can intuitively use their finger (i.e., mid-air tap gesture) to beat the character, namely, "Doctoral Mole." Finally, we include an example of smartglasses with TiPoint as a control aid for presentation in public area. Other potential uses of the fingertip-based interaction technique are as follows. The fast-repetitive mode can be extended into mid-air drawing, text typing on a virtual keyboard, and gestural (iconic) input based on hand-stroke drawing (circle, triangle, square), where the input is initiated when the user's finger taps on the hardware interface and terminated once the finger tap is released.

To sum up, TiPoint is a nonintrusive, seamless, mid-air interaction technique for human-smartglasses interaction without any additional ambient sensor and instrumental glove. We also introduce two interaction modes, namely, freehand and fast-repetitive. The bimodal interaction relies on gestures and hardware interfaces for rapid input control. Indeed, gestures are preferable in the context of navigation in a WIMP system, while a hardware click enhances the experience for repetitive tasks such as typing. TiPoint has been implemented on Google Glass and Mad Gaze, and potentially can be applied to other smartphone headsets. TiPoint enables the users to leverage their dexterous fingertips to acquire the targets inside the screen. The character selection occupies the majority of screen real estate on the smartglasses. TiPoint [14] is a real-time, affordable, and efficient mid-air interaction technique with a high level of user satisfaction. The proposed technique achieves real-time performance on the constrained hardware configuration. When submitted to real user evaluation, TiPoint proved easy to use and performed 1.82 times faster than the interaction on the hardware interface. In the next section, through applying a mid-air pointing gesture, we present a keyboard-less interface to alleviate the issue of constrained screen real estate.

10.4 HIBEY

This section introduces a convenient and unobtrusive text entry system, named HIBEY [87]. The smartglasses users with HIBEY can enter characters and words in AR without any additional proprietary sensors. HIBEY saves the majority of the screen's real estate from the bulky soft keyboards. Hence, through offering a keyboard-less text input experience, the applications in the AR environment of smartglasses (e.g., Microsoft HoloLens) result in more integrated user interaction with the physical worlds as well as the digital overlays. As people rarely invest time in learning new keyboard layouts [88], HIBEY features with fast learning layout by leveraging the advantages of putting the characters in alphabetical order, which also leads to performance improvement and better usability to novice users [89]. HIBEY encourages users to maintain continuous hand gesture interactions in a holographic environment using a keyboard-less approach. That is, HIBEY only requires the user to perform a single discrete hand gesture, i.e., mid-air tap, to initialize the character selection. The user can hold a pointing gesture throughout the text input. During the selection of characters, the user targets characters arranged along one horizontal line. To end the character selection, the user dismisses the hand gesture.

HIBEY works on Microsoft HoloLens that is a high-end smartglass enabling holographic experiences with a wide angle of user view. HIBEY consists of three key stages for typing words with freehand input in mid-air: (1) a pointing gesture can choose and grab the target character in mid-air; (2) a virtual touchscreen in mid-air in the planar boundary between the preparation zone and the fast-forward zone serves as an imaginary plane for character selection. Accordingly, users can perform text entry without a bulky soft keyboard; and (3) the system computes the coordinates on the planar boundary, and afterward the statistical decoder, containing a language model and a spatial model, provides the suitable word phrases.

10.4.1 System Design

In the text entry interface of HIBEY, three connected zones, as below explained, exist in the 3D space. The user moves his arm to choose the characters in an invisible one-line layout.

The position of the hand pointing gesture can confirm or recall the characters through traversing the three zones: (1) preparation (P-zone), (2) fast-forward (F-zone), and (3) recall zones (R-zone). In the R-zone, the user can put the mid-air hand to the position of the target character among the horizontal line of input options such as the Roman alphabet. The user confirms the character selection by forwarding his hand to the F-zone. In the F-zone, the user can speed up the character selection by moving the user's hand forward until the character is confirmed. In the confirmation process, the selected character will move to the closer edge of the interaction area from the farther edge. The relative depth of the user's hand is directly proportional to the character's movement speed toward the closer edge. In the R-zone, the user's hand serves as a backspace function for erasing some unintended inputs, e.g., mistakenly selected characters.

Character Keys

Character keys are located in their default positions at the farther edge of the AR keyboard-less text entry environment. The one-line horizontal layout consists of the 27 characters (the 26 characters from the Roman alphabet plus the white space '_'). The text entry layout follows the alphabetical order and assigns the white space on the right-hand side of the character "z." Alphabetical order layout is suggested in HIBEY as the novice users can get familiar with the horizontal layout without significant learning costs, leading to performance improvement and better usability [89]. Additionally, another prior study of one-line text entry layouts [81] shows evidence that non-alphabetical 1-line layouts such as QWERTY-ordered and ENBUD are less effective than the alphabetical-ordered layout, unless the users take tremendous efforts to learn the new layout. Furthermore, novice users feel difficult to find the character on the one-line layouts of QWERTY and ENBUD when the text entry environment goes invisible for the sake of screen space-saving [90].

Here we describe the speed control of character movement, γ is the initial moving time of character, and its speed update depends on the length L of prefix (i.e., the number of typed characters in the textbox, namely, sub-string) in a complete word, with a discounted factor η. This indicates that the basic movement speed will increase when a sub-string with more characters results in a smaller number of next possible characters. Thus, the basic movement speed of the character follows $S = \frac{D_z}{\gamma-(L*\eta)}$. At time window K, the default movement speed is further increased by the depth position of the user's hand U_k outside the P-zones (i.e., F-zone or R-zone). The value of velocity S_j is calculated as $S_k = S + \frac{S}{\gamma}\int_0^{\gamma} U_k$, where γ is the division of sub-zones in either F-zone or Recall zone. In other words, the farther distance the hand position from the F-zone and R-zone, the more accelerated character movements for either the character selection or character deletion.

The Gesture of a Pointing Hand

It is important to note that the physical input devices (e.g., mouse and stylus) own advantage of highly precise pointing. In contrast, mid-air pointing gestures are coarse [70], and accordingly direct positioning operations on small and dense characters are tedious yet erroneous. The user with HIBEY maintains a pointing gesture, as a pointing token, and hovers between the three connected zones for selecting characters. HIBEY primarily has five transition states of continuous pointing gestures: (1) *no pointing gesture is detected*, (2) *pointing gesture is detected*, (3) *pointing gesture in the P-zone*, (4) *pointing gesture in the F-zone*, and (5) *pointing gesture in the R-zone*. The transitions (a–g) between states are described as follows: Hold (a): Holding a mid-air tap gesture to begin the text entry. Enter (b): The pointing hand shifts to another zone. Select (c): The pointing hand in the P-zone selects the neighboring characters. Facilitate (d): The pointing hand in the F-zone confirms the character selection. Recall (e): The pointing hand makes a backspace function in the R-zone. Flip (f): A large horizontal hand displacement can drop the operations to the chosen character key in either the F-zone or the R-zone. Release (g): The user's pointing hand cannot be located by the camera embedded on the smartglasses.

10.4.2 Uncertainty on Keyboard-Less Environment

A pilot study is conducted to investigate the user behavior of keyboard-less typing in the holographic environment. We intend to understand the performance of text entry in different visual conditions. Accordingly, we validate the feasibility of keyboard-less text entry in the holographic environment of HIBEY.

As shown in Fig. 10.7a, we assess our text entry layouts with various degrees of visibility, represented by visual cues as follows: (1) fully visible layout (top), (2) translucent keyboard with cross marks (mid), and (3) fully invisible layout with hidden character keys (bottom). The one-line layout is configured to enable the users to select the characters in fixed position for easy memorization. We recruited 15 participants from the local university campuses. The participants had no prior experience in mid-air text input. Four out of 15 participants had tried Microsoft Kinect application. None of them is a native English speaker but all are familiar with the alphabetical order. The experiment was conducted on a AR headset named Microsoft HoloLens. We gave instructions to the participants to input word phrases as accurately

Fig. 10.7 The three experimental interfaces (**a**) and the distribution of coordinates with the three experimental interfaces (**b**)

(but not soon) as possible, i.e., locate the character keys, without correction. The output of character entry in the textbox is represented by asterisks to avoid bias toward our keyboard design and to ensure the participants focus on the next character entry without significant adjustment of their character picking positions. The three layouts were tested in an alternative order. For each layout, the participants complete 5 sessions featuring 15 phrases from MacKenzie and Soukoreff phrase set and generate 2700 phrases in total. We manually balanced the word phrases to reduce the imbalance on the least frequent characters such as q, x, and z. For each character input, the x- and y-coordinates across the planar boundary between the P-zone and F-zone are recorded.

Figure 10.7b depicts the distribution of coordinates across the planar boundary. We organize the recorded coordinates of each character within 95% confidence as ellipses. The geometric centers of character keys are shown within the squares. The three distributions represent the three visibility conditions (i.e. visible, translucent, and invisible). We calculate the horizontal and vertical offset (in pixel) by the measured coordinates minus the geometric center of the character key (the center of the square).

Regarding the horizontal offset, ANOVA shows a significant effect of the visual feedback on the horizontal offset ($F_{2,69} = 209.448$, $p < 0.0001$), and pairwise comparison between each layout shows a significant difference ($p < 0.0001$). The average offsets for visual conditions (1), (2), and (3) are respectively 22.97 (std. = 16.31), 32.12 (std. = 26.90), and 40.41 (std. = 36.51). The offset for layout (3) (invisible character keys) is 75.93% larger than for the fully visible layout, while the standard deviation of the layout (3) is 138.12% greater than the fully visible layout. According to Fig. 10.7b, we observe that layouts (1), (2), and (3) respectively display an approximate offset length of 0–1, 1–2, and 2–3. For all three layouts, the common tendency is that the measured centers of the leftmost nine characters and rightmost nine characters are shifted to the center of the screen, while the middle nine characters show random centers of measured horizontal coordinates.

Regarding the vertical offset, ANOVA reveals a significant effect of the visual feedback on the vertical offset ($F_{2,69} = 446.891$, $p < 0.0001$). Pairwise comparison between each layout shows a significant difference ($p < 0.0001$). The average offsets for visual conditions (1), (2), and (3) are 11.28 (std. = 9.77), 15.54 (std. = 15.67), and 29.16 (std. = 29.76). The vertical offset between condition (1) and (2) shows only 37.74% and 60.31% difference in the values of average and standard deviation. As expected, the offset of the condition with invisible character keys is 158.49% larger than the fully visible condition. Similarly, the standard deviation aligns as the condition with invisible character keys is two times larger than the fully visible layout. We observe that users in condition (3) have a greater vertical movement, which aligns with our findings shown in Fig. 10.7b.

This pilot study investigates the user behaviors of text input under the keyboard-less condition. Under condition (3), the overlapping of x-coordinates across keys generally shows offsets of no bigger than two character keys. An offset of fixed size two to three characters is not considered because the performance of word disambiguation will be deteriorated [81]. Therefore, a probabilistic approach is employed to handle the uncertainty issue due to imprecise hand gestural pointing.

Probabilistic Method for Handling Imprecision

As discussed in [87], the statistical decoding can alleviate the imprecise hand gestural text input. Properly designed statistical decoders for touchscreen interfaces are supported by both the language model and spatial model [91, 92]. To reduce the computational burdens on the constrained smartglasses, the coordinate system is transformed from 3D into 2D, where the planar boundaries located at either the P-zone or F-zone act as an imaginary plane in the 3D environment, while the statistical decoder handles these "touchpoints" (coordinates) created by the continuous gesture of a pointing hand traversing this 2D plane.

As shown in Fig. 10.7, all three conditions have demonstrated consistent character offsets in terms of x-coordinations. However, y-coordinations in most of the characters have demonstrated random vertical fluctuations. The primary reason is that the characters have been configured in a connected layout of a horizontal line. Thus, the participants carefully select the character position as reflected in the x-coordinates but not the vertical positions as reflected in the y-coordinates. According to the x-coordinate inputs, Bayes' theorem estimates the likelihood of a word and suggests the most probable words in the suggestive word list. Given a set of 2D coordinates in mid-air $O = \{o_1, o_2, o_3, \ldots, o_n\}$, the decoder finds for the optimal word W_{Opt} inside the lexicon X satisfying

$$W_{\text{Opt}} = \arg\max_{\forall W \in X} P(W|O) \tag{10.3}$$

According to Bayes' rule, we have

$$W_{\text{Opt}} = \arg\max_{\forall W \in X} P(O|W)P(W) \tag{10.4}$$

where $P(W)$ and $P(O|W)$ are separately calculated by an existing language model [92] and the spatial model. Given that W consists of n symbols: $S = \{s_1, s_2, s_3, \ldots, s_n\}$, the spatial model estimates the $P(O|W)$ as follows:

$$P(O|W) = \prod_{i=1}^{n} P(o_i|s_i) \tag{10.5}$$

Prior research [93] proved that the character selection on 2D interfaces follows the model of bivariate Gaussian distribution. The x- and y-coordinates of o_i are x_i and y_i and hence,

$$P(o_i|s_i) = \frac{1}{2\pi\sigma_{i_x}\sigma_{i_y}\sqrt{1-\rho_i^2}}\exp[\frac{z}{2(1-\rho_i^2)}] \tag{10.6}$$

and

$$z \equiv \frac{(x_i - \sigma_{i_x})^2}{\sigma_{i_x}} - \frac{2\rho_i(x_i - \mu_{i_x})(y_i - \mu_{i_y})}{\sigma_{i_x}\sigma_{i_y}} + \frac{(y_i - \sigma_{i_y})^2}{\sigma_{i_y}} \tag{10.7}$$

where (μ_{i_x}, μ_{i_y}) is the geometrical center of the defined symbols s_i; σ_{i_x} and σ_{i_y} are the standard deviations of x- and y-coordinates for character key s_i; and ρ_i is the correlation value of the x- and y-coordinates. The collected user behaviors from the pilot study (the respective values of mean and standard deviation as $\mu_{i_x}, \mu_{i_y}, \sigma_{i_x}, \sigma_{i_y}$), as depicted in Fig. 10.7, are applied to the above equations to compute the most probable word W_{Opt} in the text input layout of various visible cues. Eventually, suggestive words are generated from the word disambiguation process in such keyboard-less configurations, and further listed in the top-k candidates under a particular word (sub-)string.

10.4.3 Implementation and User Performance

The pilot study helps us to understand the user's behaviors with the uncertainty of keyboard-less environment and imprecision in mid-air. Accordingly, HIBEY is implemented on a Microsoft HoloLens, in which a minimum interface

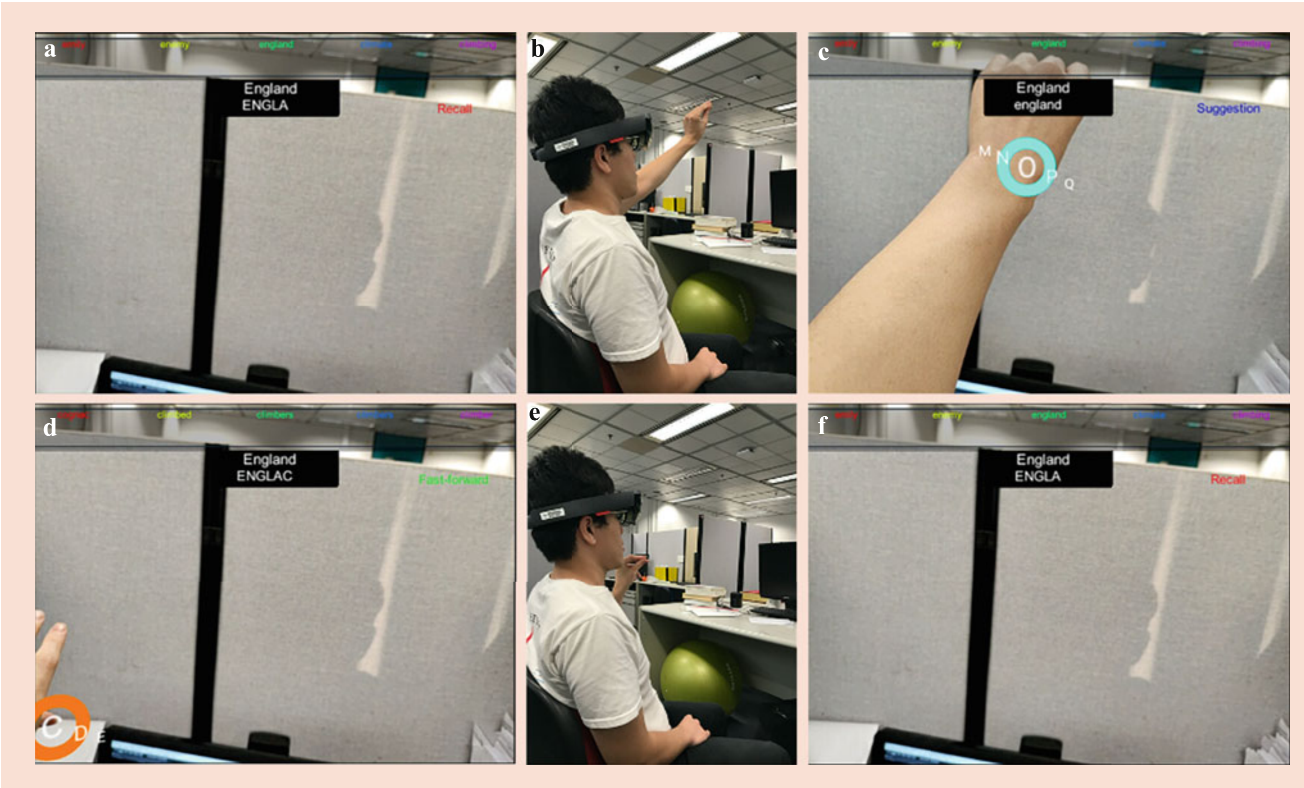


Fig. 10.8 Text entry illustration of a word "ENGLAND" (the circular panel of characters are invisible to the users, serving as illustrative purpose only): (**a**) Hand-off; (**b**) Starting mid-air tap; (**c**) Hold the hand gesture to select a character; (**d**) Mistakenly pick character "C"; (**e**) Move the hand gesture to recall zone for backspacing the character "C"; and (**f**) Removal of character "C"

with invisible character keys is presented to the users and the suggestive words is located at the upper edge of the small-size screen real estate. Figure 10.8 depicts the prototypical interfaces of HIBEY. The spectator mode with color hints is included to aid in explaining the character selection procedures through exposing the hidden character positions. In the normal use case, the characters are hidden to avoid cluttered AR cues and hence disturbance to the physical environments.

A text entry task aims to assess the system performance of HIBEY in terms of text entry speed and error rate. A total of 18 participants were involved in an 8-day text input tasks, and two text input conditions were studied: one-line (Visible, Baseline) and Non-key (Invisible, HIBEY). In the visible condition, the 27 character keys and the predicted words were shown to the participants, while the participants with invisible condition can only see the predicted words but not the hidden characters. The trials in each condition contain 25 word phrases extracted from the text material of MacKenzie and Soukoreff phrase set. The participants were told to type as fast and accurate as possible. The correction of typing mistake can be applied to the current word string only. The selection of a single character or the predicted words can both be applied in the holographic text entry interface.

Regarding text entry performance, the users are capable of performing text entry at the average rate of 11.01 WPM (std. = 2.29) with the one-line condition over the 8-day sessions. In contrast, the users are capable of perform text entry at the average rate 5.44 WPM (std. = 0.28) with the Non-key condition on the 1st day. The average rate of text entry on the 8th day increases by 142.3% to 13.19 WPM (std. =1.10). Regarding the error rate, one-line condition achieves a mean error rate of 2.51% (std. = 0.0038). Across the eight-session training, the participants improve from 2.91% (std. = 0.0032) in the first session to 2.21% (std. = 0.0033) in the final session. In comparison, the Non-key condition achieves a mean error rate of 3.94% (std. = 0.0049). As expected, the initial high error rate of 4.37% (std. = 0.0035) on the first day is mainly caused by the unfamiliarity of the layout of the hidden character keys. Throughout the second day and fifth day, we observe that the Non-key condition catches up the one-line condition. On the eighth day, the error rate of Non-key condition decreases to 3.47% (std. = 0.0032), as the participants are able to memorize the relative position of the hidden character keys.

On the other hand, we did the quantitative measure of screen real estate between HIBEY and default QWERTY

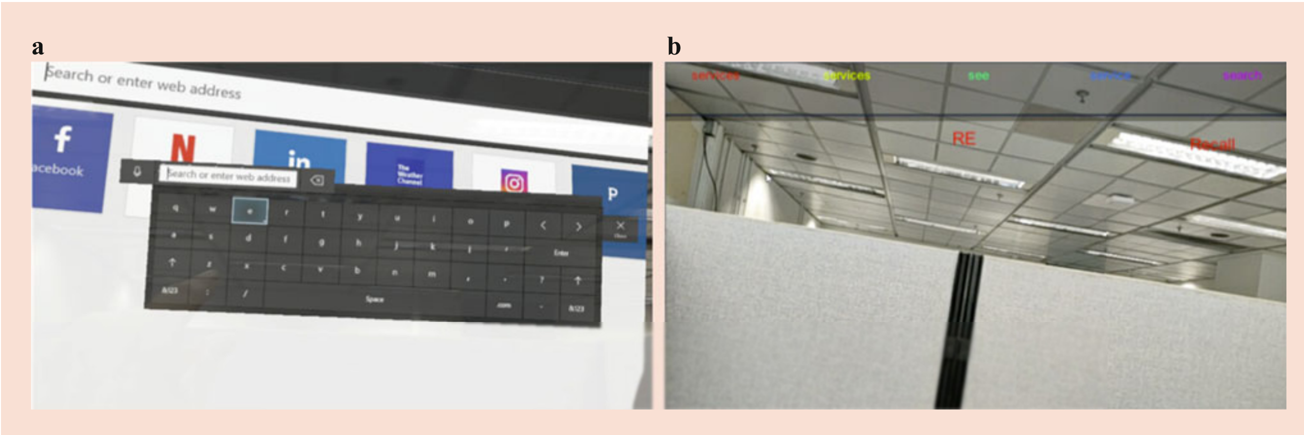


Fig. 10.9 Screen real estate between default keyboard layout (**a**) and HIBEY (**b**) on Microsoft HoloLens

keyboard on HoloLens (Fig. 10.9). It is obvious that HIBEY takes only 13.14% of screen area at the edge position, and the default QWERTY keyboard occupies 35.33% of screen area at the center position. The default QWERTY keyboard therefore needs 168.84% more space than HIBEY, and meanwhile HIBEY reserves the center position for user interactions with AR.

10.4.4 Application Scenarios

HIBEY shows a feasible demonstration of the invisible text entry layout on augmented reality smartglasses, with two prominent advantages at the blurred boundary between physical and digital environments as follows. First, HIBEY makes lower disturbance to the physical environment by removing the full QWERTY keyboard from the center screen area. Second, users with HIBEY only employ hand gestural control and no addendum sensors are involved. However, HIBEY poses several key limitations. First, the depth camera is employed to track the traversing hand across planar space between the zone boundaries (preparation/fast-forward/recall), which is costly and usually not available on the small-size augmented reality smartglasses (e.g., Google Glass). Second, the tracking sensitivity (i.e., coarse hand detection) on the pointing gesture may be unfavorable to the user performance, due to the limitations of camera sensitivity and the computation resources on AR smartglasses. If the user's hand makes a fast displacement from one character to another, to make a last-minute decision change in character selection is less feasible under the current camera configurations. Due to the camera sensitivity, more typing mistakes lead to unproductive operations in the R-zone for correcting the erroneous characters. In case of typing mistakes, users erasing the previous character in the recall zone results in unproductive times during the task. Despite the higher error rates than concurrent solutions, users manage to achieve higher typing speeds, thanks to the predictive word completion and the backspace function that allow to quickly correct typing mistakes.

For future works, the invisible text entry solution of HIBEY should be further improved by introducing deep learning techniques and more sophisticated language models. Besides, it is worthwhile to assess two additional visible/invisible conditions with magnified effects on the one-line character. Similar to the magnified application icons at the edge areas of the macOS screen, it is interesting to see any benefits of employing such magnified characters in either visible or invisible text entry layouts.

10.5 TOFI

The existing commercial solutions for text entry on augmented reality (AR) head-worn computer approaches are either socially awkward [4] or limited mobility [94], which become an obstacle to the popularity of head-worn AR computers in outdoors. Both the tangible interface on the frame of Google Glass and the mid-air gestures of Microsoft HoloLens require a hand lifting up to the eye level, which are not designed for continuous and long-term usage and catch unwanted attention from the adjacent people. In contrast, thumb-to-finger interaction has appealing properties such as discreet [4] and subtle [31] movements, comfortable yet accurate input [36], and no visual demand on the keyboard layout [95]. Nevertheless, designing text entry solution under the scenario of thumb-to-finger interaction is challenging, as the full QWERTY keyboard can barely accommodate into the small-size palm areas of two hands [7]. More importantly, two-handed thumb-to-finger interaction becomes a constraint to the user mobility as both hands are busy for user interaction, i.e., text entry.

TOFI serves as a single-handed text entry solution [16], which is designed particularly for text entry on AR headsets, where the one-handed text entry allows the users to reserve an unoccupied hand for grasping item such as shopping bags and umbrella in outdoors. Users with TOFI distinguish characters on an ambiguous layout by varying the force exerted on the keys. The design of TOFI makes each ambiguous key containing three characters. Force interaction serves as an alternative modality to disambiguate the target keys in the keypads. Thus, TOFI is able to shrink the full two-handed QWERTY keyboard into a space-saving layout for single-handed user interaction. The prominent features of TOFI are two, as follows. First, the glove-based user interaction is convenient, in which the force-assisted interaction is employed for one-handed text entry within the palm area of a single hand. Second, we optimized the keyboard layout on a prototypical glove to enhance both the text entry performance and user ergonomic, especially the easiness of force-assisted interaction.

10.5.1 System Implementation

This paragraph gives an overview of the implementation of our prototypical system and its components, and discusses the calibration and behavior of the system. The glove prototype system consists of three key components: the tactile surface of the glove, a 16-channel multiplexer, and an Arduino microcontroller. Three force sensors are allocated on the top of each of the fingers of the glove, equivalently one for each phalanx. The initial iteration of the prototype shown in Fig. 10.2 having pressure-sensitive resistors glued to a glove.

The sensors were created using small 5×5 mm squares of pressure-sensitive conductive sheets. This type of plastic was originally designed to protect items from electric discharges. To maintain constant pressure level over the plastic, the wires are connected using gaffer tapes. Other alternatives such as conductive glue lead to unpredictable fluctuations in the resistance of the material, leading to noisy measurements in our experimental setting. The sensors are mounted on the palm area using Velcro fasteners. One Velcro is employed on each phalanx, i.e., a total of 12 fasteners for 12 pressure-sensitive sensors on a single hand. This solution makes the initial setup more time-consuming, but allows customizing the system to each user during the experiment phase and easy yet swift replacement of worn-out sensors.

The sensors are connected to the 16-channel analog multiplexer that aggregates the measured pressures and sends them to the Arduino. The multiplexer converts the analog signals received to digital signals and serializes them to the Arduino. As a result, four pins are connected to the Arduino. The Arduino microcontroller has the entire logic to convert numeric measurements of force to key presses on the computer. As such, we utilize an Arduino Uno R3 offering the standard USB Human Interface Device (HID) profiles for mouse and keyboard.

The pressure sensing values on the prototypical glove are normalized between 0 and 1, where 0 corresponding to the minimum value for sensor activation and 1 being the maximum measured value. We further divide the continuous force spectrum range into three discrete options (0.000–0.299, 0.300–0.699, 0.699–1.000), representing "light tap," "shallow force touch," and "deep force touch." When a sensor's value exceeds the minimum threshold, the system awaits until that value drops below the minimum threshold and record the maximum value obtained in that time interval. When multiple sensors are activated during the movement, the system only considers the largest normalized value. Figure 10.10 depicts the process of force disambiguation for text entry, in which three force levels match with three characters for the character choices in the ambiguous key for the characters T, U, and F.

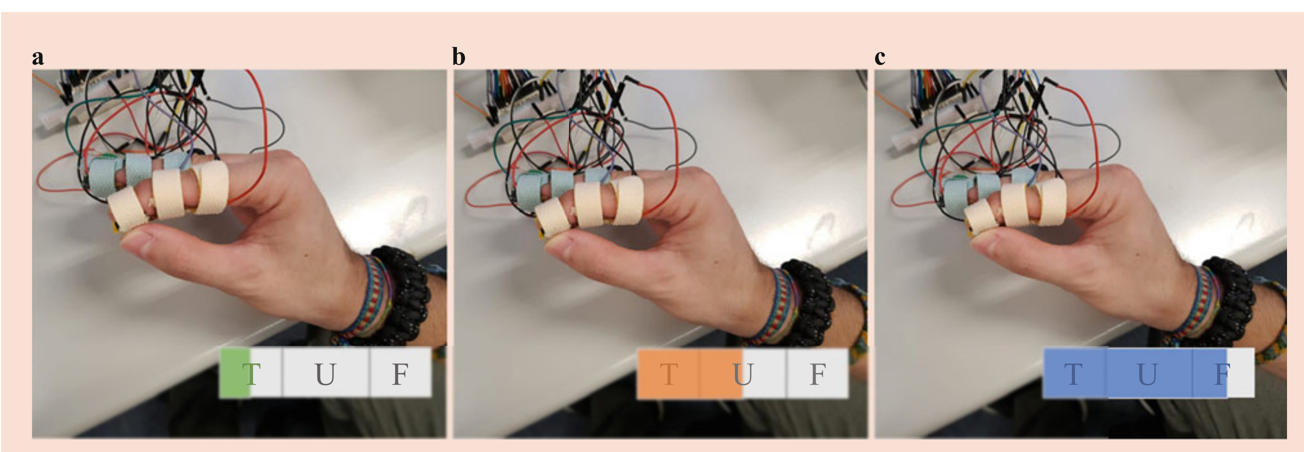


Fig. 10.10 Three force levels inside the ambiguous located in the first phalanx of the index finger: (**a**) light tap, (**b**) shallow force touch, and (**c**) deep force touch

10.5.2 Optimization of the Keyboard Layout

This section has addressed the erroneous nature of force-assisted interaction by designing a better keyboard layout error rate through the formulation of a quadmetric (i.e., four-metric) optimization problem. A previous study [36] suggests that human users can identify multiple buttons on their fingers including the *26 English alphabets*, *space*, *enter*, and *delete* keys within the palm area of a single hand, under the scenario of off-sight thumb interaction.

Our prior study [16] examined the design choices between three-level and four-level force-assisted thumb-to-finger interaction. Text entry is considered as a highly repetitive user interaction, and hence the accuracy of choosing the keys in an ambiguous key is important. Introducing additional levels can lead to significantly higher error rates. Thus, the three-level configuration is better than the four-level configuration. Although the three-level configuration results in a moderate accuracy level of 83.9%, the users in the prior study reflected a higher preference for the three-level configuration than for the four-level configuration. We therefore employ the three-level configuration as a balanced design choice in our optimization problem.

We therefore derive a keyboard of nine force touch buttons and three touch buttons, where the 26 alphabet characters are assigned to the three rows of three force touch keys (3-3-3) located on the index, middle, and ring fingers. The space, enter, and delete keys are assigned to the three touch buttons on the small finger. Our optimization problem allocates the 26 characters into the ambiguous force touch keys in the three-level design choice as above mentioned.

Our optimization problem has four metrics as follows: (1) the goodness of character pair and (2) the familiarity with the QWERTY layout facilitate the text entry performance, while (3) the easiness of force keypad interaction and (4) the comfort level of the finger space consider the human ability to input text. The optimization problem is solved by a genetic algorithm that generates the optimized 3-3-3 thumb-to-finger layout, striking a balance between the text input ability and the ergonomic constraints.

Maximizing the Goodness of Character Pair

We consider two important factors: (i) the finger movement distance and (ii) the ambiguity for auto-correction, for the goodness of a character pair in a candidate solution keyboard $k \in K$.

The frequencies $a_{i,j}$ of key pairs in English are calculated by counting the Bigram of character pairs in the selected English corpus. Let $\alpha = \{a,b,c,\ldots,x,y,z,\_\}$ be the alphabet, $Bigram_{i,j}$ denotes the probability of the finger moving from character i to character j. We established a character pairing matrix of size 676 (26*26), $Bigram_{i,j}$, by normalizing the frequencies as $Bigram_{i,j} = a_{i,j}/\sum_{\forall i,j\in\alpha} a_{i,j}$. It is worthwhile to mention that the source of the English corpus does not significantly impact the optimization of the keyboard layout [96].

Our analysis indicates that the most appeared character pair is *ER*, with 90,510,491,354 occurrences in our corpus. The probability of triggering finger movement between E and R is 3.91%. The next most common character pairs are IN with P_{IN} of 2.75%, ES with P_{ES} of 2.50%, TI with P_{TI} of 2.38%, TH with P_{TH} of 2.25%, and OR with P_{OR} of 2.24%. In contrast, the least probable character pair was QJ having a probability of 0.0000005. Next, we normalize the probability to a Bigram score, $Bi_{i,j}$ defined as follows: $Bi_{i,j} = \frac{100*Bigram_{i,j}}{\max(Bigram_{i,j})}$. For example, the character pairs *ER* and *IN* obtain 100.00 and 70.49 normalized scores, respectively.

Considering that the thumb-to-finger space interaction leads to eye-free text entry that is error-prone [97], the glove-based text entry solution heavily relies on the automatic correction algorithms. However, the auto-correction poses a key limitation on checking valid words. For instance, the words "hit" and "hat" are both valid. Therefore, the ambiguity issue emerges if the characters "A" and "I" are arranged together [73]. To alleviate the limitation, we also establish another matrix named *Badgrams* [73] recording the 676 (26*26) possible character pairs. All words of identical length in the corpus were scanned. Accordingly, the Badgram frequencies of the character pair, $b_{i,j}$, were counted by inspecting whether a character substitution in a string of the identical length produces an alternative word with valid status in the English corpus. Then, the frequencies are normalized as probabilities by $Badgram_{i,j} = b_{i,j}/\sum_{\forall i,j\in\alpha} b_{i,j}$. As a result, we list the top five Badgram, as follows: AE ($P_{\mathrm{AE}} = 0.0175$), AO ($P_{\mathrm{AO}} = 0.0146$), AI ($P_{\mathrm{AI}} = 0.0125$), EO ($P_{\mathrm{EO}} = 0.0121$), and EI ($P_{\mathrm{EI}} = 0.0118$). Putting A and E as an adjacent pair can result in numerous ambiguous words such as *end* instead of *and*, *he* instead of *ha*, *bet* instead of *bat*, and so on. Next, we normalize the probability of Badgram into the Badgram score by $Bad_{i,j} = 100 * Badgram_{i,j}/\max(Badgram_{i,j})$.

Finally, the satisfaction scores of the character pairs [74] is computed by $CP_{i,j} = Bi_{i,j} - Bad_{i,j}$. For combinations including the white space, we employed $CP_{i,j} = Bi_{i,j}$. We intend to search the most common character pairs with the least likelihoods of triggering the detection ambiguity under the limitation of the auto-correction solutions. Figure 10.11 shows the satisfaction scores of the character pairs, according to the computed Bigrams and Badgrams. We list the top eight candidate character pairs as follows: IN, NO, OR, AN, TH, TI, EN, and EH (descending order), while the bottom eight candidate character pairs are EA, OA, EO, AI, EI, LR, DS, and NR (ascending order).

We eventually calculate the goodness of the character pair configuration, M_{pair}, by aggregating the satisfaction score if the characters are adjacent, using the following objective:

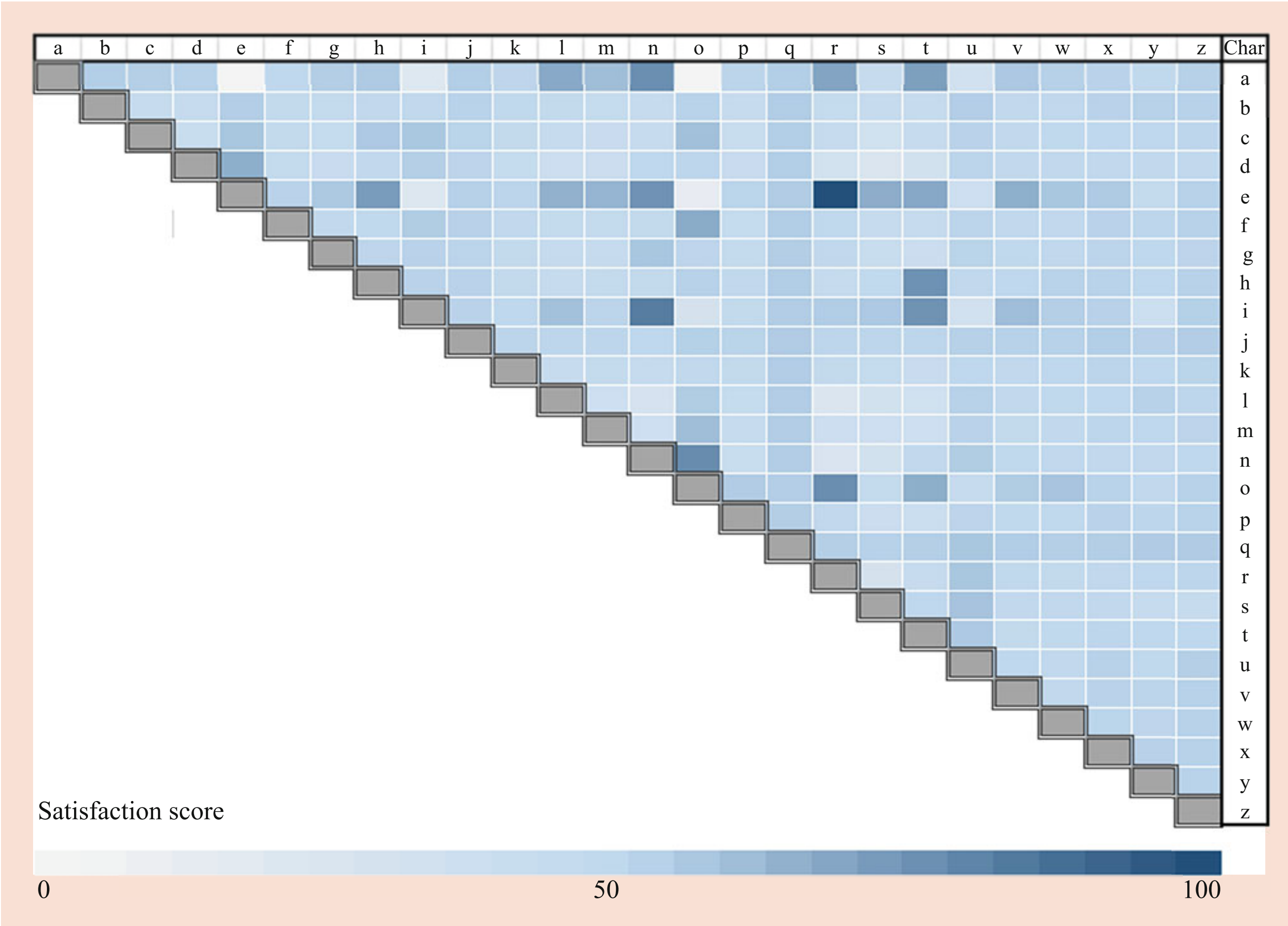


Fig. 10.11 Satisfaction score matrix of 26 × 26 character pairs

$$O_1 : M_{\text{pair}} = \sum_{\forall i,j \in \alpha} \begin{cases} CP_{i,j} & \text{if character } i \ \& \ j \text{ are adjacent keys} \\ 0 & \text{otherwise} \end{cases}$$

The M_{pair} is normalized into a comparable score among candidate solutions, M_{pair_n}, ranging from 0 to 1 by $M_{\text{pair}_n} = M_{\text{pair}}/1.1 * M_{\text{pair\_best}}$, where $M_{\text{pair\_best}}$ is defined as the keyboard of the highest score to the criterion. To facilitate the optimal search, we defined an offset index value of 1.1 to smooth the iteration of searching candidate layouts.

Maximizing the Familiarity with the QWERTY Layout

As the QWERTY layout is a *de facto* standard, the most computer and smartphone users are accustomed to the standard layout. Prior works show that users become more competent in QWERTY-like layouts than alphabetical layouts, due to the prominent advantage that QWERTY-like layouts alleviate the user's efforts on the visual search for characters [98]. The standard QWERTY layout, regardless of physical and virtual keyboards, has a trapezoid-shaped QWERTY configuration containing ten, nine, and eight characters on the top row, the middle row, and the bottom row, respectively. In our 3-3-3 QWERTY-like configuration, a minor modification has been made to fit all the characters, where each row contains nine characters.

We assess the familiarity between the candidate keyboard and a reference layout named **QWERTY starter layout**, through calculating the aggregated Euclidean distance of all the characters $i \in \alpha$ between the geometrical center of a given key in the candidate solution k_i and the center of the corresponding character position in the **QWERTY starter layout** s_i, as the reference of home positions. The character deviating from their home position in the candidate solution will be penalized. We assess the aggregated distance of a candidate solution, k_{dist}, in the candidate solution set K as

$$k_{\text{dist}} = \sum_{\forall i \in \alpha} \text{Distance}(k_i, s_i)^2$$

We obtain the normalized familiarity score as follows:

$$O_2 : M_{\text{familiarity}} = 1 - \frac{k_{\text{dist}}}{\max_{\forall k \in K}(k_{\text{dist}})}$$

The shorter the distance generated from the character alternation, the more the familiarity in this criterion.

Maximizing the Easiness of Force Keypad Interaction

In our prior study [16], the participants reflected reasonable user performance (i.e., the accuracy of character disambiguation) and user acceptance to the design configuration having three characters in one key, formulating the 3-3-3 keyboard configuration. The criterion considers the above user feedback as follows. As the light tap is more accessible than the deep force touch in the force-assisted keys, the most frequent character tier is assigned to the light tap force level, while the second most frequent character tier is assigned to the deep force touch force level. Finally, the least frequently used characters match with the shallow force touch. From our English corpus, the 26 characters, CF_i for all i,j $\in \alpha$ are sorted into three tiers based on their frequencies. We assign the characters into the force spectrum, as follows: $Tier_1$ [*t, a, s, c, i, p, o, b, m*], light tap; $Tier_2$ [*f, w, d, r, h, l, e, n, g*], deep force touch; and $Tier_3$ [*u, y, v, j, k, q, x, z*], shallow force touch. The characters of the $Tier_1$ link to 79.56% of the occurrence, and hence the most common interaction, known as light tap, enables most of the text entry. Prior works confirm that the force-assisted interaction is error-prone [77, 78]. Thus, the optimization criterion attempts to allocate the less frequently used character in the lower tiers to reduce the likelihoods of making erroneous user interaction.

The ease of use of our force keyboard is represented by the summed weighted score of character-tier matches. We establish a reward scheme to generate a score to those configurations if the characters in the candidate solution are assigned to the matched tiers (as listed above). The weighted score is summed by the CF_i and no normalization is needed. Our design encourages the most frequently used character to be matched with the user-preferred positions that are the two ends of the force spectrum. Accordingly, the user performance in terms of error rate will improve. We attempt to achieve a theoretical low error rate during the design phase for the force-assisted interaction. Thus, this optimization criterion is considered as a constraint, and hence the normalized value of the metric has the default value of 1.00.

$$O_3 : M_{\text{force}} = \sum_{\forall i \in \alpha} \begin{cases} CF_i & \text{if character } i \text{ in the matched tier} \\ 0 & \text{otherwise} \end{cases}$$

Maximizing the Comfort Level of the Finger Space

It is well known that the thumb owns the unique feature of reaching all the finger phalanxes within the palm space. The criterion addresses the issue that the individuals reflect various preferences to the phalanxes of the fingers from the thumb reach. We utilize the findings from an existing work [36], which supports our quantification of the comfort level of thumb-to-finger interaction.

We calculate the comfort level of the finger space by obtaining the summed product of all the characters $i \in \alpha$ and the button positions $z \in Z$ in the 3-3-3 configuration. And the comfort level is represented by $Comf_z$. The frequently used characters positioned in a comfortable position will be rewarded in the candidate solution k_i. The assessment of the candidate solution k_i in the candidate solution set K follows the below equation.

$$O_4 : M_{\text{Comfort}} = \sum_{\forall z \in Z} \sum_{\forall i \in \alpha} \begin{cases} CF_i * Comf_z & \text{if character } i \text{ in key } z \\ 0 & \text{otherwise} \end{cases}$$

Optimized Keyboard Layouts

Table 10.1 lists the performance of the layout in the perspective of quadmetric optimization. Among the layouts, the symbol "_" indicates a dummy character in the keyboard having 27 possible position as we only have 26 characters in the English dictionary. Our optimization model intends to select an ideal layout that addresses all the metrics and maximizes the overall score as shown in the rightmost column in Table 10.1. The keyboards named **Character Pair OPT Layout**, **QWERTY Starter Layout**, and **Finger Space OPT Layout** focused on a single criterion such as the goodness of character pair (Fig. 10.12a), the familiarity with the QWERTY layout (Fig. 10.12b), and the level of comfort in the finger space (Fig. 10.12c). It is important to note that the sole focus on one criterion can hurt the overall performance, i.e., the remaining three criteria are neglected, leading to sub-optimal layouts. The keyboards named **QWERTY OPT Layout** and **Finger Space OPT Layout** give excellent examples showing conflicting design, reflected by the full scores of individual criteria, i.e., the sub-optimal layout prioritizes only a single criterion without balancing the remaining criteria. The above two keyboard layouts reflect that satisfying the full score of the criterion for the familiarity with the QWERTY layout will remarkably hurt the criterion of comfort level within the finger space, and they work contrariwise. Nevertheless, the above sub-optimal

Table 10.1 Optimization results under the consideration of four criteria. The four criteria are O_1, O_2, O_3, O_4 where the higher the value of the normalized score between 0 and 4, the better the objective function has been achieved

Character layout	O_1	O_2	O_3	O_4	O_1+O_2+O_3+O_4
Character pair OPT layout (Fig. 10.12a)	1.00	0.38	1.00	0.53	2.91
QWERTY starter layout	0.61	1.00	0.15	0.64	2.40
QWERTY OPT layout (Fig. 10.12b)	0.73	1.00	1.00	0.52	3.25
Finger space OPT layout (Fig. 10.12c)	0.83	0.48	1.00	1.00	3.31
Quadmetric OPT (Fig. 10.12d)	**0.83**	**0.84**	**1.00**	**0.92**	**3.59**
Alphabetical constrained Layout	0.91	0.16	0.31	0.98	2.36

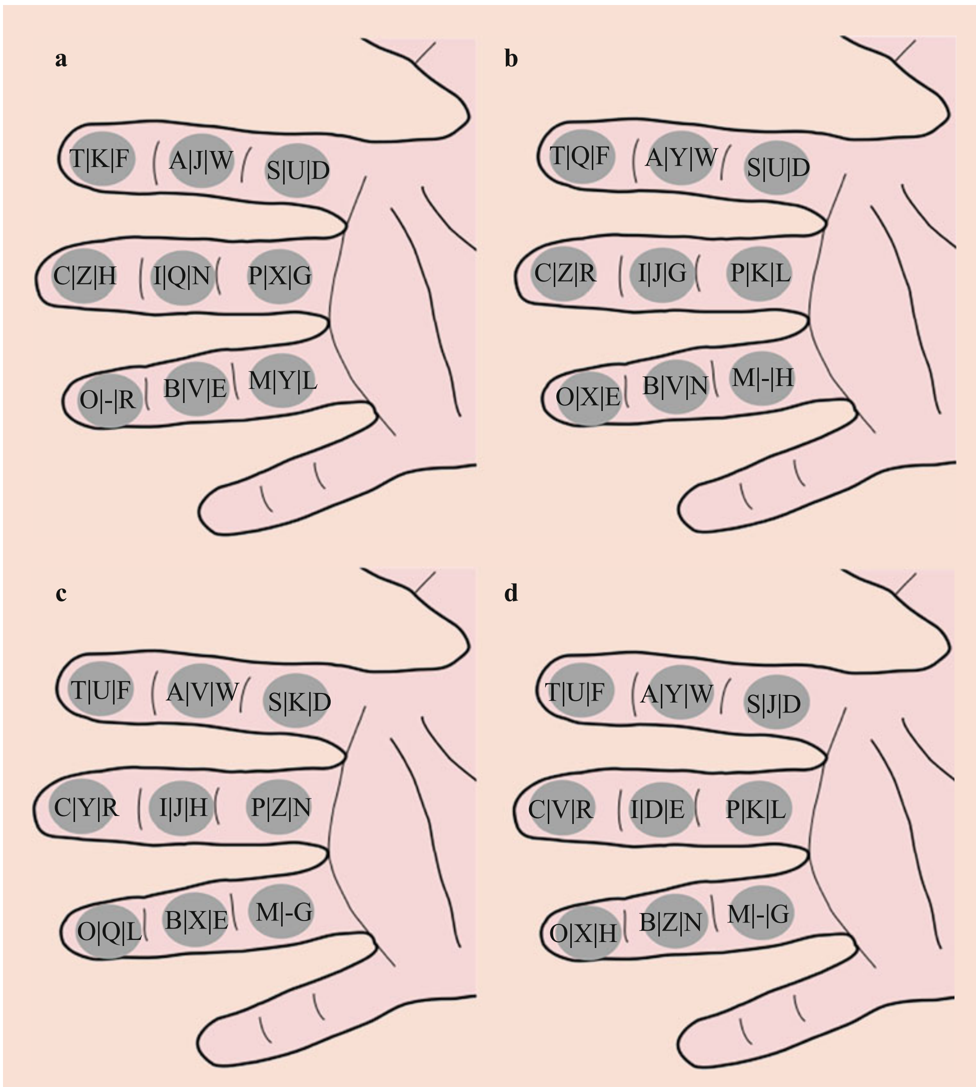


Fig. 10.12 Resultant layouts from the optimization problem: (**a**) The layout considering the goodness of character pair as top priority; (**b**) The layout considering the similarity to QWERTY keyboard as top priority; (**c**) The layout considering thumb-to-finger interaction as top priority; (**d**) The layout considering the four metrics in holistic view

keyboard layouts generate design cues on the feasibility of selecting a layout that balances all the four criteria. In the optimization model, the trade-offs among four criteria have been made incrementally. As a result, our model suggests a keyboard layout named **Quadmetric OPT**, i.e., the quadmetric optimized keyboard (Fig. 10.12d), which reaches an equilibrium between all the four criteria and achieves the highest scores among all the candidate keyboard layouts. The existing literature [36] suggests that the user interaction within the finger space is subject to the level of comfort within the thumb reaches to other fingers. In TOFI, broader considerations on the goodness of character pair (O_1) as well as the familiarity with the QWERTY layout (O_2) have been made in the text entry problem. As shown in our proposed optimization problem, the quadmetric optimized keyboard layout can reserve both the goodness of character pair and easiness of force keypad interaction. Although the optimized quadmetric layout sacrifices the goodness of character pair by 17.00% (0.83 in O_1), the normalized score for the familiarity with the QWERTY layout increases drastically by 121.05% from 0.38 to 0.84. Also, the normalized score for the comfort level of the finger space improves by 73.58% from 0.53 to 0.92. Therefore, the overall score of the quadmetric layout outperforms the keyboard named **Character Pair OPT Layout** by 23.37% through enabling a beneficial trade-off among the criteria of O_1, O_2, O_4. The quadmetric optimized layout also results in the scores 10.46% and 8.46% higher than the sub-optimal layouts named **QWERTY OPT Layout** and **Finger Space OPT Layout**, respectively. Not surprisingly, the alphabetically constrained layout and the QWERTY starter layout have made less satisfied scores than the aforementioned layouts due to their negligence of easiness for force keypad interaction.

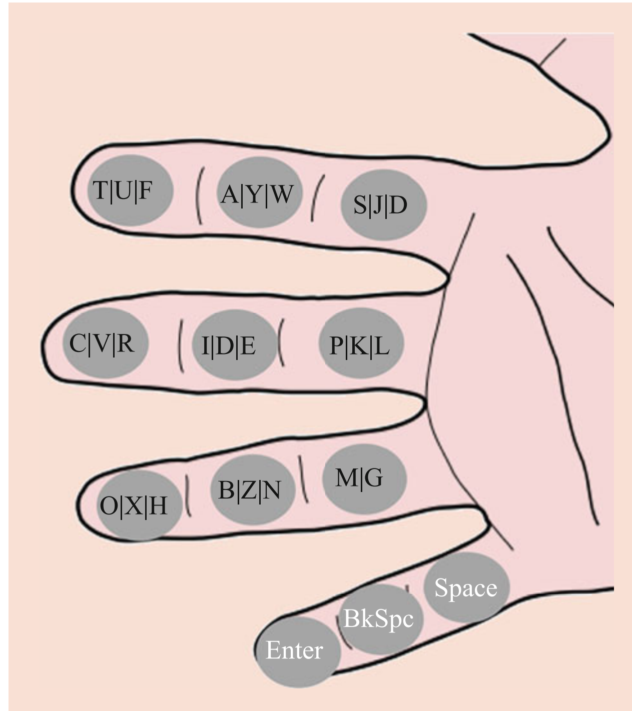


Fig. 10.13 The testing quadmetric optimized layout, where the ambiguous keys (English alphabets) are colored in black and the discrete keys (Enter, Backspace (BkSpc), and White Space (Space)) are colored in white

10.5.3 User Performance

Once we have confirmed the quadmetric optimized layout, the user performance of the layout is tested. As shown in the previous sections, the layout on the prototypical glove has been updated to the quadmetric optimized layout. Similarly, the ambiguous keys are distinguished by the force exertion between a normalized range from 0 to 1, where the force spectrum is divided on the basis of the three-level configuration, as follows: level 1, 0.000–0.299; level 2, 0.300–0.699; and level 3, 0.700–1.00. Three characters are contained in each key on the index, middle, and ring fingers. Also, the three keys on the small finger represent the space, enter, and backspace functions. Figure 10.13 depicts the layout as mentioned above on the prototypical glove. The force-assisted interaction on the glove employed the technique of quick release [77]. Users receive visual feedback displaying the force level in the three-level spectrum, and the appropriate level in the spectrum helps the users to get the target characters in the ambiguous key. In the evaluation, the quadmetric optimized layout on the prototypical glove had been tested with 5,400 characters (5 average characters in one word * 15 word phrases * 6 sessions * 12 participants).

Regarding the user performance, TOFI across the six-session evaluation achieved a mean text entry rate of 5.12 WPM (SD = 1.09) under the quadmetric optimized layout and 4.30 WPM (SD = 0.95) under the alphabetical layout, respectively. The text entry performance of the quadmetric optimized layout significantly improves from 3.53 WPM (SD = 0.17) in the first session to 6.47 WPM (SD = 0.53) in the final session, showing an 83.36% speed enhancement. Additionally, the error rate of the quadmetric optimized layout increases significantly from 14.88% (SD = 0.0058) in the first session to 5.64% (SD = 0.0120) in the final session. The quadmetric optimized layout achieves a mean error rate of 10.35% (SD = 0.0304). As the users were learning the unfamiliar layout (i.e., quadmetric optimized layout), the high error rate in the first session was unsurprisingly reflected. More importantly, the error rates decrease significantly in the latter sessions, as the most frequent characters are placed at the two ends of the three-level force spectrum (i.e., “light tap” and “deep force touch”), minimizing the likelihood of triggering the erroneous middle key (i.e., “shallow force touch”) inside the force spectrum. After the six-session training, the users are more familiar with the new layout and hence alleviate the erroneous inputs.

FingerT9 [99] ($\tilde{4}$.7 WPM) is a comparable single-handed text entry solution implemented on a glove. FingerT9 employs a disambiguation technique of multiple taps on a single key, instead of a single force-assisted tap in TOFI. The text entry rate of FingerT9 reaches 5.42 WPM in the final session, which is 16.23% slower than TOFI (6.47 WPM). The difference of text entry rates demonstrates that the force-assisted quadmetric optimized layout performs better than the multi-tap T9 layout, with the following reasons. The multi-tap T9 layout demands the user to do *no less than one* taps for the second- and third-level characters, while the force-assisted quadmetric optimized layout allows one single force-assisted touch to select the second- and third-level characters directly. For instance, the word phrase "FOOL" needs 12 taps with multi-tap T9 layout but only 4 taps with force-assisted layout (3 light tap and 1 deep force touch). Other longer words will worsen the situation with multi-tap T9 layout but not force-assisted layout. Moreover, the second- and third-level characters in multi-tap T9 can make more taps than expected. That is, the users sometimes over-tap on the key and do unnecessary taps for the correction of the mistakenly inputs. For example, when a character "B" is over-tapped as an unwanted character "C," the user requires to do additional taps (C – A – B) to get back to "B." leading to the degraded user performance.

At the end of the study (the sixth session), we conducted a mini-survey with all the participants and asked their intention of using the quadmetric optimized keyboard for the text entry task on augmented reality headsets. In a 10-point scale (1 meaning "definitely no" and 10 "definitely yes"), the majority of participants (seven out of ten) indicates positive responses to TOFI (more than five). The mean score of the question is 6.5 (SD = 2.4). Five participants reflected that the quadmetric optimized layout was a counterintuitive layout when comparing with the alphabetical layout at the beginning of the evaluation. Also, 7 out of 12 participants gave the comment of "the quadmetric optimized layout enables easier character entry than the alphabetical counterpart." Thus, the improved user acceptance of the quadmetric keyboard during the six-session period can be summarized as follows. At the beginning, the participants were not familiar with the quadmetric optimized layout, but the alphabetical constrained layout is deeply rooted in the mind of the participants. After the six-session practice, they felt more confident with the new layout and achieved more faster yet reliable text entry.

10.5.4 Application Scenarios

Between the studies of two-handed text entry systems leveraging the thumb interaction within the finger space, PalmType [31] (1.74%) utilized a highly precise motion tracking sensor in a controlled laboratory environment. However, the sensor setup may not be suitable for both mobility and scalability. Digitouch [7] (15.8%) leads to a significantly higher error rate than TOFI. The primary reason is that the users with Digitouch interact on a continuous rectangle-shaped strip of conductive fabric on each finger for character selection. Accordingly, the keyboard becomes highly ambiguous, as the tapping positions are converted to character selection through computing its relative position on the strip. Again, the text entry system of two-handed gloves deteriorates the mobility.

In comparison, TOFI has the advantages from the feedback channel of human proprioception as an additional perception mechanism [1]. The keypads within the finger space of TOFI on the glove can achieve accurately disambiguation among the ambiguous keys once the users are familiar with the quadmetric optimized layout. Also, TOFI features with one-handed and off-sight text entry, which facilitates the user interaction in mobile scenarios. For example, one hand can hold a handrail in commuting, and simultaneously the visual attention is preserved for the dynamic physical environment and other digital contents in AR. TOFI encourages users to do text entry through a series of small and unnoticeable finger movements in off-hand postures. Other systems lead to obvious and large movements such as lifting up the user's arm [100], waving the arm [69], and constantly tapping the touchscreen on wristbands [101], which can be viewed as socially awkward [4].

10.6 Take-Home Message for AR Interaction Techniques

In this chapter, we argued that by addressing the constrained interfaces on wearable computers of the augmented reality smartglasses, redesigning interaction techniques for target acquisition and text entry are conducive to their wide adoption. The three most recent works supported that embodied interaction can serve as a promising strategy of leveraging the resources on human users to overcome these constraints on wearable computers. TiPoint, HIBEY, and TOFI considered the unique resources on our body including finger dexterity, the memory of lexicographical order, proprioception, and opposable thumb, and demonstrated the corresponding interaction techniques and user interfaces on augmented reality smartglasses. Figure 10.14 depicts the mapping of the three discussed works to the task acquisition and text entry tasks on augmented reality smartglasses.

The three discussed works can map into the gap of the existing input options on augmented reality smartglasses. As the nowadays commercial standards have not addressed the target acquisition and text entry comprehensively, these works can seize this opportunity for commercialization. We summarized the highlights of the three discussed works as follows.

10.6.1 Conclusions

1. **TiPoint**: Smartglasses barely reserved the hardware interfaces such as touchpad and buttons, which are cumbersome and often counterintuitive to use. To overcome these

Fig. 10.14 The usage scenarios of the three discussed works on the augmented reality smartglasses

issues, TiPoint, a freehand mid-air interaction technique, leverages the camera mounted on the smartglass to detect the user's hand without relying on gloves, markers, or sensors, enabling intuitive and nonintrusive interaction. We introduce a computationally fast and lightweight algorithm for fingertip detection, which is especially suited for the limited hardware specifications and short battery lifetime of smartglasses. TiPoint processes pictures at real time with a high detection accuracy. Our evaluation and user study show that TiPoint as a mid-air nonintrusive interface delivers a better experience for interaction between users and smartglasses, with users completing typical tasks 1.82 times faster than when using the original hardware.

2. **HIBEY**: Text input is a very challenging task in augmented reality (AR). On non-touch AR headsets such as Microsoft HoloLens, virtual keyboards are counterintuitive and character keys are hard to locate inside the constrained screen real estate. We present the design, implementation, and evaluation of HIBEY, a text input system for smartglasses. HIBEY serves as a fast, reliable, affordable, and easy-to-use text entry solution through vision-based freehand interactions. Supported by a probabilistic spatial model and a language model, a three-level holographic environment enables users to apply fast and continuous hand gesture to pick characters and predictive words in a keyboard-less interface. Through the thorough evaluations lasting 8 days, we show that HIBEY leads to a mean text entry rate of 9.95 word per minute (WPM) with 96.06% accuracy, which is comparable to other state-of-the-art approaches. After 8 days, participants can achieve an average of 13.19 WPM. In addition, HIBEY only occupies 13.14% of the screen real estate at the edge region, which is 62.80% smaller than the default keyboard layout on Microsoft HoloLens.
3. **TOFI**: Augmented reality head-worn computers often feature small-sized touch interfaces that complicate interaction with contents, provide insufficient space for comfortable text input, and can be awkward to use in social situations. We present a novel one-handed thumb-to-finger text entry solution for augmented reality head-worn computers. A glove composed of 12 force-sensitive nodes featuring an ambiguous keyboard layout is designed and implemented. We first explore the viability of force disambiguation to evaluate the force division within the force spectrum. We select a three-level force division as it allows to considerably reduce the number of keys while featuring a high (83.9%) accuracy. Following this pilot study, we map the 26 English characters onto the nine nodes located on the index, middle, and ring fingers in a 3-3-3 configuration, and attribute the space, enter, and backspace keys to the remaining three nodes. We consider text entry performance as a quadmetric optimization problem considering the following criteria: goodness of character pairs, layout similarity to the QWERTY keyboard, easiness of force interaction, and comfort level of thumb reach. The resulting layout strikes a balance between performance and usability. We finally evaluate

the quadmetric optimized layout over six sessions with 12 participants. The participants achieve an average text entry rate of 6.47 WPM with 6.85% error rate in the final session, which is significantly faster than existing thumb-to-finger solutions. In addition, our one-handed text entry system enhances the user mobility compared to other state-of-the-art solutions by freeing one hand while allowing the user to direct his visual attention to other activities.

10.6.2 Future Outlook: Toward the Miniature and Subtle Interfaces

We conclude this chapter by extrapolating the shrinking sized wearable computers into a future in which the mass population uses the head-worn computers to handle the tasks being regularly managed on personal computers and smartphones. The shrinking size of wearable computers is facilitated by the advent of the high-speed network, as the resource-hungry applications can be off-loaded to the cloud servers. We envision that the size of wearable computers will be even smaller such as a coin-sized interface attached on the body. Minimum size of interfaces will be reserved for instant interactions as noticeable latency will deteriorate the user experience. For example, a finger-worn device can accommodate the proposed text acquisition technique of DupleFR [102] that demonstrates the feasibility of a miniature interface as small as the users' thumbs. Alternatively, an ambient camera can apply to detect the user fingertip through TiPoint or arm movements through HIBEY. Therefore, a finger-worn computing device or an ambient camera can serve as a token to communicate with streamed digital contents from the cloud and edge servers shown by the head-worn display, which enable the users to interact with the contents through direct manipulation (tap and select) as well as text entry.

Acknowledgments This research has been supported in part by project 16214817 from the Research Grants Council of Hong Kong, and the 5GEAR project from the Academy of Finland ICT 2023 programme.

References

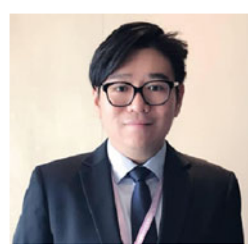
1. Lee, L., Hui, P.: Interaction methods for smart glasses: a survey. IEEE Access **6**, 28712–28732 (2018)
2. Singh, G., Delamare, W., Irani, P.: D-swime: a design space for smartwatch interaction techniques supporting mobility and encumbrance. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pp. 634:1–634:13, Montreal QC, Canada. ACM, New York (2018)
3. Blattgerste, J., Strenge, B., Renner, P., Pfeiffer, T., Essig, K.: Comparing conventional and augmented reality instructions for manual assembly tasks. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '17, pp 75–82, Island of Rhodes, Greece. ACM, New York (2017)
4. Hsieh, Y.T., Jylhä, A., Orso, V., Gamberini, L., Jacucci, G.: Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 4203–4215, San Jose, California, USA. ACM, New York (2016)
5. Bispham, M.: Security vulnerabilities in speech recognition systems. Technical report, 2016
6. Moggridge, B.: Designing Interactions. The MIT Press, Cambridge (2006)
7. Whitmire, E., Jain, M., Jain, D., Nelson, G., Karkar, R., Patel, S., Goel, M.: Digitouch: Reconfigurable Thumb-to-Finger Input and Text Entry on Head-Mounted Displays, vol. 1, pp. 113:1–113:21, Maui, Hawaii, USA. ACM, New York (2017)
8. O'Sullivan, D., Igoe, T.: Physical Computing: Sensing and Controlling the Physical World with Computers. Course Technology Press, Boston (2004)
9. Leigh, S.-W., Sareen, H., Kao, H.-L.-C., Liu, X., Maes, P.: Body-borne computers as extensions of self. Computers **6**(1) (2017)
10. Friston, K.J., Daunizeau, J., Kilner, J., Kiebel, S.J.: Action and behavior: a free-energy formulation. Biol. Cybern. **102**(3), 227–260 (2010)
11. Jacob, R.J., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: Reality-based interaction: a framework for post-wimp interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, pp. 201–210, Florence, Italy. ACM, New York (2008)
12. Garg, A.B.: Embodied cognition, human computer interaction, and application areas. In: Kim, T.-H., Mohammed, S., Ramos, C., Abawajy, J., Kang, B.-H., Ślęzak, D. (eds.) Computer Applications for Web, Human Computer Interaction, Signal and Image Processing, and Pattern Recognition, pp. 369–374, Berlin, Heidelberg. Springer, Berlin (2012)
13. Dourish, P.: "Being-in-the-World": Embodied Interaction. MITP (2004)
14. Lee, L.H., Braud, T., Bijarbooneh, F.H., Hui, P.:. Tipoint: detecting fingertip for mid-air interaction on computational resource constrained smartglasses. In: Proceedings of the 23rd International Symposium on Wearable Computers, ISWC '19, pp. 118–122, New York, NY, USA. ACM, New York (2019)
15. Yau, Y.P., Braud, T., Lee, L.-H., Lam, K.Y., Hui, P.: Hibey: Hide the keyboard in augmented reality. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications, PERCOM '19, Kyoto, Japan. IEEE (2019)
16. Lee, L.H., Lam, K.Y., Li, T., Braud, T., Su, X., Hui, P.: Quadmetric optimized thumb-to-finger interaction for force assisted one-handed text entry on mobile headsets. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **3**(3), 94:1–94:27 (2019)
17. Tung, Y.-C., Hsu, C.-Y., Wang, H.-Y., Chyou, S., Lin, J.-W., Wu, P.-J., Valstar, A., Chen, M.Y.: User-defined game input for smart glasses in public space. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pp. 3327–3336, Seoul, Republic of Korea. ACM, New York (2015)
18. Grossman, T., Chen, X.A., Fitzmaurice, G.: Typing on glasses: adapting text entry to smart eyewear. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '15, pp. 144–152, Copenhagen, Denmark. ACM, New York (2015)
19. Kienzle, W., Hinckley, K.: Lightring: always-available 2d input on any surface. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, UIST '14, pp. 157–160, Honolulu, Hawaii, USA. ACM, New York (2014)
20. Yang, X.-D., Grossman, T., Wigdor, D., Fitzmaurice, G.: Magic finger: always-available input through finger instrumentation. In:

Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, UIST ’12, pp. 147–156, Cambridge, Massachusetts, USA. ACM, New York (2012)

21. Ogata, M., Sugiura, Y., Osawa, H., Imai, M.: iring: intelligent ring using infrared reflection. In: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, UIST ’12, pp. 131–136, Cambridge, Massachusetts, USA. ACM, New York (2012)
22. Ashbrook, D., Baudisch, P., White, S.: Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’11, pp. 2043–2046, Vancouver, BC, Canada. ACM, New York (2011)
23. Rekimoto, J.: Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In: Proceedings of the 5th IEEE International Symposium on Wearable Computers, ISWC ’01, pp. 21–, Zurich, Switzerland. IEEE Computer Society (2001)
24. Ham, J., Hong, J., Jang, Y., Ko, S.H., Woo, W.: Smart Wristband: Touch-and-Motion–Tracking Wearable 3D Input Device for Smart Glasses, pp. 109–118. Springer, Berlin (2014)
25. Schneegass, S., Voit, A.: Gesturesleeve: using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches. In: Proceedings of the 2016 ACM International Symposium on Wearable Computers, ISWC ’16, pp. 108–115, Heidelberg, Germany. ACM, New York (2016)
26. Dobbelstein, D., Hock, P., Rukzio, E.: Belt: an unobtrusive touch input device for head-worn displays. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI ’15, pp. 2135–2138, Seoul, Republic of Korea. ACM, New York (2015)
27. Gustafson, S.G., Rabe, B., Baudisch, P.M.: Understanding palm-based imaginary interfaces: the role of visual and tactile cues when browsing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’13, pp. 889–898, Paris, France. ACM, New York (2013)
28. Harrison, C., Benko, H., Wilson, A.D.: Omnitouch: wearable multitouch interaction everywhere. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST ’11, pp. 441–450, Santa Barbara, California, USA. ACM, New York (2011)
29. Harrison, C., Tan, D., Morris, D.: Skinput: appropriating the skin as an interactive canvas. Commun. ACM **54**(8), 111–118 (2011)
30. Weigel, M., Lu, T., Bailly, G., Oulasvirta, A., Majidi, C., Steimle, J.: iskin: flexible, stretchable and visually customizable on-body touch sensors for mobile computing. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI ’15, pp. 2991–3000, Seoul, Republic of Korea. ACM, New York (2015)
31. Wang, C.-Y., Chu, W.-C., Chiu, P.-T., Hsiu, M.-C., Chiang, Y.-H., Chen, M.Y.: Palmtype: using palms as keyboards for smart glasses. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI ’15, pp. 153–160, Copenhagen, Denmark. ACM, New York (2015)
32. Wang, C.-Y., Hsiu, M.-C., Chiu, P.-T., Chang, C.-H., Chan, L., Chen, B.-Y., Chen, M.Y.: Palmgesture: using palms as gesture interfaces for eyes-free input. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI ’15, pp. 217–226, Copenhagen, Denmark. ACM, New York (2015)
33. Azai, T., Ogawa, S., Otsuki, M., Shibata, F., Kimura, A.: Selection and manipulation methods for a menu widget on the human forearm. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA ’17, pp. 357–360, Denver, Colorado, USA. ACM, New York (2017)
34. Ogata, M., Sugiura, Y., Makino, Y., Inami, M., Imai, M.: Senskin: adapting skin as a soft interface. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST ’13, pp. 539–544, St. Andrews, Scotland, United Kingdom. ACM, New York (2013)
35. Lin, S.-Y., Su, C.-H., Cheng, K.-Y., Liang, R.-H., Kuo, T.-H., Chen, B.-Y.: Pub-point upon body: Exploring eyes-free interaction and methods on an arm. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST ’11, pp. 481–488, Santa Barbara, California, USA. ACM, New York (2011)
36. Huang, D.-Y., Chan, L., Yang, S., Wang, F., Liang, R.-H., Yang, D.-N., Hung, Y.-P., Chen, B.-Y.: Digitspace: designing thumb-to-fingers touch interfaces for one-handed and eyes-free interactions. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI ’16, pp. 1526–1537, San Jose, California, USA. ACM, New York (2016)
37. Yoon, S.H., Huo, K., Nguyen, V.P., Ramani, K.: Timmi: finger-worn textile input device with multimodal sensing in mobile interaction. In: Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, TEI ’15, pp. 269–272, Stanford, California, USA. ACM, New York (2015)
38. Serrano, M., Ens, B.M., Irani, P.P.: Exploring the use of hand-to-face input for interacting with head-worn displays. In: Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI ’14, pp. 3181–3190, Toronto, Ontario, Canada. ACM, New York (2014)
39. Lissermann, R., Huber, J., Hadjakos, A., Mühlhäuser, M.: Earput: augmenting behind-the-ear devices for ear-based interaction. In: CHI ’13 Extended Abstracts on Human Factors in Computing Systems, CHI EA ’13, pp. 1323–1328, Paris, France. ACM, New York (2013)
40. Jones, E., Alexander, J., Andreou, A., Irani, P., Subramanian, S.: Gestext: accelerometer-based gestural text-entry systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’10, pp. 2173–2182, Atlanta, Georgia, USA. ACM, New York (2010)
41. Shimizu, J., Chernyshov, G.: Eye movement interactions in google cardboard using a low cost EOG setup. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, UbiComp ’16, pp. 1773–1776, Heidelberg, Germany. ACM, New York (2016)
42. Schuchert, T., Voth, S., Baumgarten, J.: Sensing visual attention using an interactive bidirectional HMD. In: Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction, Gaze-In ’12, pp. 16:1–16:3, Santa Monica, California. ACM, New York (2012)
43. Saponas, T.S., Kelly, D., Parviz, B.A., Tan, D.S.: Optically sensing tongue gestures for computer input. In: Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology, UIST ’09, pp. 177–180, Victoria, BC, Canada. ACM, New York (2009)
44. van Krevelen, D.W.F., Poelman, R.: A survey of augmented reality technologies, applications and limitations. Int. J. Virtual Reality **9**(2), 1–20 (2010)
45. Aigner, R., et al.: Understanding mid-air hand gestures: a study of human preferences in usage of gesture types for hci. Technical report, November 2012
46. Haque, F., Nancel, M., Vogel, D.: Myopoint: pointing and clicking using forearm mounted electromyography and inertial motion sensors. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI ’15, pp. 3653–3656, Seoul, Republic of Korea. ACM, New York (2015)
47. Hsieh, Y.-T., Jylhä, A., Jacucci, G.: Pointing and Selecting with Tactile Glove in 3D Environment, pp. 133–137. Springer, Berlin (2014)

48. Rosenberg, R., Slater, M.: The chording glove: a glove-based text input device. Trans. Sys. Man Cyber Part C **29**(2), 186–191 (1999)
49. Markussen, A., Jakobsen, M.R. and Hornbæk, K.: Vulture: a mid-air word-gesture keyboard. In: Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, pp. 1073–1082, Toronto, Ontario, Canada. ACM, New York (2014)
50. Chan, L., Chen, Y.-L., Hsieh, C.-H., Liang, R.-H., Chen, B.Y.: Cyclopsring: enabling whole-hand and context-aware interactions through a fisheye ring. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, UIST '15, pp. 549–556, Charlotte, NC, USA. ACM, New York (2015)
51. Huang, Z., Li, W., Hui, P.: Ubii: towards seamless interaction between digital and physical worlds. In: Proceedings of the 23rd ACM International Conference on Multimedia, MM '15, pp. 341–350, Brisbane, Australia. ACM, New York (2015)
52. Lee, T., Hollerer, T.: Handy ar: markerless inspection of augmented reality objects using fingertip tracking. In: 2007 11th IEEE International Symposium on Wearable Computers, pp. 83–90, Boston, MA, USA (2007)
53. Bailly, G., Müller, J., Rohs, M., Wigdor, D. and Kratz, S.: Shoesense: a new perspective on gestural interaction and wearable applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, pp. 1239–1248, Austin, Texas, USA. ACM, New York (2012)
54. Hollerer, T., et al.: Vision-based interfaces for mobility. In: The First Annual International Conference on Mobile and Ubiquitous Systems, vol. 00, pp. 86–94, Los Alamitos, CA, USA. IEEE Computer Society (2004)
55. Guimbretière, F., Nguyen, C.: Bimanual marking menu for near surface interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, pp. 825–828, Austin, Texas, USA. ACM, New York (2012)
56. Ha, T., Feiner, S., Woo, W.: Wearhand: Head-worn, rgb-d camera-based, bare-hand user interface with visually enhanced depth perception. In: 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 219–228 (2014)
57. Heun, V., Kasahara, S., Maes, P.: Smarter objects: Using ar technology to program physical objects and their interactions. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI EA '13, pp. 2939–2942, Paris, France. ACM, New York (2013)
58. Pino, A., Tzemis, E., Ioannou, N., Kouroupetroglou, G.: Using kinect for 2d and 3d pointing tasks: performance evaluation. In: International Conference on Human-Computer Interaction, pp. 358–367, Las Vegas, NV, USA. Springer, Berlin (2013)
59. Ren, G.,O'Neill, E.: Freehand gestural text entry for interactive tv. In: Proceedings of the 11th European Conference on Interactive TV and Video, EuroITV '13, pp. 121–130, Como, Italy. ACM, New York (2013)
60. Hamette, P., Lukowicz, P., Troster, G., Svoboda, T.: Fingermouse: A wearable hand tracking system. In: Adjunct Proceedings of the 4th International Conference on Ubiquitous Computing, Tokyo, Japan (2002)
61. Huang, Y., Liu, X., Zhang, X., Jin, L.: A pointing gesture based egocentric interaction system: dataset, approach and application. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 370–377, Las Vegas, USA. IEEE Computer Society (2016)
62. Simion, G., David, C., Gui, V., Caleanu, C.D.: Fingertip-based real time tracking and gesture recognition for natural user interfaces. Acta Polytech. Hung. **13**, 189–204 (2016)
63. Howell Istance, Richard Bates, Aulikki Hyrskykari, and Stephen Vickers. Snap clutch, a moded approach to solving the midas touch problem. In: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, vol. 38, pp. 221–228, Savannah, Georgia, USA (2008)
64. Ren, G., O'Neill, E.: 3d selection with freehand gesture. Comput. Graph. **37**(3), 101–120 (2013)
65. Heo, G., Lee, D.W., Shin, H.C., Jeong, H.-T., Yoo, T.-W.: Hand segmentation and fingertip detection for interfacing of stereo vision-based smart glasses. In: 2015 IEEE International Conference on Consumer Electronics (ICCE), pp. 585–586, Taipei, Taiwan (2015)
66. Li, S., Ashok, A., Zhang, Y., Xu, C., Lindqvist, J., Gruteser, M.: Whose move is it anyway? Authenticating smart wearable devices using unique head movement patterns. In: 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 1–9, Sydney, Australia (2016)
67. Kollee, B., Kratz, S., Dunnigan, A.: Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing. In: Proceedings of the 2Nd ACM Symposium on Spatial User Interaction, SUI '14, pp. 40–49, Honolulu, Hawaii, USA (2014)
68. Myers, B.A.: A brief history of human-computer interaction technology. Interactions **5**(2), 44–54 (1998)
69. Nirjon, S., Gummeson, J., Gelb, D., Kim, K.H.: Typingring: a wearable ring platform for text input. In: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '15, pp. 227–239, Florence, Italy (2015)
70. Ens, B., Byagowi, A., Han, T., Hincapié-Ramos, J.D. and Irani, P.: Combining ring input with hand tracking for precise, natural interaction with spatial analytic interfaces. In: Proceedings of the 2016 Symposium on Spatial User Interaction, SUI '16, pp. 99–102, Tokyo, Japan (2016)
71. McCall, R., Martin, B., Popleteev, A., Louveton, N., Engel, T.: Text entry on smart glasses. In: 2015 8th International Conference on Human System Interaction (HSI), pp. 195–200, Warsaw, Poland, (2015)
72. Yi, X., Yu, C., Zhang, M., Gao, S., Sun, K., Shi, Y.: Atk: enabling ten-finger freehand typing in air based on 3d hand tracking data. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, UIST '15, pp. 539–548, Charlotte, NC, USA (2015)
73. Dunlop, M.D., Durga, N., Motaparti, S., Dona, P., Medapuram, V.: Qwerth: an optimized semi-ambiguous keyboard design. In: Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services Companion, MobileHCI '12, pp. 23–28, San Francisco, California, USA. ACM, New York (2012)
74. Dunlop, M., Levine, J.: Multidimensional pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, pp. 2669–2678, Austin, Texas, USA. ACM, New York (2012)
75. Smith, B.A., Bi, X., Zhai, S.: Optimizing touchscreen keyboards for gesture typing. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, pages 3365–3374, Seoul, Republic of Korea. ACM, New York (2015)
76. Gong, J., Xu, Z., Guo, Q., Seyed, T., Chen, X.A., Bi, X., Yang, X.D.: Wristext: one-handed text entry on smartwatch using wrist gestures. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, pages 181:1–181:14, Montreal QC, Canada. ACM (2018)
77. Mizobuchi, S., Terasaki, S., Keski-Jaskari, T., Nousiainen, J., Ryynanen, M., Silfverberg, M.: Making an impression: force-controlled pen input for handheld devices. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems, CHI EA '05, pp. 1661–1664, Portland, OR, USA. ACM, New York (2005)

78. Wilson, G., Stewart, C., Brewster, S.A.: Pressure-based menu selection for mobile devices. In: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI ’10, pp. 181–190, Lisbon, Portugal. ACM, New York (2010)
79. Yin, J., Ren, X., Zhai, S.: Pen pressure control in trajectory-based interaction. Behav. Inform. Technol. **29**, 137–148 (2010)
80. Hsiu, M.-C., Huang, D.-Y., Chen, C.-A., Lin, Y.-C., Hung, Y.-P., Yang, D.-N., Chen, M.: Forceboard: using force as input technique on size-limited soft keyboard. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI ’16, pp. 599–604, Florence, Italy. ACM, New York (2016)
81. Zhong, M., Yu, C., Wang, Q., Xu, X., Shi, Y.: Forceboard: subtle text entry leveraging pressure. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI ’18, pp. 528:1–528:10, Montreal QC, Canada (2018)
82. Hincapié-Ramos, J.D., Guo, X., Moghadasian, P., Irani, P.: Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’14, pp. 1063–1072, Toronto, Ontario, Canada. ACM, New York (2014)
83. Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI ’17, pp. 3328–3339, Denver, Colorado, USA. ACM, New York (2017)
84. Vernier, F., Nigay, L.: A framework for the combination and characterization of output modalities. In: Proceedings of the 7th International Conference on Design, Specification, and Verification of Interactive Systems, DSV-IS’00, pp. 35–50, Limerick, Ireland. Springer, Berlin (2001)
85. Sobottka, K., Pitas, I.: A novel method for automatic face segmentation, facial feature extraction and tracking. Signal Process. Image Commun. **12**(3), 263–281 (1998)
86. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables, vol. 55. Courier Dover Publications (1972)
87. Lee, L.H., Yung Lam, K., Yau, Y.P., Braud, T., Hui, P.: Hibey: hide the keyboard in augmented reality. In: 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, pp. 1–10 (2019)
88. Bi, X., Smith, B.A., Zhai, S.: Quasi-qwerty soft keyboard optimization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’10, pp. 283–286, Atlanta, Georgia, USA (2010)
89. Gong, J., Tarasewich, P.: Alphabetically constrained keypad designs for text entry on mobile devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’05, pp. 211–220, Portland, Oregon, USA. ACM, New York (2005)
90. Walmsley, W.S., Xavier Snelgrove, W., Truong, K.N.: Disambiguation of imprecise input with one-dimensional rotational text entry. ACM Trans. Comput.-Hum. Interact. **21**(1), 4:1–4:40 (2014)
91. Bi, X., Zhai, S.: Bayesian touch: A statistical criterion of target selection with finger touch. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST ’13, pp. 51–60, St. Andrews, Scotland, United Kingdom (2013)
92. Goodman, J., Venolia, G., Steury, K., Parker, C.: Language modeling for soft keyboards. In: Proceedings of the 7th International Conference on Intelligent User Interfaces, IUI ’02, pp. 194–195, San Francisco, California, USA (2002)
93. Azenkot, S., Zhai, S.: Touch behavior with different postures on soft smartphone keyboards. In: Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services Companion, MobileHCI ’12, pp. 251–260, San Francisco, California, USA (2012)
94. Yu, C., Gu, Y., Yang, Z., Yi, X., Luo, H., Shi, Y.: Tap, dwell or gesture? Exploring head-based text entry techniques for HMDs. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI ’17, pp. 4479–4488, Denver, Colorado, USA. ACM, New York (2017)
95. Peshock, A., Duvall, J., Dunne, L.E.: Argot: a wearable one-handed keyboard glove. In: Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program, ISWC ’14 Adjunct, pp. 87–92, Seattle, Washington. ACM, New York (2014)
96. Zhai, S., Hunter, M., Smith, B.A.: The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design. In: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology, UIST ’00, pp. 119–128, San Diego, California, USA. ACM, New York (2000)
97. Lu, Y., Yu, C., Yi, X., Shi, Y., Zhao, S.: Blindtype: eyes-free text entry on handheld touchpad by leveraging thumb’s muscle memory. In: ACM Interactive Mobile Wearable Ubiquitous Technology, vol. 1, pp. 18:1–18:24, Maul, Hawaii, USA. ACM, New York (2017)
98. Hwang, S., Lee, G.: Qwerty-like 3×4 keypad layouts for mobile phone. In: CHI ’05 Extended Abstracts on Human Factors in Computing Systems, CHI EA ’05, pp. 1479–1482, Portland, OR, USA. ACM, New York (2005)
99. Wong, P.C., Zhu, K., Fu, H.: Fingert9: leveraging thumb-to-finger interaction for same-side-hand text entry on smartwatches. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI ’18, pp. 178:1–178:10, Montreal QC, Canada. ACM, New York (2018)
100. Yu, C., Sun, K., Zhong, M., Li, X., Zhao, P., Shi, Y.: One-dimensional handwriting: Inputting letters and words on smart glasses. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI ’16, pp. 71–82, San Jose, California, USA. ACM, New York (2016)
101. Grossman, T., Chen, X.A., Fitzmaurice, G.: Typing on glasses: adapting text entry to smart eyewear. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI ’15, pp. 144–152, Copenhagen, Denmark (2015)
102. Lee, L.H., Zhu, Y., Yau, Y.P., Braud, T., Su, X., Hui, P.: One-thumb text acquisition on force-assisted miniature interfaces for mobile headsets. In: 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom, pp. 1–10, (2020)



Lik-Hang Lee received a PhD degree from SyMLab, Hong Kong University of Science and Technology, and the Bachelor’s and M.Phil. degrees from the University of Hong Kong. He is currently an Assistant Professor (tenure-track) at the Korea Advanced Institute of Science and Technology (KAIST), South Korea, and the Head of the Augmented

Reality and Media Laboratory, KAIST. He has built and designed various human-centric computing specializing in augmented and virtual realities (AR/VR).

Tristan Braud got his PhD degree from Université Grenoble Alpes, France, in 2016. He is currently Assistant Professor at the Division of Integrative Systems and Design at the Hong Kong University of Science and Technology. His research interests revolve around human-centered pervasive system design, including mobile computing, cyber foraging, and AR.

Pan Hui is a Professor of Computational Media and Arts and Director of the Centre for Metaverse and Computational Creativity at the Hong Kong University of Science and Technology. He is also the Nokia Chair in Data Science at the University of Helsinki, Finland. He is a world-leading expert in Augmented Reality and has been recognized as an International Fellow of the Royal Academy of Engineering, a member of Academia Europaea, an IEEE Fellow, and an ACM Distinguished Scientist.

Networking and Cyber Foraging for Mobile Augmented Reality

11

Tristan Braud, Wenxiao Zhang, Benjamin Finley, and Pan Hui

Contents

11.1 Mobile Augmented Reality Requirements and Cyber-Foraging . . . 274
11.1.1 Requirements of MAR Applications . . . 274
11.1.2 Network Capabilities and MAR . . . 275
11.1.3 Cyber-Foraging for AR . . . 276
11.1.4 Arising Challenges . . . 276
11.1.5 Performance Models . . . 277
11.1.6 Highlights of This Chapter . . . 278

11.2 Related Works . . . 278
11.2.1 Mobile Augmented Reality . . . 279
11.2.2 Generic Cyber-Foraging Systems . . . 279
11.2.3 Cyber-Foraging for MAR . . . 279
11.2.4 Network Protocols . . . 280
11.2.5 Discussions . . . 281

11.3 Network Access . . . 281
11.3.1 Wireless Networks . . . 282
11.3.2 Future Wireless Architectures and 5G: Promises and Challenges . . . 283
11.3.3 Upload to Download Ratio on Asymmetric Links: A Delicate Balance . . . 284

11.4 Infrastructure and Transport Protocol for MAR . . . 285
11.4.1 Classful Traffic . . . 286
11.4.2 Congestion Control, Fairness, and Graceful Degradation . . . 287
11.4.3 Enforcing Low Latency with Loss Recovery . . . 289
11.4.4 Multipath . . . 289
11.4.5 Multi-Server and Distributed Computations . . . 290
11.4.6 Security and Privacy . . . 291
11.4.7 Implementation Notes . . . 291

11.5 Improving Latency at the Application Layer . . . 292
11.5.1 Mobile AR Pipeline . . . 292
11.5.2 Commercial SDK Cloud Offloading Procedures . . . 292
11.5.3 Commercial SDK End-To-End Latency Analysis . . . 295
11.5.4 Discussion . . . 296

11.6 Conclusion . . . 296
11.6.1 Access Link . . . 297
11.6.2 Network and Transport Layer . . . 297
11.6.3 Application Layer . . . 297
11.6.4 Future Challenges . . . 297

References . . . 298

T. Braud
Division of Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong, China
e-mail: braudt@ust.hk

W. Zhang
Computational Media and Arts Thrust, Hong Kong University of Science and Technology, Hong Kong, China
e-mail: wzhangal@ust.hk

B. Finley
Edge Cloud Innovation Lab, Huawei Technologies Co. Ltd., Shenzhen, China
e-mail: benjamin.finley@helsinki.fi

P. Hui
Computational Media and Arts Thrust, Hong Kong University of Science and Technology, Hong Kong, China

Division of Integrative Systems and Design, Hong Kong University of Science and Technology, Hong Kong, China
e-mail: panhui@ust.hk

Abstract

Mobile augmented reality (MAR) applications are gaining popularity due to the wide adoption of mobile and especially wearable devices such as smartglasses. These devices often strike a compromise between mobility, energy efficiency, and performance. On the other hand, MAR applications rely on computationally intensive computer vision algorithms with extreme latency requirements. Cyber-foraging allows resource-constrained devices to leverage the computing power of nearby machines, and has often been proposed as a solution for increasing the computing capabilities of constrained devices. However, this process introduces new constraints in the application, especially in terms of latency and bandwidth. MAR applications are so demanding that current network infrastructures are barely ready for such traffic. Such resource-hungry applications may rapidly saturate future wireless networks such as 5G. As such, developing cyber-foraging solutions for MAR applications requires a delicate balance in network resource usage to ensure seamless user experience. In order to identify the opportunities for further improving the Quality of Experience (QoE) for mobile AR, we break down the end-to-end

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_11

latency of the pipeline for typical MAR applications and pinpoint the dominating components in the critical path, from the physical transmission to the application. We then derive a set of actions to enhance the possibilities of cyber-foraging, taking into account the specificities of future networking technologies.

Keywords

Mobile augmented reality · Cyber-foraging · Computation offloading · Transport layer protocol · Application layer · Wireless networks · Mobility · Multipath networking · Distributed computing

11.1 Mobile Augmented Reality Requirements and Cyber-Foraging

Mobile augmented reality (MAR) is the mobile counterpart of augmented reality (AR) [1]. Typical scenarios involve user mobility and transport of the device [2–4]. When running on mobile devices, a typical application uses a video flow to detect and locate objects and regions of interest and augments the user's perception of the physical world with one or several virtual layers [5, 6]. More precisely, a MAR application exhibits the following data flow:

- **Input**: In general, a MAR application relies principally on the video flow generated by the device's camera. This video flow is essential to augment the users' perception of the environment. Other sensors' data, such as microphone, accelerometer, gyroscope, and GPS, may complement the video flow. Video from several cameras can also be combined, for instance, using a second camera for depth perception or gaze tracking. In some cases, a MAR application may use data from a companion device's sensors [7].
- **Processing**: The application analyzes the data from the sensors to determine what to render on the screen. In the video flow, the application locates regions of interest by recognizing and tracking objects and renders the corresponding virtual objects onto an overlay. To perform these operations, the application may require a database containing the features of objects to recognize, as well as the various virtual objects to overlay. This database may be embedded in the application or stored on a remote machine.
- **Output**: The application overlays the virtual objects either on top of the video on a screen or directly onto the user's field of vision in the case of more advanced headsets.

As mobile devices operate in mobility scenarios, they often are the result of a compromise between portability (size, shape, battery life) and performance (processing power and storage). Often, the components are low energy in order to accommodate for the limited battery supply. Thus, the limited form factor considerably limits the performance of the embedded components.

However, AR applications are computation-intensive. AR applications should function with little to no prior knowledge of the users' environment in real time. Additionally, aligning the virtual content with physical objects requires heavy use of complex image processing algorithms, involving object recognition, mapping, or tracking. Finally, the vast number of potential objects to recognize requires large-size datasets or models.

As such, mobile devices are often too constrained to address such requirements. Running a full application on-device will lead to either limited functionality or a severe degradation in the Quality of Experience (QoE).

In order to compensate for their constrained hardware, mobile devices may transfer the heavy computational tasks to other devices and servers, exploiting the growing ubiquity of computational resources. This process is also known as computation offloading or cyber-foraging. In the rest of this chapter, we will use both terms interchangeably. Although cyber-foraging allows us to leverage distant resources to decrease computation times, the additional network transmission times may lead to a significant increase in latency. Besides, video flows require large bandwidth for real-time transmission.

11.1.1 Requirements of MAR Applications

An important parameter of MAR applications is the motion-to-photon latency, that is, the delay between user motion and its repercussion on the displayed view. In the case of AR applications, even the slightest increase in such latency may have dramatic consequences. For example, alignment problems are a direct consequence of motion-to-photon latency. Alignment problems arise when the computation results do not match the current view (see Fig. 11.1), leading to a sharp decrease in QoE.

Michael Abrash, Chief Scientist at Oculus, claims that the limit for a seamless experience might be as low as $7\,ms$ [8]. Additionally, MAR often supports other applications and sometimes these are also heavily latency-constrained [9]. More realistically, we consider the inter-frame interval as the maximum acceptable latency: 33 ms for a 30 frames per second (FPS) video flow (interframe time $f = 1/30$), or 17 ms for a 60 FPS flow (interframe time $f = 1/60$).

The motion-to-photon latency τ for a MAR application fully executed on-device is the sum of several components:

$$\tau = \tau_{\text{camera}} + \tau_{\text{processing,device}} + \tau_{\text{display}} \tag{11.1}$$

Offloading tasks to a remote device adds several components:

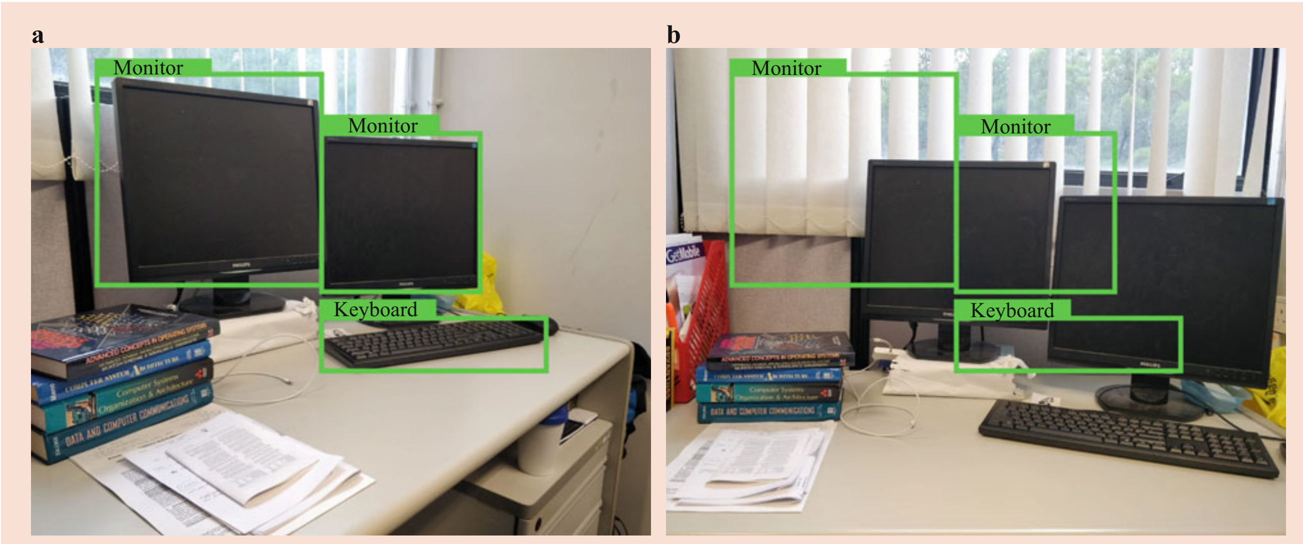


Fig. 11.1 Alignment problem in AR and motion-to-photon latency. (**a**) Original frame, the images are correctly recognized and the boundary is aligned. (**b**) User motion occurs. Due to latency, the computation results do not match the new view

$$\tau = \tau_{\text{camera}} + \tau_{\text{network,up}} + \tau_{\text{processing,server}} + \tau_{\text{network,down}} + \tau_{\text{display}} \quad (11.2)$$

where τ_{network} is the sum of the *propagation delay* (time for the first bit of the transmission to reach the destination), the *transmission delay* (time for all the packet's bits to be transmitted on the link), and the *queuing delay* (amount of time spent in queues). MAR applications should therefore strike a compromise between the processing time on-device $\tau_{\text{processing,device}}$, and the added network delays $\tau_{\text{encoding}} + \tau_{\text{network,up}} + \tau_{\text{network,down}}$ to minimize the motion-to-photon latency. Offloading tasks to a remote device should therefore satisfy the following requirements:

$$\tau_{\text{network,up}} + \tau_{\text{processing,server}} + \tau_{\text{network,down}} < \tau_{\text{processing,device}} \quad (11.3)$$

The bandwidth requirements of MAR are closely related to the quality of the video flow. Despite being composed of millions of rods, the human eye can see accurately only at the center of the retina, with a diameter of approximately 2 degrees, corresponding to 6 to 10 Mb/s of information. In practice, without gaze tracking, it is impossible to isolate this area on video frames. A typical smartphone camera has a field of view between 60 and 70 degrees, thus generating a raw video flow between 9 and 12 Gb/s at a similar resolution as the human eye. In practice, the bitrate of a 4K video flow (3840 × 2160) at 60 FPS with 12 bits per pixel is around 700 Mb/s uncompressed and 20–30 Mb/s after compression. In order for an AR application to gather enough information to operate, the bandwidth should be at least *10 Mb/s* for a compressed video flow. This value represents a minimum, as with the increasing sensing capabilities of mobile devices, we can soon expect to handle much larger flows, including stereoscopic or infrared video. When exploiting such features, the required bandwidth would grow proportionally, up to hundreds of megabits per second.

In the previous paragraph, we explained how compressing a video allows for a significant decrease in the required bandwidth. However, although video compression allows for dramatically decreasing the motion-to-photon latency when the network cannot accommodate the bandwidth requirements of raw video flow, it introduces new latency components:

$$\tau = \tau_{\text{camera}} + \tau_{\text{compression}} + \tau_{\text{network,up}} + \tau_{\text{decompression}} + \tau_{\text{processing,server}} + \tau_{\text{network,down}} + \tau_{\text{display}} \quad (11.4)$$

Encoding the flow is thus profitable if:

$$\tau_{\text{compression}} + \tau_{\text{network,up,compression}} + \tau_{\text{decompression}} < \tau_{\text{network,up,raw}} \quad (11.5)$$

Besides pure motion-to-photon latency, jitter is another critical parameter. Latency variations lead to stutter and choppiness in the display and have a dramatic impact on the perceived QoE. We consider that jitter starts to have an impact when it causes frames to be skipped. As such, the maximum tolerable jitter is around 33 ms for a 30 FPS video.

11.1.2 Network Capabilities and MAR

As highlighted in the previous section, cyber-foraging may represent a useful solution to the limited capabilities of mobile devices. However, the stringent latency constraints of AR applications combined with the substantial requirements in computation power and storage significantly complicate the operation. The closest resources from the user tend to be the

Table 11.1 Average network round-trip time measured for different offloading mechanisms from a university network in Hong Kong (September 2018)

D2D	Edge (Wi-Fi AP)	Edge (LTE BS)	Alibaba (HK)	Alibaba (HK)	Google (TW)	Google (TW)
Wi-Fi D	Wi-Fi	LTE	Wi-Fi	LTE	Wi-Fi	LTE
3.5 ms	3.7 ms	19.9 ms	5.5 ms	24.9 ms	42.2 ms	52.4 ms

least powerful, while cloud servers provide near-unlimited computation power and storage at the cost of a significant increase in latency.

Table 11.1 illustrates measurements of network round-trip latency performed in Hong Kong using an LG Nexus 5X on a local university's network to various destinations using several different technologies. These measurements were performed during daytime on a weekday in September 2018. In each case, we send 100 ICMP Echo Request packets and consider the average measured latency. Device-to-device (D2D) network latency using Wi-Fi Direct can be as low as 3.5 ms, and 3.7 ms for an edge server located right after a Wi-Fi access point. For Alibaba's cloud server in Hong Kong, the latency is still relatively low (5.5 ms with Wi-Fi). However, at the time of these measurements, the closest Google Cloud server was located in Taiwan, over 1000 km away. As such, the latency increases dramatically to 24.9 ms. Wi-Fi has the advantage of often being set up in a controlled environment. Such a solution is useful in more sedentary use cases such as MAR for production lines, where the mobility is limited to a predefined geographic area. In the case of fully mobile scenarios, cellular communications such as LTE networks remain the only available option for seamless operation. However, contacting the servers using LTE adds a minimum of 10 ms of latency, and up to 20 ms.

Therefore, it is critical to optimize link utilization in order to maximize the bandwidth usage while minimizing the latency. Multiple protocols aim at maximizing the usage of network resources, whether generic protocols targeted toward better link utilization or specific protocols for multimedia content. However, none can thoroughly address the requirements of augmented reality.

11.1.3 Cyber-Foraging for AR

Cyber-foraging can take multiple forms. Offloading computation to the cloud is currently the most common paradigm, and significant research [10, 11] has been performed in the area of mobile cloud computing (MCC). Specifically, in the case of MAR, some tasks are so heavy that offloading to the cloud results in a motion-to-photon latency reduction compared to on-device execution. However, since cloud offloading performs poorly in several delay-sensitive real-time mobile applications, the research community has considered moving the computing resources closer to the users. Mobile edge computing (MEC) aims at moving the servers as close as possible to the user's access link [12, 13], ideally at the access point itself, a single hop away from the user. Edge computing is at the core of the upcoming fifth generation of cellular networks (5G). This new computing paradigm enables improving QoE for mobile users, but is still in the infancy stage and has many open research challenges, including deployment cost, mobility management, and scalability.

Another alternative resides in device-to-device (D2D) offloading. In 2018, the worldwide penetration of mobile subscriptions reached 67% (over 80% in Europe and North America) [14]. The majority of the world's population carries at least one mobile device with them. It would thus make sense to exploit such resources in order to provide the additional performance boost needed to run AR applications. D2D offloading consists of distributing the computations among a pool of available devices located near the user. However, such techniques increase the complexity of the system while requiring users to give idle resources to the system voluntarily.

11.1.4 Arising Challenges

Computation offloading has been widely studied [10, 11, 15], but most of the literature focus on generic task offloading to the cloud, which targets computation-intensive applications with simple context information and loose latency requirements. We cannot blindly apply the previous computation offloading solutions to edge-based multimedia systems since these systems have some special characteristics:

- **Complex pipeline.** Generic computation offloading has a simple pipeline, with the primary process usually finished in a few function calls. However, a typical multimedia system involves several different subsystems, with each subsystem running its own pipeline, and different kinds of data traversing the whole pipeline after a few transformations. For example, an AR application involves at least computer vision and computer graphics subsystems, making the overall pipeline quite complicated. We have to choose wisely which parts to offload, and we should have systemic considerations.
- **Heavy data dependency.** Unlike the numerical calculation functions which have simple inputs and outputs [10], the execution of multimedia systems relies on large

volumes of data [16]. For example, computer vision tasks require frequent access to massive image databases, and computer graphics tasks render massive 3D models and textures. The storage and retrieval of these multimedia data require a complex subsystem itself, not to mention the effort in data migration.

- **Stringent latency requirement.** Generic computation tasks do not require frequent user interaction so the users can wait for the results for 1 or 2 s without major QoE problems. However, such latency is unacceptable in multimedia systems. AR applications involve frequent user interaction, and users are frustrated by slow system reactions. Additionally, slow execution in multimedia systems leads to less accurate results. For example, a delayed recognition results in an AR application cannot align precisely with the physical scene (as shown in Fig. 11.1).
- **High bandwidth demand.** Besides the rich contextual data dependency described above, the execution of multimedia applications generates a large volume of live data streams, the transmission of which is a heavy burden for the network infrastructure. To make the situation even more complex, there are clear trade-offs between the bandwidth consumption and latency when considering which part of the pipeline to offload. Offloading a specific step to the edge may decrease the overall processing time, but the input of this step would require a large volume of traffic compared to sending the output of the current step after local execution. We have to strategically decide which part of the pipeline to offload in order to achieve a cost-performance balance.
- **High-throughput demand.** For multimedia systems, the tasks are generated in a streaming manner, that is, the computation requests happen frame by frame. If the processing of a frame is not finished by the arrival of the next frame, the system has to either queue the tasks, making the latency higher, or drop the newly arrived frame, making the results less fluent or even incorrect. This streaming characteristic of multimedia requests requires the system to have a very high throughput, where the processing of each request should complete within a small inter-frame interval ($\sim$33 ms for a 30 FPS system, $\sim$17 ms for a 60 FPS system).
- **Scalability issue.** Edge computing is promising in terms of low-latency computation offloading, but the edge server also faces severe scalability issues. People are not concerned with the scalability issue when offloading tasks to the cloud since the cloud resources are usually elastic and can be dynamically allocated to satisfy the increasing demands of the users. However, this is not the case for edge computing, as the edge server attached to each base station has fixed and limited processing capability. MAR applications are so computation-intensive that each instance takes a toll on the available resources and considerably worsens the situation.
- **Geometric locality.** The geometric vicinity of the connected users of the edge server brings a unique characteristic of MEC, geometric locality. On the one hand, the requests may correlate with the physical environment (e.g., AR applications) and so that the edge server receives the same requests from time to time. On the other hand, users within the same vicinity may have common interests. This geometric locality gives us a chance to cache the computation and data on the edge server, which not only saves processing time but also helps to solve the scalability issue described above.

11

11.1.5 Performance Models

People build MCC or MEC systems to improve the performance of mobile applications, and the performance metrics are of crucial importance to evaluate the effectiveness of the task offloading methods. Some popular metrics in existing work [10, 11, 15, 17, 18] include execution time savings, energy savings, thermal level, and data usage. These intuitive metrics show the performance gain of the mobile applications when offloading specific tasks to the center or edge cloud. However, they reflect the performance gain of only one end user. The above metrics for generic task offloading cannot reflect the unique characteristics of both MEC and multimedia systems since they ignore the fact that massive concurrent application users exist in the real environment, and they do not consider performance from the service provider's point of view.

In order to guide the design of edge-assisted mobile systems as well as evaluate the multimedia applications with MEC specific metrics, we need some systemic performance models, which not only consider the user-perceived QoE factors in the production environment but also help with resource deployment planning from the service provider's point of view.

User-centric performance metrics:

- **Latency.** Latency is the most critical metric for mobile systems, and we define the end-to-end latency as the time period from the task generation on the client to the result displaying on the client.

  For each offloaded task in a production environment, the end-to-end latency consists of four parts: *client processing time, dynamic network delay, request queuing delay*, and *dynamic server processing time*.

$$\tau_{\text{production}} = \tau_{\text{client}} + \tau_{\text{networkDynamic}} + \tau_{\text{serverQueuing}} + \tau_{\text{serverDynamic}} \qquad (11.6)$$

In order to compare the latency with other systems, we can evaluate the system in a controlled environment, with controlled network and a dedicated edge server for a single user. The latency formula is simplified as follows:

$$\tau_{\text{simplified}} = \tau_{\text{client}} + \tau_{\text{networkControlled}} + \tau_{\text{serverDedicated}} \quad (11.7)$$

- **Throughput**. Throughput reflects the frequency of successful requests from a single client, and we usually use the term *frames per second (FPS)*. Throughput is determined by the client processing speed, the network bandwidth, or the server processing speed, whatever is the slowest part in the whole pipeline:

$$\text{throughput}_{\text{client}} = \min\left(\frac{1}{\text{time}_{\text{client}}}, \frac{\text{bandwidth}_{\text{network}}}{\text{dataConsumption}_{\text{request}}}, \frac{1}{\text{time}_{\text{server}}}\right) \quad (11.8)$$

- **Quality.** The service quality metric has different meanings for different tasks. For object recognition tasks, it is represented by {*recognition accuracy, recognition scope*}; for video processing tasks, it is represented by {*image quality(SSIM, PSNR), resolution*}; and for graphics rendering tasks, it is represented by {*scene complexity, rendering quality level*}.
- **Efficiency.** Similar to generic task offloading scenarios, the efficiency metric indicates the cost of the task execution, which consists of *data consumption, energy consumption, thermal level*.

System performance metrics:

- **Scalability.** The limited computational resource edge server has to process massive requests from multiple concurrent users. We can utilize a multithreading feature and set up multiple software instances, but the processing time of each request usually increases with the number of parallel processed requests. We can define the scalability as the number of processed requests in a unit of time:

$$\text{throughput}_{\text{server}} = \frac{1}{\text{time}_{\text{serverDynamic}}} * \#_{\text{parallelInstances}} \quad (11.9)$$

- **Cache performance.** Considering the special characteristics of MEC, caching data and computation is an important function of an edge server. Cache performance consists of *cache hit ratio (CHR)* and *cache hit precision (CHP)*.

In the remainder of this chapter, we apply these performance metrics in the design of our guidelines.

11.1.6 Highlights of This Chapter

In this chapter, we focus on how cyber-foraging and networking protocols can improve the QoE of AR applications. After reviewing some of the most influential related works, we study the current state of wireless networks and evaluate the challenges facing future networks. We then proceed to formulate guidelines related to cyber-foraging for MAR applications. Specifically, we target the following topics:

- **Network Access for MAR**: We first review current wireless network technologies and their potential for MAR applications. We show that despite some appreciable characteristics, none of the currently available wireless networks can accommodate the stringent requirements of MAR. Even promising future solutions such as 5G may not be able to unleash MAR's full potential.
- **Transport Protocols for MAR**: In order to optimize the link utilization, it is necessary to design new transport protocols adapted to the constraints of MAR systems. MAR traffic is more complex than most applications. It combines multiple sub-flows with diverse requirements and priorities, including video, audio, sensor data, and connection metadata. Even within a flow, not all data has the same worth for the application, e.g., I-frames and P-frames in a video flow. It is thus necessary to adapt the transmission to these multiple sub-flows while enforcing fairness toward other connections sharing the same link. Additionally, such a protocol should account for multihoming and distributed execution of tasks, to maximize the usage of all available resources.
- **Application-Level Networking and Computation Offloading**: A transport protocol dedicated to MAR computation offloading needs to constantly communicate with the application to assign the different sub-flows. Additionally, there are multiple possible strategies at the application level that allow improving the motion-to-photon latency. After reviewing the capabilities of current commercial MAR SDKs, and breaking down the latency for different tasks, we formulate guidelines on how to optimize the performance and QoE of MAR.

11.2 Related Works

In this section, we briefly summarize the major works related to cyber-foraging and mobile augmented reality. We first review the main technologies behind MAR applications, before describing generic cyber-foraging systems. We then move on to MAR-specific cyber-foraging solutions, and network protocols supporting such applications.

11.2.1 Mobile Augmented Reality

Augmented reality is discussed by many research works. In this section, we focus only on the most significant pieces of literature, the founding works, and other highly innovative or exciting studies.

Overlay [6] was one of the first works targeting deployment of MAR applications on smartphones. This work uses the smartphone's sensors to reduce the object recognition search space by taking into account the objects' geometric relationship. VisualPrint [19] reduces the uplink bandwidth demand by transmitting compressed feature points from a camera frame (i.e., their fingerprints) to the cloud for object recognition. Nestor [20] is a real-time planar shape recognition system for mobile phone. Once a shape is recognized, the system estimates their 6DoF pose and augments them in AR. Augmented vehicular reality (AVR) [21] allows vehicles to share visual information. Each vehicle can then align the visual information within its own field of view and eventually display it to the user through an AR head-up display (HUD) system [9, 22].

Image recognition is at the core of augmented reality. We summarize below a few interesting or innovative applications related to this topic. CrowdSearch [23] is an image search system exploiting crowdsourcing systems such as Amazon Mechanical Turk to validate automated search based on image recognition algorithms. ThirdEye [24] track the users' gaze in physical retail stores using smartglasses. The system then identifies the items in the user's field of view with an online search service in order to identify the users' browsing behaviors. Gabriel [25] uses Google Glass for assisting users in cognitive decline. It captures images using Google Glass' embedded camera and combines it with cloud processing for real-time scene interpretation.

Collaborative AR has been investigated in the AR research community for around two decades. AR^2Hockey (Augmented Reality AiR Hockey) [26] is a collaborative AR system with a real-time vision-based registration algorithm. Studierstube [27] is an AR architecture and platform for enabling interactive scientific visualization among multiple users. Reitmayr and Schmalstieg [28] describe a descendant of the Studierstube platform that supports stereoscopic rendering and 6DoF manipulation of annotation content. MagicMeeting [29] is an AR system for group discussion in a design zone, aiming to integrate 2D and 3D information in a shared environment seamlessly. Instead of using head-mounted displays in the above systems, Henrysson et al. [30] develop a collaborative AR application, a tennis game, on mobile phones to investigate the impact of multisensory feedback (e.g., visual, audio and haptic). Most of those collaborative AR concepts were proposed decades ago, and do not utilize the state-of-the-art recognition, tracking, and rendering techniques and cannot fit into mobile platforms.

11.2.2 Generic Cyber-Foraging Systems

Cloud offloading has been a hot topic since the emergence of cloud computing. MAUI [11] does code partitioning on mobile devices. Virtual machines perform code execution in remote clouds. Developers annotate which methods to offload. The offloading decision is made at run time. Similarly, ThinkAir [10] offloads annotated codes to Android virtual machines in remote or nearby servers, and the offloading decision is also dynamic with the help of a coordination server. CloneCloud [15] also utilizes virtual machines to do computation offloading, but there is no need for code annotation since they clone the whole device.

Edge computation offloading has been proposed to address the latency and scalability issues of cloud offloading. Femto clouds [31] provide a dynamic and self-configuring cloud system for many-device situations whereby the Femto cloud coordinates the usage of close-by underutilized mobile devices at the network edge. Thus, edge computing reduces network latency for computation offloading relative to pure cloud solutions. Similarly, CloudAware [32] provides a programming model for MEC for creating scalable MEC applications. Through analyzing the potential networks, network strength, potential compute resources for offloading, and server utilization, CloudAware splits the apps into components and creates an offloading strategy through solving an optimization problem (such as minimizing computation time) while also considering network connectivity.

These generic task offloading mechanisms are designed for computation-intensive tasks which do not require context information and timely response, while we focus on the efficiency of edge-assisted applications to support real-time user interactions. With the substantial data dependency characteristic of mobile multimedia applications, we cannot simply apply code partition mechanism in our systems, as the runtime transmission of massive multimedia databases leads to a considerable delay and data consumption. Application clone solution does not require the runtime data migration, but the duplicated storage of the multimedia database on the mobile devices is a massive waste of the limited mobile resource. Instead of these code partition or application clone mechanisms, we utilize heterogeneous client/server architecture to achieve flexible development and high efficiency. To address the scalability issue of the edge server, the mobile devices in our proposed systems would also utilize nearby mobile devices as backup processing and storage nodes to provide basic service.

11.2.3 Cyber-Foraging for MAR

Related research projects can be split into noncontinuous and continuous AR groups. In noncontinuous AR devices, must wait for results from the cloud before moving on.

Noncontinuous Case The Overlay [6] AR system, detailed in the prior section, labels and annotates indoor scenes using sensor and camera input. These inputs are also used to restrict the search area for object recognition by identifying the geometric and spatial relationships between objects. The system result is basic annotation text sent back by the cloud after object recognition. VisualPrint [19] is an extension of Overlay that lowers data usage by sending carefully selected image fingerprints to the cloud rather than the actual images. Unfortunately, the VisualPrint authors did not report an effect on the total latency when both feature extraction and fingerprinting tasks are performed on the mobile device.

Continuous Case Continuous object recognition is a significant challenge because, as mentioned, end-to-end latency is often longer than the inter-frame time of the camera. LiKamWa et al. [33] suggests two methods for improving the energy consumption of continuous mobile vision: aggressive standby and optimal clock scaling, which are both energy-proportional methods. The Glimpse [5] system performs continuous object recognition for road signs and faces. Relatedly, Glimpse [34] studies a hardware solution for smartly discarding unneeded video frames for efficiency gains in continuous mobile vision. The solution involves rethinking the traditional video pipeline for mobile devices. Recently, deep learning has also become a more conventional method in continuous mobile vision. The DeepEye [35] system, as an example, performs a detailed evaluation of periodic images through local deep learning with multiple large models. DeepMon [36] performs inference in continuous mobile vision through leveraging GPUs on mobile devices for convolutional (layer) processing.

11.2.4 Network Protocols

Mobile augmented reality offloading combines stringent requirements with increased network traffic. Current network infrastructures can hardly satisfy these requirements with the additional load. Even promising new technologies such as 5G or 802.11ax alone may not be up to the task of providing consistent performance [37], especially since the release of these technologies will be accompanied by new applications that will further increase the existing network load. According to Speedtest [38], the average household in the United States experienced a downlink bandwidth around 50 Mb/s in August 2016. However, analysts at CISCO [39] predict that household applications will soon require 1 Gb/s while considering 10 Gb/s “not excessive.” With such predictions, we can envision that the full potential of 5G will be required to satisfy future applications, leaving minimal leeway for MAR. Waiting for 6G, another decade from now, for providing seamless MAR is also not an option. We believe that seamless MAR cannot be achieved by purely brute-forcing the resources at the physical layer. Instead, it requires not only smart but also respectful use of the existing technologies. In other words, it is necessary to maximize bandwidth usage while sharing the link equitably with other connections. Such a balance is the role of transport layer protocols. There is currently no transport protocol that focuses on or even considers the specificities of MAR. However, some protocols developed for other applications may become a source of inspiration for MAR-specific networking. In particular, multiple protocols target multimedia applications, and more precisely, audio/video transfer and streaming. Other generic transport protocols also bring exciting concepts, among which congestion control over wireless links or multi-homing.

Audio and Video Protocols

Since the 1990s, audio and video transmission has been a major focus, with the creation of new services including video streaming, digital and satellite television, or voice-over IP. As such, various protocols were developed for multimedia content. Offloaded MAR applications generate a considerable amount of video and images. Such protocols develop concepts that could, therefore, be employed in the design of AR-specific transport protocols.

Resource Reservation Protocol (RSVP) [40]. Hosts use RSVP to request routers for QoS guarantees on specific flows. Although this protocol is rarely encountered in current networks, providing QoS guarantees on AR flows is critical to enable a seamless experience. Even though 5G introduces the concept of network slicing that allows specific applications to run within guaranteed parameters, allowing different sub-flows to request and define their expected QoS would lead to a finer partition of the network.

Real-Time Protocol (RTP) and Real-Time Control Protocol (RTCP) [41] are two complementary protocols that were developed for audio and video transmission. Both protocols operate on top of UDP with a slightly different focus. RTP provides a minimal layer for transmission of real-time content. As such, it includes several features on top of UDP. Such features include in-order delivery and synchronization of the payload between the media source and the client. RTCP adds QoS functionalities through the continuous transmission of reports to the media source. This pair of protocol provides several useful features for MAR applications, among which:

1. Media synchronization and jitter compensation.
2. Collection and synchronization of content between multiple sources.
3. Dialogue between the protocol and the application to adapt the media flow (video quality, bitrate, buffering) to the current status of the network.

Real-Time Streaming Protocol (RTSP) [42] does not directly handle the transmission of video data. Instead, it usually runs on top of RTP/RTCP or TCP. However, RTCP allows for a better dialogue between the application and the transport protocol by providing a set of directives, similarly to HTTP. Among other things, RTSP allows the client to set transport and playback parameters finely.

MPEG Transport Stream (MPEG-TS) () [43] is a digital container format for media transmission. It is used in current digital TV services. Similarly to RTP/RTCP, MPEG-TS enables stream synchronization, as well as multiple stream inter-synchronization. Besides these features, forward error correction (FEC) allows the client to recover lost or damaged frames without the need for retransmission.

D2D Multimedia Protocols

Although D2D communication has been a very active research field with the advent of mobile computing, most protocols target general data transmission without a specific focus on multimedia content. Chatzopoulos et al. [44] evaluate the feasibility of D2D task offloading on mobile networks. Nevertheless, the ad hoc network community proposed a few solutions for end-to-end connections. TFRC [45] is a TCP-friendly protocol for multimedia applications. Luo et al. [46] extends this protocol to a complete real-time video streaming solution for ad hoc networks. Although these solutions enable distributing media content over a pool of accessible devices, they are not suitable for the extensive requirements of MAR applications, both in terms of network resources and computational power. A MAR-specific protocol should also provide facilities for more traditional client-server communication.

Improving General Performance

Besides the works mentioned above, many transmission protocols have been developed for general performance optimization. Many TCP variants aim at maximizing link utilization in various environments, including large bandwidth-delay product links (TCP Cubic [47]), wireless links (TCP Westwood [48]), high-latency links (TCP Hybla [49]), mobile edge computing (NATCP [50]), or data center networks (Elastic-TCP [51]). However, these variants focus on reliable, in-order transmission of data, which interferes with the real-time constraints of MAR applications. However, several other alternatives introduce exciting functionalities.

Multipath TCP (MPTCP) [52] *and SCTP (Stream Control Transmission Protocol)* [53] enable connections to exploit multiple paths to the same server. For instance, a mobile device may exploit its cellular connection in parallel to a Wi-Fi connection for a larger bandwidth. Besides, it allows to always choose the path with the lowest latency. Finally, as Wi-Fi networks are not always available, the cellular connection can provide backup to avoid interrupting the service. This use case is particularly interesting during Wi-Fi handovers, which can take up to a second. The mobile device can instantaneously switch to the cellular connection and switch back to Wi-Fi when the new connection has been established [54].

Quick UDP Internet Connections (QUIC) [55] is a protocol developed by Google that aims at providing functionalities from the transport layer to the application layer in an integrated solution. QUIC operates on top of UDP and includes features from TCP and MPTCP while also providing functions similar to TLS and HTTP. This protocol is currently part of Chrome, Google's web browser.

Datagram Congestion Control Protocol (DCCP) [56] is an interesting take on congestion control. Contrary to its lossless counterparts, DCCP is designed for applications where new data is preferred to retransmission or reordering. As such, it allows real-time applications to operate while providing congestion control and fairness toward other connections.

11.2.5 Discussions

None of the protocols discussed in this section exactly match the constraints of MAR applications. Multimedia protocols often focus on a single type of multimedia application (e.g., video streaming), and do not account for the diversity of traffic generated by MAR applications. On the other hand, generic transport protocols often target non-real-time applications in a specific scenario. Finally, some protocols consider particular use cases that are not necessarily compatible with more traditional client-server applications. However, these protocols all present exciting functionalities that can be combined in a MAR-specific protocol, targeting both current and future network infrastructures [57].

11.3 Network Access

MAR applications are both computationally and storage intensive, and mobile devices are often too constrained to address these requirements. As such, it is necessary to offload the most computation-intensive tasks to remote devices in parallel with accessing data in large databases. In order to provide a seamless experience to the user, the additional latency brought by the network needs to be low enough to prevent the apparition of stutter and alignment problems. We thus consider that, for a 30 FPS video flow, both the motion-to-photon latency and the jitter should be below 33 ms so that no frame is skipped during processing. Although these values are already very low, considering the current state of networks, they are in no way a lower bound. MAR applications should strive for motion-to-photon latencies and jitters as low as possible in order to account for prolonged usage. In this section, we summarize the capabilities and limitations of current network architectures and establish the requirements for future networks to handle MAR traffic.

11.3.1 Wireless Networks

Among the multitude of wireless network technologies available, only a few can pretend to address the requirements of MAR applications in terms of bandwidth and latency. As such, in this section, we only consider 3G, 4G, and Wi-Fi networks. We already presented in the introduction section an estimation of the latency with various technologies in a controlled environment. In this section, we expand on this simple experiment by presenting measurements from large-scale network data collection companies, including Speedtest [58] and OpenSignal [59]. These companies have gathered a considerable amount of data around the world, reaching millions of users and millions of data points. We complement these measurements with peer-reviewed studies for more insights on the root causes of these results.

HSPA+ (High Speed Packet Access)

In theory, HSPA+ can achieve high throughput, with a bandwidth of up to 168 Mb/s on the downlink, and an uplink bandwidth up to 22 Mb/s. However, most of the commercial implementations provide only 21–42 Mb/s. In practice, only a sheer portion of this bandwidth is available to the users. In the United States, OpenSignal reports low download throughputs (0.66–3.48 Mb/s), and high latencies (109.94–131.22 ms) [60]. A study of three local ISPs in Singapore corroborates these results [61]. This work exhibits maximum throughput of 7 Mb/s on the downlink, with uplink throughputs no higher than 1.5 Mb/s. Besides these low values, both the uplink and downlink throughputs vary abruptly over time. Throughput variations may reach several orders of magnitude in a timespan of a few milliseconds. In some extreme cases, latency can skyrocket up to 800 ms. With such high latencies, HSPA+ is thus not adapted to real-time applications, as any latency over 100 ms starts to be directly noticeable by the user. Throughput is low enough to cripple any multimedia application. As such, HSPA+ is not suitable for seamless MAR operation. However, there are still areas not covered by LTE [62], and 5G is still in its infancy. HSPA+ would allow keeping the connection running (e.g., sending connection metadata and low-volume sensor data such as GPS or accelerometer) while waiting for the device to reconnect to a more robust network. Such behavior would allow keeping the application running in degraded conditions (as most operations would run on-device), and prevent service interruption.

LTE (Long-Term Evolution)

LTE is built on top of HSPA to improve both the throughput and the latency. LTE with 4×4 MIMO can potentially achieve a downlink throughput of 326 Mb/s, and uplink throughput of 75 Mb/s, while halving the latency compared to HSPA+ [63]. LTE was originally thought to reach latencies as low as 10 ms [64]. In practice, LTE provides a noticeable improvement in throughput compared to the previous generation of mobile networks – in the United States, the downlink throughput ranges between 6.56 and 12.26 Mb/s. However, the actual latency remains relatively large, between 66.06 and 85.03 ms, according to OpenSignal [60]. Speedtest reports slightly higher throughputs, with an average downlink throughput of 19.61 Mb/s, and an average uplink throughput 7.95 Mb/s [38]. LTE also uses lower frequencies than HSPA, which allows for a higher range, and thus a broader deployment in rural areas. In the United States, 98% of the population is covered [65]. A recent study comparing the LTE latency carried over three operators in Hong Kong and four operators in Helsinki provides slightly more positive results [62], with average latencies between 40 and 60 ms, and availability over 95%. However, this study also showed a strong geographic dependency in the results, with some areas in Hong Kong displaying latencies over 120 ms, as well as a noticeable increase in latency when used in a mobile setting.

Even though real-life measurements of LTE latency and throughput are not as high as the theoretical capabilities of the technology, the improvement compared to HSPA remains significant. LTE allows for some real-time applications to operate seamlessly. With careful design, it may even be possible to run real-time multimedia applications such as MAR and mobile cloud gaming.

LTE Direct

The LTE standard features the possibility for D2D communication with LTE Direct [66]. LTE Direct introduces *in bound D2D*, where mobile devices can communicate with each other within the same frequency spectrum as LTE without going through a conventional cellular infrastructure. The theoretical communication range of this technology can reach up to 1 km, with a throughput of 1 GB/s, and lower latencies [67].

However, to the best of our knowledge, LTE Direct is only a theoretical use case that had not been deployed in real life. With the advent of 5G, it is most probable that LTE Direct will remain a promising, yet unexplored, technology.

Wi-Fi

The main advantage of mobile broadband networks resides in their quasi-ubiquity in urban environments, at the cost of performance and stability in densely populated areas. On the other hand, Wi-Fi can provide more reliable Internet access, with higher throughput, lower bandwidth, and lower jitter, at the cost of a restricted range and longer handover times.

802.11n and 802.11ac are the most deployed versions of Wi-Fi nowadays. Both present a high maximum bandwidth, 600 Mb/s for 802.11n, up to 1300 Mb/s for 802.11ac. However, in practice, OpenSignal reports much lower download bitrates, around 7 Mb/s and 33 Mb/s for 802.11n and 802.11ac, respectively [68]. Several factors can explain this discrepancy. First of all, the values mentioned in the standard correspond to the maximum bandwidth that can only be reached in ideal conditions—client close to the base station, high SNR, controlled environment. On the other hand, the OpenSignal values correspond to measurements "in the wild," for all users. Therefore, they include users at home with low interference, but other users who try to access Wi-Fi from a crowded coffee shop in far from ideal conditions. Besides, OpenSignal's measured bandwidth corresponds to the bandwidth to distant servers. With Wi-Fi access points reaching bandwidths of a gigabit per second, it is possible that the bottleneck is located further in the network, for instance, a Wi-Fi router connected to a DSL network, or even a 100 Mb/s Ethernet network. Regarding latency, OpenSignal reports an average of 150 ms [69]. However, latency can drop to a few milliseconds in a controlled environment, as shown in Table 11.1, where we managed to achieve a round-trip latency below 5 ms on the university network.

Mobility is much more limited when using Wi-Fi than when using broadband networks. As mentioned previously, the limited transmission range and the long (up to several seconds) handover times when changing the access point can severely cripple any real-time mobile application. Moreover, public access points are sparsely deployed, and often require signing up for each individual connection. As such, it is not possible to continuously transmit data over only Wi-Fi, even during short trips. A study [70] published in 2012 showed that, in a medium-size French city, open Wi-Fi was accessible 98.9% of the time, similarly to 3G, with a 99.23% availability. However, due to the problems mentioned above, such Wi-Fi access points only provided an Internet connection 53.8% of the time, with extremely low average throughput: 55 Kb/s. As a comparison, the average 3G throughput measured in this study was around 90 Kb/s.

Wi-Fi networks also display by design a significant weakness, known as the performance anomaly problem [71, 72]. Due to how Wi-Fi changes the modulation and coding scheme, and thus the available data rate, depending on the SNR, users with a low SNR may degrade the actual throughput of users with a higher SNR. Figure 11.2 displays a typical example of the performance anomaly, with two users located at a different distance from the access point. When both User A and User B are close to the access point, they communicate with the access point using the same modulation and coding scheme and experience similar throughputs. However, when User B moves further away from the access point, its modulation and coding scheme will change, and the experienced data rates will decrease. User A's throughput will also drop as User B will occupy the channel longer. With the channel busy, User A cannot transmit his data, and thus experiences degraded service.

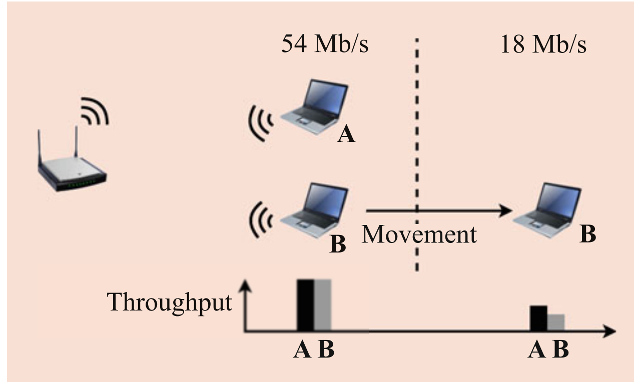


Fig. 11.2 The Wi-Fi performance anomaly problem. When B moves further from the Wi-Fi access point, the data rates experienced by both A and B decrease

Wi-Fi Direct and Wi-Fi Ad Hoc

The Wi-Fi standard also allows for D2D communication, similarly to LTE. Wi-Fi Direct, or Wi-Fi peer-to-peer [73], allows devices to communicate with each other without going through an access point. Contrary to Wi-Fi ad hoc, Wi-Fi Direct is a single-hop communication scheme, similar to Bluetooth. Wi-Fi Direct works by establishing an access point to which the other device connects. As such, only the device acting as a router has to be compatible with Wi-Fi Direct.

Wi-Fi Direct is useful to establish a one-to-one connection to another device, for instance, when connecting smartglasses with a companion device in the user's pocket. However, it is not adapted to more complex device-to-device scenarios. Wi-Fi ad hoc allows creating a multi-hop network of several devices, allowing many users to contribute their computational power. However, most mobile OS (Android and IOS) do not support Wi-Fi ad hoc by default, thus preventing its application at large scale.

Wi-Fi Direct has a transmission range of up to 200 m, and a maximum bandwidth of 500 Mb/s. Similarly to Wi-Fi, this bandwidth represents a theoretical maximum in ideal conditions, and actual data rates vary depending on the distance, the SNR, and the mobility [74]. Compared to Wi-Fi Direct, LTE Direct was designed to be more energy-efficient for a large number of users, with a higher transmission range, and better nearby devices discovery. However, for small amounts of data, Wi-Fi Direct is more energy-efficient [75] and is cheaper to deploy. Furthermore, Wi-Fi Direct is nowadays available to virtually all mobile devices, while LTE Direct will probably never be released to the customer market.

11.3.2 Future Wireless Architectures and 5G: Promises and Challenges

As we have shown in the above sections, none of the current wireless architectures can completely accommodate the requirements of MAR. Availability, bandwidth, and latency can only be satisfied in specific scenarios and controlled setups. 5G is considered as a game changer for heavily constrained applications, including vehicular networks, 360 video streaming, and MAR. MAR is acknowledged as one of the typical use cases by the 5G White Paper [76] published by the NGMM alliance, a group of interest regrouping international mobile operators. MAR is defined as a *pervasive and part of everyday life* application. As such, MAR applications should receive stable and uninterrupted service, at least in urban areas. This white paper's general recommendations include a minimum bandwidth of 50 Mb/s. Such bandwidth should be obtained 95% of the time in 95% of the locations. Besides, 5G should exhibit data rates up to 1 Gb/s, with an average latency of 10 ms for general applications, down to 1 ms for specifically constrained applications. For MAR, the white paper recommends downlink data rates of 300 Mb/s, uplink data rates of 50 Mb/s, and a 10 ms end-to-end latency. Besides, service should be seamless in mobility scenarios up to 100 km/h.

Our recommendations slightly differ from the values proposed by the white paper. MAR applications heavily rely on video flows coming from the device. Such video flows require large available bandwidth, with guaranteed minimum data rates. Besides, video flows represent the most bandwidth-hungry flows for typical MAR applications and are transmitted on the uplink. We thus consider that the uplink bandwidth should be equal to the downlink bandwidth to provide an acceptable QoE to the users. Video flows are generated at a fixed constant rate and need to be processed close to this rate to prevent alignment problems and stuttering. As such, substantial variations in the round-trip latency may significantly impact the QoE. Wireless networks typically present abrupt changes in throughput and latency, that congestion control algorithms cannot address fast enough. The network architecture itself should ensure that the round-trip latency does not vary by more than what would be acceptable for the application (30 ms for a 30 FPS flow).

Regarding the computing infrastructure, edge computing is at the core of current 5G proposals. However, contrary to cloud computing, such servers will present fixed and limited resources. Computation-heavy applications such as MAR will thus require to complement edge computing with cloud computing. Cloud servers should be deployed as close as possible from the user, preferably within the same city in order to minimize the additional latency. Major cloud providers already follow this practice in some areas [77–79].

Finally, the deployment of micro base stations promises to significantly increase the availability while reducing the number of users connected to the same base station.

5G represents a radical change in the landscape of mobile broadband networks. The announced bandwidth and latency open the door to many new applications that up to now were limited to sedentary usage on controlled wireless environments. However, despite these promises, we foresee that usage will most likely catch up. Similarly to 4G, the explosion of applications exploiting the power of 5G will progressively occupy and overburden the network. For instance, nowadays, seamless 4K video streaming over mobile networks is, at best, delusive. However, in Sect. 11.1.1, we estimated that video flows could reach up to several Gb/s. Similarly, new usages will involve 360 video streaming, which requires a 16K quality for an acceptable VR experience. Other technologies such as holograms may also arise, putting further strain on the networks. Network slicing may partially respond to this issue. However, the physical deployment of base stations needs to follow the increasing demand on the network.

5G is not the only promising wireless technology. 802.11ax, also known as Wi-Fi 6, promises data rates up to 10 Gb/s, which would allow for high-quality, high QoE multiuser MAR applications to run in semisedentary scenarios such as AR for the industry.

11.3.3 Upload to Download Ratio on Asymmetric Links: A Delicate Balance

MAR traffic often follows unusual network usage patterns. Indeed, most of the heavy MAR traffic will take place on the uplink. The client device transmits video flows, audio flows, sensor data, and model updates to the server, which returns computation results and metadata. There is thus a significant disproportion between the uplink and the downlink traffic. As stated previously, a MAR application requires a larger uplink bandwidth than the downlink bandwidth. Nevertheless, usual network access links are asymmetrical and feature the opposite profile. In such networks, the downlink presents a significantly higher bandwidth than the uplink, often over several orders of magnitude. In this section, we study the history of access networks and try to forecast the tendencies of future access networks. We then discuss how MAR traffic can integrate within such a landscape.

History and Future of Access Networks

Internet access started to become available for end consumers around 1992 with the development of dial-up modems [80]. Wide adoption, however, only started around 2000 with the advent of the broadband Internet (ADSL and Cable) commercialization. These technologies provide much higher data rates and lower latencies than dial-up modems. However, they also feature asymmetric capacity, that is, a higher bandwidth on the downlink than on the uplink. In 2005, broadband access surpassed dial-up in the United States, and the number of American users kept growing until 2014 [81]. Optical fiber is nowadays replacing this broadband infrastructure. Although optical fiber is theoretically symmetrical, it is still sold with asymmetrical capacity in multiple countries. For example, at the time this chapter was written, Orange was still selling asymmetrical optical fiber in France, with up to 2 Gb/s download and 600 Mb/s upload [82]. Similarly, it is not uncommon to encounter cable access offer with fiber-like capacity on the downlink. However, the uplink capacity is usually much lower. Such an access medium is prevalent in the United States.

Mobile broadband networks follow the same trend. The emergence of 3G, followed by 4G, enabled usages to converge between desktop and mobile Internet. This convergence got to the point where most of the Internet traffic is nowadays mobile. More than half of the web traffic in 2015 was coming from smartphones, tablets, and other mobile devices [83]. In some regions, the mobile Internet penetration is much higher than desktop Internet [84]. 5G networks seem to lean toward providing dynamic asymmetric connection, despite being capable of symmetric data rates through full-duplex transmission.

According to Speedtest [38], in August 2016, the measured download throughputs were, on average, three times higher than the upload throughputs. The average American broadband Internet connection displays a downlink-to-uplink ratio of 2.92. Out of the six fastest ISPs in the United States, one achieves symmetric data rates. The other five display an asymmetry factor of 3.31 to 8.22 between their downlink and uplink bandwidth. Similarly, for mobile Internet providers, the average connection in the United States experience a downlink-to-uplink ratio of 2.47, while the four fastest mobile providers achieve values between 1.81 and 3.20.

802.11 currently provides low delay/high throughput first-hop link with a theoretically symmetric link. However, the quality of communications also depends on the broadband access link to which the access point is connected. Moreover, some problems such as the performance anomaly [71, 72] can severely impact the overall performance. Still, with the development of optical fiber, the gain in throughput and latency compared to mobile networks makes 802.11 the best available candidate for MAR offloading.

Are Symmetric Links Really Necessary?

In the early ages of the consumer-market Internet, the download volumes were much larger than the uploads, up to ten times. Typical traffic included email, web surfing, and other server-to-client applications. In the second half of the 1990s, the advent of peer-to-peer, and later cloud computing, inverted this trend. In 2012, download volumes were only three times larger than upload volumes, down to 2.70 times in 2016 [85]. Nowadays, download-to-upload ratios are again increasing, thanks to the emergence of new bandwidth-heavy entertainment services, such as streaming services (Netflix, YouTube), that also lead to the decline of peer-to-peer usage. Researchers, academics, and analysts foresee the future of network usage to be more asymmetrical [85, 86], with both download and upload usage growing significantly. Nonetheless, we observe that most network access links nowadays are far more asymmetric than the current usage (see section "History and Future of Access Networks"). Besides, uplink usage tends to increase inconsistently with a high variance

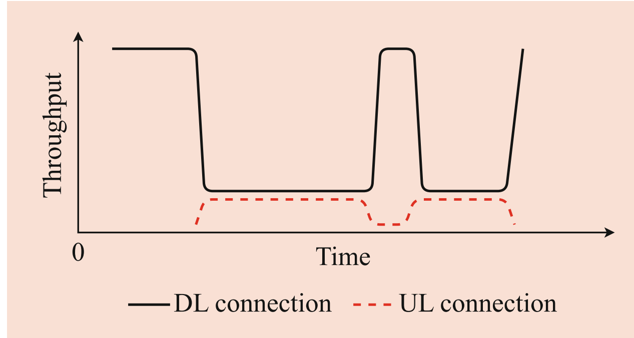


Fig. 11.3 Typical impact of an uplink connection (red dotted) on the throughput of a TCP downlink connection (continuous line) on an asymmetric link with oversized uplink buffer. The uplink connection leads to a sharp decrease of performance for the downlink connection

[87]. It is thus difficult to predict what the future traffic asymmetry will be.

Typically, Digi-Capital predicts that the AR consumer market revenue will quadruple by 2023 [88], following the current market increase trend. This increase would involve an explosion of applications in the upcoming years, in particular mobile applications. However, by that time, the hardware of mobile devices will most likely still be too limited for the most computation-intensive applications. In order to provide support for such applications, it will be necessary to offload computation to a remote machine. Such a machine can be in the cloud, the edge, or the device of another user. Offloading computation for AR applications will lead to a significant increase in the traffic on the uplink due to the volume of image and video flows. On asymmetric links, such uncommon traffic patterns are likely to cause problems. Due to the increased volume of upload traffic, the uplink capacity is directly linked to the Quality of Service (QoS) and the Quality of Experience (QoE). Besides, this traffic will increase the chances of congestion on the uplink. When uplink buffers are oversized, uplink congestion will directly affect TCP traffic in the other direction, due to the retention of acknowledgments in the uplink queue [89, 90]. The effect of a single TCP download on an asymmetrical congested uplink with an oversized buffer (bufferbloat) is shown in Fig. 11.3. A transport protocol focusing on MAR applications should thus concentrate primarily on maximizing the uplink usage while at the same time minimally impacting cross traffic.

11.4 Infrastructure and Transport Protocol for MAR

Due to the massive requirements of MAR in both network and computation resources, designing a transport protocol for offloading MAR applications represents a significant challenge.

The current offloading model favors transmission through a single link to a single server. However, current communication technologies are too limited to provide satisfying performance. Similarly, computing resources either are too far in the network or may present limited computation capabilities. It is therefore necessary to rethink the entire communication paradigm in order to meet MAR latency and bandwidth constraints. Exploiting every single available communication channel available will be critical for optimal performance. Similarly, computations may be distributed around multiple servers located close to the user. In this section, we provide guidelines for a new generation of transport protocols focused on latency-constrained, bandwidth, and computation-hungry multimedia applications. Although the concepts developed in this section were written with MAR applications in mind, they can also apply to any multimedia computation offloading, including VR and mobile cloud gaming.

A transport protocol for MAR applications should have the following properties:

1. *Classful traffic*: Classful traffic allows the protocol to account for the diversity of data that MAR applications generate: video, sensor data, or metadata. Such traffic has extremely different requirements in latency, bandwidth, and reliability.
2. *Fairness*: Transport protocols should maximize their bandwidth usage while remaining fair to other connections. Despite the strict requirements of MAR applications, a connection cannot exploit the available resources to the point of being detrimental to other connections.
3. *Low-latency*: MAR being latency-sensitive, it is primordial to optimize the transmission in order to keep close-to-real-time communication.
4. *Fault-tolerant*: To enforce the low-latency property, the protocol should exhibit a certain level of tolerance to losses in order to avoid the additional latency exhibited by packet retransmissions.
5. *Distributed*: Computation offloading may exploit multiple devices, from nearby companion devices with low computation power to powerful servers in the cloud, going through intermediate solutions at the edge of the network.
6. *Secure*: MAR relies heavily on video flows. These videos may contain elements that infringe users' privacy.

In this section, we consider a MAR application transmitting a VP8- or VP9-encoded flow. VP8 and VP9 can generate three types of frames: (1) Reference frames or *I-frames*. Frames that do not depend on any other frames. Other frames are encoded as a difference with the I-frame. (2) *Golden frames*. Intermediary frames that depend on the I-frame. These frames participate in the encoding of P-frames. (3) Predicted frames or *P-frames*. These frames depend on both I-frames and P-frames. Besides these video frames, the application also generates: (5) *Sensor data*. Various data coming from the device's sensors: IMU, GPS, light sensor, etc. (6) *Connection metadata*. Data sent periodically to keep the connection running, QoS measurements, and application signaling. Various types of *Sensor Data*, as well as *Connection Metadata*.

11.4.1 Classful Traffic

MAR applications are complex applications that rely on many types of data. As such, MAR offloading to a distant machine requires transmitting a wide range of information, including audio, video, sensor status, computation results, and connection metadata. Each type of traffic presents its own requirements in terms of latency, bandwidth, and reliability. For instance, video and audio flows require large bandwidth and close-to-real-time latency. However, for such flows, new data is often preferred to retransmission. On the other hand, connection metadata or computation results may need to reach their destination at all cost while requiring minimal bandwidth. It is therefore necessary to define internal that rule the transmission behavior.

We consider three base classes:

1. *Best effort*: Heavily latency-constrained data and data for which losses are not consequential to the application. This class primarily consists of sensor data, audio flows, and video I-frames. These flows are made of periodic data that present interest at the moment they are generated. The application is most likely to generate new data faster than the retransmission process.
2. *Best effort with loss recovery*: Data sensitive to both losses and latency. This class may be further divided depending on whether latency or loss recovery is the most critical parameter. For instance, a VP8 flow can include three types of frames: P-frames, I-frames, and golden frames. I-frames belong to the Full Best Effort class. These frames are decoded using a combination of P-frame and golden frame. As P-frames are generated less frequently than golden frames, a P-frame loss will significantly alter the rest of the video flow. On the other hand, latency is more critical for golden frames.
3. *Critical data*: Data favoring reliable, in-order delivery. Connection metadata and some types of computation results dependent on previous computations may fall into this category.

As MAR applications generate a vast variety of traffic, it is possible to define other classes depending on the specific requirements on latency bandwidth and reliability. For instance, a flow may require in-order delivery but no loss recovery, while another flow can be sensitive to only certain losses.

Besides traffic classes, a MAR application should also provide some mechanisms for graceful degradation. In the case of congestion, intermittent connectivity, or general poor network performance, the protocol may discard or delay certain traffic flows until the situation improves. In order to determine which flows to affect first, the protocol should define per-flow priorities. A transport protocol for MAR should define at least five priority classes:

1. *High priority* data should never be delayed nor discarded. As soon as the network conditions force high priority data to be delayed or discarded, it is better to stop the application completely. This includes the P-frames in the video flow.
2. *Medium priority 1* data would rather be delayed than discarded. Flows belonging to this priority class include most of the *critical data* traffic class.
3. *Medium priority 2* data can be discarded but not delayed. Data belonging to this priority class include all flows for which new data is preferred retransmission or reordering. For instance, I-frames in a video flow are constantly replaced by new data. The video flow can thus conserve its integrity if the reference frames are transmitted.
4. *Medium priority 3* data can withstand a certain amount of delay, after which it can be discarded. Such a priority class is useful for flows where new data replace older data at a low pace. For instance, some sensors' data may be sent every second. It may thus be interesting to delay the data for up to 900 ms in case of temporary degradation in the network conditions.
5. *Low priority* data belongs to flows that can entirely be discarded would congestion arise. Such flows are application-dependent and thus depend on the actual implementation. For instance, an application using GPS for coarse location and accelerometer data for fine location may revert to using only the GPS data and discarding the accelerometer signal until the situation improves.

In order to allow fine-grained control, each priority class should include an internal priority list in order to choose which service should be degraded first. A typical example would be a sudden decrease in available bandwidth, without an increase in latency. Such a scenario may happen on bottlenecks employing queuing strategies designed to combat bufferbloat such as CoDel or FQ CoDel [91]. In these conditions, it is necessary to reduce the bandwidth usage of the application and flows belonging to the low priority and medium priority 2 classes. Not all medium priority 2 flows may be impacted. The application developer may thus define which flows in the medium priority 2 class to discard first. We represent the sub-flows of our example MAR application and the typical traffic and priority classes in Table 11.2.

11

11.4.2 Congestion Control, Fairness, and Graceful Degradation

As stated in the previous sections, MAR applications generate a considerable amount of data unconventionally. Unregulated data transmission may, therefore, have a catastrophic impact on other connections transmitting on the same links. Despite the diverging goals of MAR transmission and regular traffic, it is critical to implement some sort of congestion control algorithm to ensure the MAR application does not use more than its fair share of the link capacities. Most congestion control algorithms focus on reliable, in-order delivery. Meanwhile, MAR applications require in-time task processing, with stringent latency constraints. It is thus not possible to apply the concepts typically used in current congestion control algorithms. For instance, most TCP variants rely on a congestion window representing the amount of data that can be transmitted on the network without being acknowledged. This congestion window progressively inflates until congestion—generally signaled by a loss—happens. The algorithm then decreases the size of the congestion window, indifferently impacting the total traffic of the connection. However, as we have seen in the previous section, MAR traffic is composed of various sub-flows with diverse requirements and priorities. Some sub-flows can be delayed or discarded if the available bandwidth is not sufficient, while others should be transmitted at all times in order not to interrupt the service. As such, congestion should not impact all sub-flows in the same way. This concept is known as *graceful degradation*. By introducing *graceful degradation* into the congestion control algorithm, it is possible to keep a

Table 11.2 Typical AR traffic flows and the associated traffic classes and priority classes

Data type	Traffic class	Priority class
Video I-frames	Best effort with loss recovery	High priority
Video golden frames	Best effort with loss recovery	Medium priority 1 or medium priority 3
Video P-frames	Best effort	Medium priority 2 or medium priority 3
Connection metadata	Critical data	High priority or medium priority 1
Sensor	Best effort	Medium priority to low priority

seamless, yet slightly degraded, experience for the user while preserving fairness toward other connections.

Such a system can only be achieved through a constant dialogue between the application and the protocol. By using the classes defined in the previous section, the protocol knows which data it can discard or delay when congestion happens. Meanwhile, it can preserve critical information flows. The protocol can also communicate QoS information back to the application. Indeed, some countermeasures to congestion cannot happen at the protocol level. For instance, if the available bandwidth is particularly low, the application may directly alter the quality of the video—resolution, bitrate, and frames per second. Such a dialogue avoids interrupting the user experience for both short-term and long-term network perturbations.

We summarize these ideas in Fig. 11.4. In this figure, we represent a typical additive increase multiplicative decrease congestion window similar to what can be encountered with TCP, to which we add the multiple sub-flows composing MAR traffic. Our example application described in the introduction of this section presents five sub-flows to which we assign the following priority classes:

1. *I-frames*. *Best effort with loss recovery* and *high priority*. All the following frames depend on the I-frame for decoding. Besides, the common practice consists of generating one I-frame every 30 frames. In a 30 FPS video flow, losing a single I-frame would, therefore, result in losing one entire second of video.
2. *Golden frames*. *Best effort with loss recovery* and *medium priority 3*. As golden frames participate in the encoding of P-frames, losing a golden frame also results in losing the following frames, although to a lesser extent. As such, it may be worth delaying the transmission of such frames if the delay is shorter than the time to transmit a few P-frames.

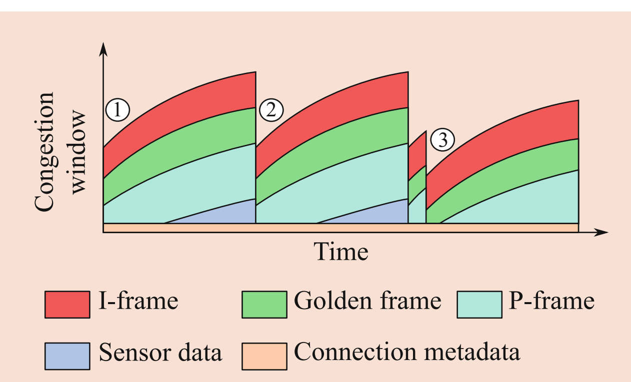


Fig. 11.4 Graceful degradation for MAR congestion control. The flows critical to the application's operation are prioritized

3. Predicted frames or *P-frames*. *Best effort* and *medium priority 1*. P-frames do not require retransmission as a new frame will almost always be generated within the same time interval. They can thus be discarded would the need arise.
4. *Sensor data*. *Best effort* and *lowest priority*. In the case of our application, we consider the other sensor data only to provide additional information to the video flow. For instance, an application performing simultaneous location and mapping (SLAM) only requires the camera images to locate the user within his environment, and data from the IMU and the GPS may only slightly refine this estimation.
5. *Connection metadata*. *Critical data* and *high priority*. Connection metadata is primordial for the operation of the application. Data should reach its destination at all costs, preferably on time. If connection metadata has to be interrupted, it is preferable to shut down the application.

According to these permissions, we can see that sensor data and P-frames are the primary adjustment variables, followed, would the need arise, by the golden frames. I-frames and connection metadata should be transmitted at all times. We consider the three following scenarios, noted 1, 2, and 3 on Fig. 11.4:

1. In the absence of losses, the congestion window grows linearly. The algorithm progressively sends more data, following the priorities of the traffic flows mentioned above. I-frames and connection metadata are sent at all times, while the quality and amount of P-frames and golden frames increase progressively. Once the connection can send all the data from the I-frames, golden frames, and P-frames, it starts sending sensor data.
2. A loss happens. Similarly to TCP, the congestion window is halved, requiring to select the data to send depending on the predefined priorities. The algorithm keeps prioritizing the I-frames and connection metadata while throttling the flow of golden frames and P-frames.
3. Two consecutive losses happen, following a sudden drop of the available bandwidth on the link—for instance, another connection starts sharing the link. The algorithm keeps transmitting the I-frames and connection metadata while throttling the golden frames. The available bandwidth is initially too low to keep transmitting the P-frames or the sensor data. In practice, in such a scenario, the algorithm may request the application to reduce the quality or resolution of the I-frames and golden frames to keep transmitting some P-frames, allowing the user to experience a seamless, yet degraded, experience.

The congestion control algorithm continually monitors the network conditions and reacts with changes in the observed bandwidth and latency. Contrary to traditional congestion control solutions, we cannot consider losses as the sole indicator of congestion. Abrupt variations in delay and may also signal a change in the network conditions that impacts the transmission of some of the sub-flows. Besides, due to the large amount of data transmitted on the uplink, it is necessary to strike a balance between optimizing the link utilization without disturbing other connections on the downlink. The congestion control algorithm should, therefore, compromise between bandwidth, latency, and fairness.

11.4.3 Enforcing Low Latency with Loss Recovery

Recovery from losses is a costly process that may interrupt a connection for several round-trip times. Typical TCP variants wait for the reception of three duplicate acknowledgments before retransmitting the packet. The complete retransmission process can take up to three round-trip times (RTT). In the introduction of this chapter, we considered the maximum acceptable latency to be 33 ms for a 30 FPS video flow and 17 ms for a 60 FPS video flow. In case of occasional losses, it may be possible to push these boundaries to 66 ms and 34 ms, respectively. However, even such values require a very low round-trip time to perform retransmission. If we consider two RTT as the average retransmission time, retransmission can happen only when the average RTT is below 33 ms and 17 ms, respectively. However, as we showed on Table 11.1, such latencies require ideal conditions, and cannot be obtained if the server is located too far away from the user, or if the access link's latency is too high. Recovering from a loss would thus lead to a noticeable choppiness in the display, which would severely degrade the user experience.

However, not all traffic is sensitive to losses. As we already discussed in the previous sections, for some types of flows such as audio, video, or sensor data, new data is often preferred to retransmission. This is particularly true for P-frames, which are not used to decode any further frame in the flow. On the other hand, loosing too many of such frames would also result in choppiness and thus further degrade the user experience. An alternative option consists of introducing redundancy in the flow so that data can be recovered without retransmission. Network coding and forward error correction are popular techniques used in the lower layers of the OSI model to protect data during wireless transmission. Nevertheless, several studies successfully introduced this concept to the transport layer [92, 93]. Multipath transmission may also allow replicating information over multiple links in order for data to reach its destination during a potential link failure [94,95]. Nonetheless, most of these approaches add a notable amount of data to the connection in an environment where resources are scarce. Once again, there is a need for balancing bandwidth, latency, and fairness toward other connections.

11.4.4 Multipath

MAR applications present stringent constraints in terms of available bandwidth and low latency. On the other hand, current wireless networks are highly variable. Wi-Fi networks generally present a high bandwidth and low latency in controlled environments but are not available everywhere. Mobile broadband networks, on the other hand, are almost always accessible in urban environments, at the cost of much more unreliable bandwidth and latency. As such, it may not be possible to attain the desired QoE by exploiting a single type of network. Using multiple paths at the same time would allow for significantly improved performance.

Using 4G and Wi-Fi in parallel would present multiple benefits. First of all, by combining the bandwidth of both technologies, we can significantly increase the resulting available bandwidth. Another exciting characteristic of using multiple paths is the possibility to replicate sensitive data over both links, thus ensuring the integrity of data while preventing costly retransmission. Finally, using two links also resolves the problem of temporary link unavailability or high-latency episode. In this scenario, only one link is used at a time, while the other one acts as a failover to select the link with the best conditions. A new derivative of this property is the possibility to handle handovers [96]. When one of the networks switches between two access points, the application can instantaneously use the other network until the situation resolves, thus efficiently negating the handover time.

However, 4G networks are far from optimal for multipath transmission. Not only is there a severe imbalance between Wi-Fi and 4G in terms of both latency and bandwidth, but also, 4G plans are expensive, and often constrained by the amount of data transmitted over some time. 5G may partly solve this issue by presenting latencies and bandwidth similar to current days' Wi-Fi networks while significantly inflating the limits of data plans. Meanwhile, we cannot yet consider mobile broadband networks as the core of a multipath solution. Therefore, we consider using mobile networks in the following scenarios:

1. Using the Wi-Fi network only. For users that wish to experience AR in a mostly sedentary environment, with occasional interruptions due to handovers.
2. Using the mobile broadband network only for handover while relying mainly on Wi-Fi. Such a scenario allows users with constrained mobile plans to experience MAR applications in places where the Wi-Fi access point density is high enough.
3. Using the mobile broadband network for handover and redundancy of critical data while relying primarily on Wi-Fi. It allows for a more reliable connection while minimizing the usage of the mobile network.

4. Using the mobile broadband network when Wi-Fi is not available. A balanced approach to provide a seamless MAR experience while minimizing mobile broadband usage.
5. Using the mobile broadband network and Wi-Fi in parallel. Such a behavior significantly improves the overall performance.

Such scenarios allow the user to finely select his preference for mobile network usage while enabling users with unlimited plans or 5G phones to enjoy the increase in performance brought by multipath networking fully.

11.4.5 Multi-Server and Distributed Computations

Using multiple paths allows us to increase the network capacity significantly. A logical evolution of this paradigm would be to use multiple servers to reduce the task completion time further [37]. Such a system can take multiple forms, from transmitting on a single link to multiple servers located in the cloud to complex scenarios involving companion devices, edge servers, and cloud servers. Figure 11.5 displays several examples of how such multi-server offloading can take place as an extension of the edge computing paradigm.

Figure 11.5a represents the most straightforward scenario for multipath, multi-server communication. The edge computing paradigm often considers servers located either one hop away from the client or at least within the same locality in the operator's network. When using Wi-Fi and a mobile broadband network in parallel, it is improbable that a single server can be close enough on both links. A solution to this issue consists in using a different edge server for each link. However, such an architecture raises the question of the dependency between the tasks distributed between the two servers. This concern can be addressed either by designing the MAR application to offload independent tasks on the different links or by exploiting the fact that the latency between the two edge servers is most likely lower than the latency on the wireless link, thus allowing for in-time synchronization between the servers. However, the latter solution is not possible in every case. For instance, many institutions present a firewall between their own network and the Internet, adding significant latency to the inter-server synchronization. Interconnection agreements between operators may also reduce the available bandwidth or increase the latency between the servers. Finally, the complexity of synchronization exponentially increases with the number of servers involved. Device-to-device () communication is another challenging paradigm. D2D offloading displays the shortest latencies, at the cost of limited computation capabilities. However, in the case of heavily constrained MAR hardware such as smartglasses, even a simple companion smartphone can significantly increase the computation capabilities of the system. Combined with traditional edge and cloud computing, D2D communication allows for multiple applications, as presented in Fig. 11.5b–d. In Fig. 11.5b, we depict how a user may use his laptop and smartphone as companion devices connected to the home Wi-Fi while connecting to a cloud server through both Wi-Fi and LTE. Meanwhile, Fig. 11.5c, d respectively display examples of how LTE Direct and Wi-Fi Direct can

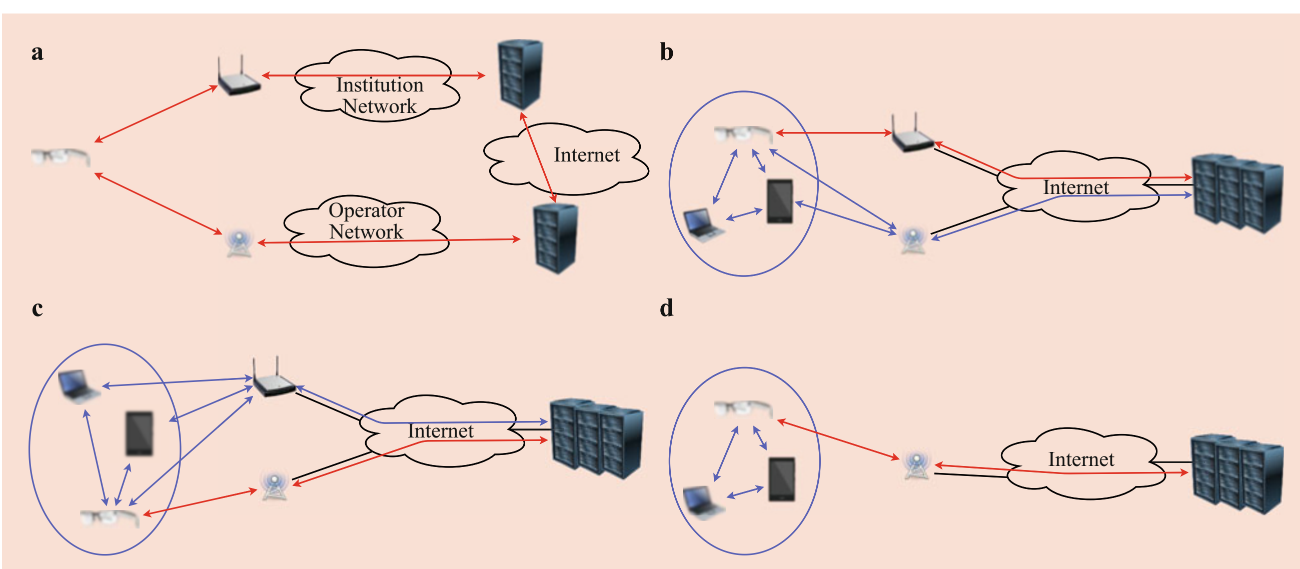


Fig. 11.5 Different scenarios involving offloading computation to multiple resources in parallel. (**a**) Multiple edge servers. (**b**) D2D with cloud computing. (**c**) D2D using LTE Direct and cloud computing. (**d**) D2D using Wi-Fi Direct and LTE connection to the cloud. The AR headset is heavily hardware-constrained, and offloads to companion devices, nearby devices, edge servers, and cloud servers in parallel

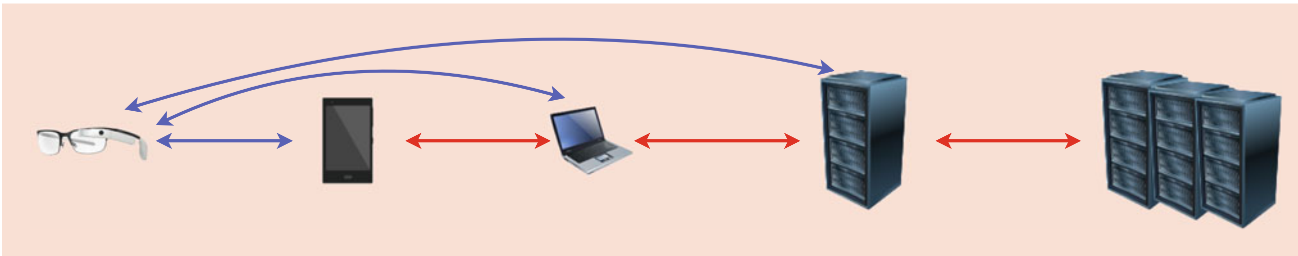


Fig. 11.6 Chaining computation between servers

be used in conjunction with cloud computing for improved performance. Besides the use of companion devices, it may be possible to combine the computational power of the local community to distribute the computations.

Finally, an exciting direction for multi-server computing consists of chaining the servers. As nearby devices and edge servers may be shared with multiple users and applications without presenting the elasticity of the cloud, they may not be able to provide the necessary computing power to complete the task in time. However, the server may prioritize tasks depending on their deadlines, and offload tasks with the farthest deadlines to a server located further in the network. We represent such an architecture in Fig. 11.6. Although the smartglasses can offload tasks to every machine individually, a given machine may offload the computation one or several steps beyond the network.

11.4.6 Security and Privacy

MAR applications continuously generate audio and video flows. As such, they pervasively sense the surroundings of the user. When offloading such data to remote machines, whether edge, cloud, or even nearby foreign devices, the application exposes the user to severe privacy risks. We consider three major threat scenarios:

1. Data intercepted by an attacker eavesdropping on the network.
2. Data gathered by a malicious device in a D2D scenario or a compromised server.
3. Exploitation of the data by the service provider for profiling, advertising, or surveillance.

In order to prevent interception of the data by a malicious user eavesdropping on the network, end-to-end encryption is critical. Besides encryption, it is necessary to anonymize the data to prevent recovery if an element in the transmission is compromised. Every element leading to recovering personal information on the user needs to be removed. However, anonymization may significantly lower the utility of the application. For instance, an AR recommender system may not function without fine localization information on the user.

Many studies focus on confidential information leaked by digital pictures and videos. `PrivateEye` and `WaveOff` [97] offer the possibility to annotate areas on the captured picture before transmitting the picture to the application. The application will only see the areas the users explicitly marked as safe. `PrivacyTag` [98] uses visible tags to automatically blur users' faces depending on the community they belong to. `Cardea` [99] builds on top of this concept by enabling users to define contexts in which they agree for their picture to be taken while defining gestural overrides. Similarly, `I-PIC` [100] enables the user to define the visual information they are willing to transmit through a secure protocol. Finally, `Vanish` [101] uses the vanishing property of data in a hash table to transmit images. After a certain amount of time, the images become unreadable.

Besides directly protecting the images, architectural choices may also significantly increase the privacy of the users. Federated learning allows the users to propagate only model updates to the server instead of the whole data. However, federated learning is a complex task that requires substantial computation power. Edge computing can also partially increase privacy, as the data remains on the edge server, instead of being centralized on a cloud server.

However, all the solutions listed above limit the utility of the data to some extent. It is, therefore, necessary to use them parsimoniously and a combination of techniques depending on the data, and the type of computation is the most likely scenario. Overall, we need to find a trade-off between application functionality and privacy using an inherently privacy-affecting technology.

11.4.7 Implementation Notes

The protocol described in this section can be implemented on top of UDP. Such an implementation not only makes the protocol easier to embed within any MAR application as a library but also allows more active dialogue between the application and the protocol by directly providing the primitives and interfaces. On the other hand, integrating the protocol directly within the kernel, as it is the case for traditional transport protocols, may be interesting. By incorporating the protocol within the kernel, one may avoid the latency induced by the communication between the application layer (in the application) and the transport layer (in the kernel). Besides, it allows optimizing the algorithms at a lower layer. However, for mobile devices relying, for instance, on Android, developing a kernel module for MAR transmission severely reduces the potential devices on which the protocol can be deployed. Indeed, modifying the kernel can only be done by the phone's manufacturer without significant effort from the user.

Another modification to consider within the kernel concerns the network buffers. The mobile network's uplink buffers are generally largely oversized, in the order of thousands of packets, thus significantly impacting the latency. As MAR applications heavily rely on large volumes of latency-sensitive data, such design choice can considerably affect the application's performance. Such a phenomenon can be avoided by employing low-latency queuing strategies such as FQ CoDel [91]. However, as much as such strategies improve general performance [102], they may also delay or starve longer flows, among which MAR application flows. Developing an appropriate queuing policy favoring MAR traffic while protecting connections coming on the downlink, for instance, by prioritizing acknowledgments, is thus the next logical step for improving MAR performance at the network layers of the OSI model.

11.5 Improving Latency at the Application Layer

In this section, we look at typical mobile augmented reality systems and current scenarios for cloud offloading of parts of these systems. We examine two commercial AR SDKs (SDK-A and SDK-B). More particularly, we focus on their offloading procedures, including looking at the latency of various subtasks (e.g., preprocessing, detection) for several AR tasks. We finally conclude by discussing recommendations for providing high performance and high QoE for mobile AR tasks.

11.5.1 Mobile AR Pipeline

Figure 11.7 illustrates a standard mobile AR system pipeline. The initial stage is *object detection* whereby objects in the phone camera perspective are detected, and regions of interest (ROI) related to each object are identified. In the subsequent stage, *feature extraction* is applied to each ROI so that relevant feature points for each target are collected. Finally, *object recognition* is utilized to identify the object based on a collection of original images of targets in a database.

Given an original frame and a present frame, the stage of *template matching* certifies the recognition of the object and determines the target pose. The target pose is then inputted to *object tracking*, which helps avoid costly object recognition in every frame. Finally, the *annotation rendering* stage displays the content associated with the recognized target (from the database). For illustration, in Fig. 11.7, the color of a task refers to the computational demand (i.e., darker denotes heavier computation), with demand meaning the time taken to finish a cloud or mobile device task. For instance, feature extraction takes far less time than object recognition given the same compute resources.

For the mobile devices of today, computing power and battery size are significant restrictions for the performance of many tasks. To avoid these restrictions, AR systems can offload some tasks (from those previously described) to the cloud. Figure 11.7a, b denote two typical scenarios for cloud offloading (i.e., lines denote the separation of tasks performed on cloud and mobile). In scenario one, the images are directly sent for object detection [6] to the cloud, whereas in scenario two, the mobile device performs the object detection and feature extraction. As such, the device only sends the feature points for the server to perform object recognition [19] to the cloud. In an additional scenario, all tasks can be performed on the mobile device (using a limited local database), and if the object recognition is unsuccessful, then apply one of the two prior scenarios (with a more extensive cloud database) to complete the recognition.

11.5.2 Commercial SDK Cloud Offloading Procedures

We analyze the cloud offloading procedures of two commercial AR SDKs (SDK-A and SDK-B) by using a man in the middle (MITM) proxy (Fiddler [103]) to intercept and then decrypt traffic between the device and the cloud. This MITM proxy technique is necessary because the source code of the SDKs is not openly available. The Android demonstration application of SDK-A and the Unity Android-based demonstration application of SDK-B are employed for the experimentation.

An examination of MITM network traces allows for the illustration of, for example, the cloud-based object recognition procedure of SDK-A in Fig. 11.8a. The procedure consists of a single HTTPS session with a request that includes a camera view frame in JPEG format with gray scale and 640 × 360 resolution (compared to a camera resolution of 1920 × 1080). The server then returns the recognized frame from the database in JPEG format with gray scale, 320 width, and varying height. The Android app, after getting the recognized image, then performs feature extraction and template matching for pose identification. We note that

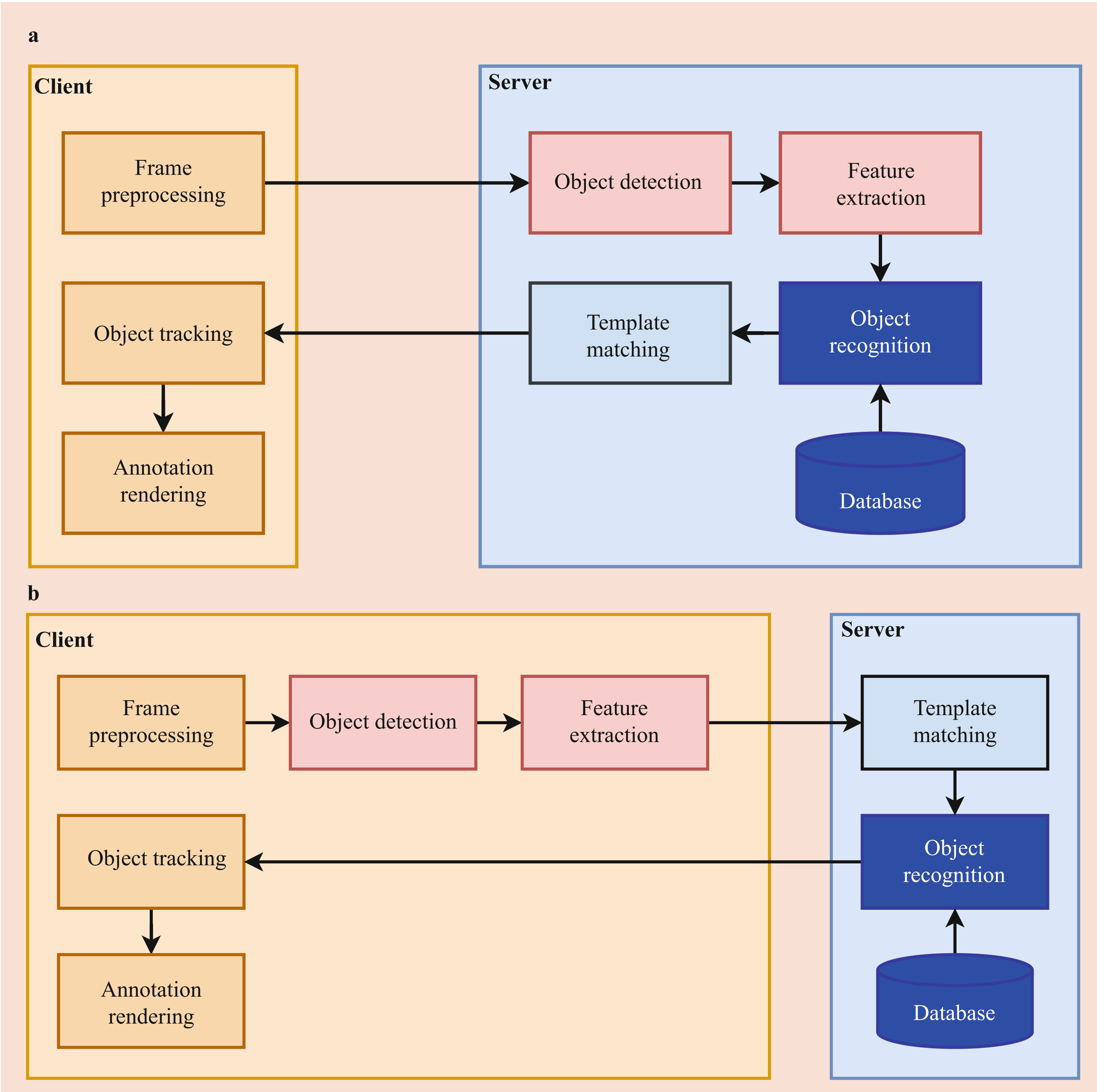


Fig. 11.7 Pipeline for common offloaded mobile AR systems. (**a**) Scenario 1: the smartphone offloads most of the operations to the cloud. (**b**) Scenario 2: the smartphone offloads only the heaviest computations to the cloud, and transmits feature points instead of full images. Object detection, feature extraction, object recognition, and template matching are the most demanding tasks

at the time, SDK-A did not support the storage of annotations, and thus this part is included in the Android app itself.

For SDK-B, the analogous procedure is also illustrated in Fig. 11.8a. The procedure consists of two HTTPS sessions for image detection, followed by a single HTTP session for annotation. The first HTTPS session request includes a similar JPEG as SDK-A but with color instead of gray scale, while the response contains a URL to the original frame, a key for a cloud database containing feature-point sets, and the annotation content filename. The following HTTPS session request includes the key to the set of feature points (the feature extraction is not performed on the client like in SDK-A). The server then uses template matching for pose identification. The last HTTP session request includes the annotation for rendering on the client.

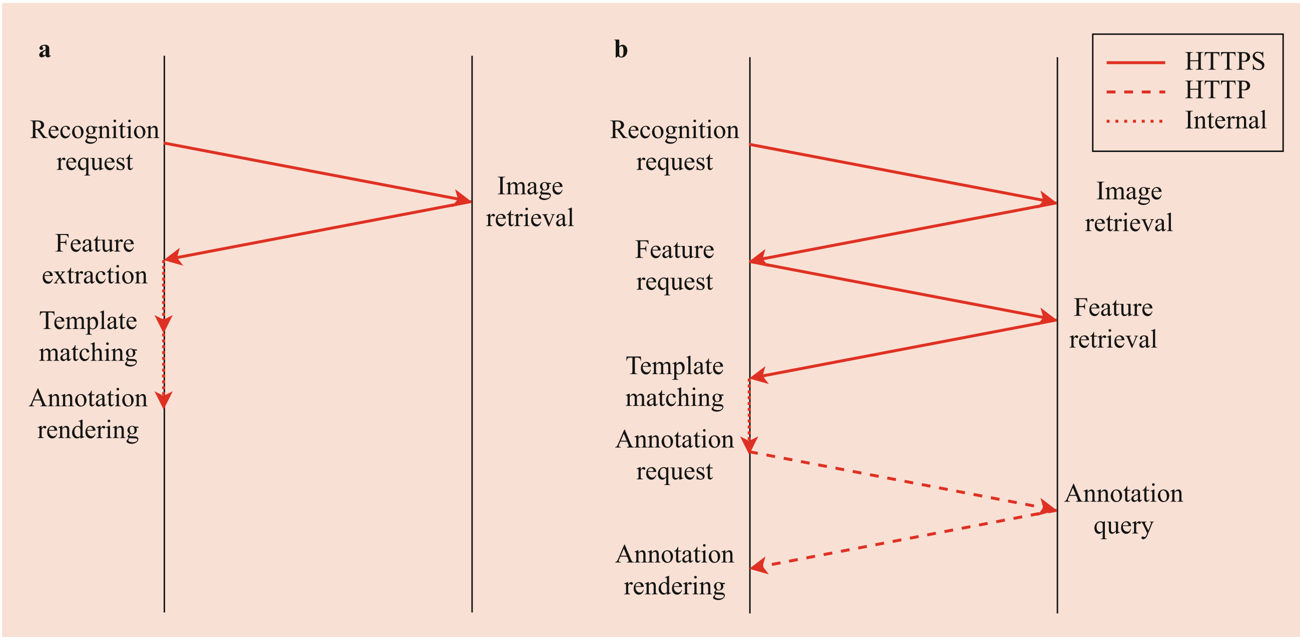


Fig. 11.8 Offloading process of SDK-A (**a**) and SDK-B (**b**). SDK-A uses a single HTTPS request, while SDK-B uses two HTTPS requests followed by a single HTTP request

Table 11.3 Experimental performance of SDK-A and SDK-B and standard deviation. For SDK-B, values include only the two HTTPS sessions for a fair comparison

SDK	Upload		Download		Latency (ms)	Energy (mAh)
	Volume (KB)	Type	Volume (KB)	Type		
SDK-A	16.3 ± 2.0	Image	33.5 ± 8.2	Image	1410 ± 14	0.45 ± 0.04
SDK-B	26.6 ± 2.6	Image	160.0 ± 55.0	Features	$19,320 \pm 5280$	7.91 ± 2.1

A third commercial SDK known as SDK-C is also examined, but as no trial license exists, the information is only extracted from documentation of the API. The procedure uses only a single HTTPS session like SDK-A.

Fine-Grained Measurement Study

Next, we analyzed the data use, latency, and energy consumption of the SDK-A and SDK-B systems. Unfortunately, analyzing, for example, Overlay [6], VisualPrint [19], or Glimpse [5] is not possible as these systems are not yet open source.

The analysis procedure included the following steps and notes. First, 100 movie posters were collected from an online source, uploaded to SDK-B and SDK-A clouds, and the image databases registered to the respective applications. Then we verified that all posters could be recognized correctly by both apps. Packet traces were then collected at the MITM proxy for analyzing upload and download data usage, and a Monsoon power monitor was used to analyze the energy usage of the procedures on the mobile device. We also note that duration between capturing a camera frame and the annotation rendering is defined as the end-to-end latency (also motion-to-photon latency).

In terms of results, Table 11.3 details the traffic, latency, and energy results. Firstly, we note that SDK-A uses less data than SDK-B; this is because SDK-A uses gray scale instead of color and the SDK-B feature-point set is much larger and not optimal for network transfer (i.e., even larger than the original image of mean size 73.56 KB). Second, the end-to-end SDK-A latency is also much smaller than SDK-B (though both systems are significantly above the inter-frame duration for 30 FPS (33 ms)). The SDK-B cloud processing is slower partly because the initial SDK-B request is not sent immediately by the application after triggering but instead is delayed by at least a few seconds. Besides, latency increases with every new object SDK-B detects. Unfortunately, confirming this hypothesis is not possible because SDK-B is closed-source.

Also, the RTT between the server of SDK-A and the mobile client is significantly less than the similar RTT for SDK-B (48 vs. 162 ms). The measurements were taken from the client in Hong Kong to SDK-A servers (in Singapore) and SDK-B servers (in Shanghai). In other words, even though the physical distance to the SDK-A server is significantly larger, the RTT is less than for SDK-B. This observation emphasizes that deploying compute resources in locations with low RTT values for users is essential.

Additionally, SDK-B uses an additional HTTPS session compared to SDK-A, thus adding extra latency. This high overall latency is a big cause of the more significant energy usage of SDK-B versus SDK-A (7.91 versus 0.45 mAh).

Takeaway: *Cloud-based commercial SDKs are relatively immature with little optimizing of latency and data traffic (in free versions).*

11.5.3 Commercial SDK End-To-End Latency Analysis

We now analyze the time required for the different tasks from Fig. 11.7 with the help of our prototype cloud-based mobile VR implementation. Specifically, in the prototype, the computer vision functions use OpenCV,[1] the client uses a custom android app on a Hong Kong-based Xiaomi MI5, and the server uses a custom C++ program on a Taiwan-based Google Cloud VM (8 vCPUs with 2.5 GHz and 32 GB memory). The implementation of each task is further described below. We note that the tasks were also briefly detailed in Sect. 11.5.1.

Frame preprocessing downscales size of the frame from the camera from 1920×1080 to 480×270 pixels, converts to gray scale color space, and encodes the frame as a JPEG.

Object detection finds the regions of interest for potential targets within the frame. In the implementation, a sliding window runs through the frame. The system then calculates the gray scale pixel variance within the window. Areas with high variances are assumed to be foreground textures and are extracted as regions of interest.

Feature extraction should then select features from the regions by applying ORB [104]. However, the ORB algorithm is not yet implemented on the client. Therefore, features are instead selected from the entire frame to ensure a good comparison.

Object recognition is an image recognition process that finds the most similar image to the query image in an image database. For the process setup, first, a Bernoulli mixture model (BMM) is created offline from the features from images from the dataset (the images from section "Fine-Grained Measurement Study"). Applying the BMM allows us to encode each image's features into a Fisher vector, which is then stored with locality-sensitive hashing (LSH) for quick access. Recognition then uses the nearest neighbor search in the LSH hash table to find the most similar image.

Template matching certifies the recognition of the object and determines the target pose in 6DOF (six degrees of freedom). A brute force method using the features from the database image and the camera image finds the matching feature pairs. Using these feature pairs, it is possible to determine the target homography within the view of the camera (i.e., a 2D change of position). *Object tracking* tracks a given object between frames using its feature points, which helps avoid costly object recognition in every frame. The lightweight optical flow [105] object tracking method is used to allow real-time processing.

Annotation rendering recalculates the object's 6DOF pose, determines the pose of the annotation, and renders the 3D annotation.

Table 11.4 details the analysis results as a mean over 100 trial runs. We do not report the results for annotation rendering, object tracking, or image preprocessing on the cloud as these tasks should be executed on the mobile device for logistical reasons. We do report the results for template matching, object recognition, feature extraction, and object detection on both mobile and cloud as these tasks are offloadable.

Results-wise, object recognition is the most compute- and storage-intensive task (in the server case 58% of the time for the four offloadable tasks). For mobile devices, according to [106], object recognition takes longer than 2 s on mobile CPU (Snapdragon 800). Therefore, with mobile device energy considerations, object recognition should take place on a cloud server (in the near term, at least).

The object detection task finds regions of interest. As such, it significantly reduces the area on which feature extraction and object recognition are performed. In the server case, the object detection time is short but non-negligible at about 11.49 ms on average. Unfortunately, object detection is not yet implemented on the mobile client, but we expect the time to be significantly longer. For feature extraction, the mobile device takes about three times longer than the cloud (131 ms vs. 47.41 ms).

The template matching time on the cloud is also substantially less than for the mobile device (9.22 versus 143.1 ms). This is partly because the cloud only needs to transfer simple pose data to the mobile device rather than the original image. Unfortunately, this pose data may be out of date due to offloading latency as the server pose is from before the object recognition request. Additionally, if the object tracking on the mobile client fails, then recovery is not possible (as the original image is not available). The SDK-A and SDK-B systems

Table 11.4 Latency analysis of distinct AR tasks. Values succeeding ± are standard deviations

AR task	Mobile (ms)	Cloud (ms)
Image preprocessing	34.7 ± 11.0	N/A
Object detection	N/A	11.5 ± 0.4
Feature extraction	131.0 ± 43.8	47.4 ± 1.2
Object eecognition	>2000	92.4 ± 19.5
Template matching	143.1 ± 31.7	9.2 ± 4.1
Object tracking	15.1 ± 6.3	N/A
Annotation rendering	19.3 ± 0.5	N/A

[1] http://opencv.org/.

use client-based template matching to improve robustness against such failures.

The network latency under various network conditions is also measured with the lowest latency on a nearby testbed (over LTE) of about 14 ms. Combining the network latency with the processing latency of the prototype system, the end-to-end latency for the prototype system is a minimum of 244 ms.

For visualization purposes, Fig. 11.9 illustrates the components of end-to-end latency in a percentage-based pie chart with the assumption of a reasonable network latency of 50 ms and offloading to the cloud template matching, object recognition, feature extraction, and object detection. Interestingly, even with offloading object recognition and feature extraction, these tasks dwarf other tasks in terms of latency.

Takeaway: *Mobile cloud AR delay is significantly larger than the inter-frame interval for 30FPS of 33 ms, and object recognition is a major part of that delay. However, there are still multiple opportunities to improve the performance and experience of mobile AR.*

11.5.4 Discussion

This section details the potential to improve the QoE and performance of systems for mobile AR.

What to Offload? As discussed prior, when offloading to the cloud, the offloaded tasks can start at the object detection task (thus transmitting images) or the object recognition task (thus transmitting feature points). Advantages and disadvantages exist for both approaches.

For transmitting images, there are also two options: video (e.g., H.264) or single frames. Video can utilize inter-frame compression in addition to the inter-frame compression used by both approaches, therefore providing higher-quality frames with the same data use and enhancing recognition accuracy. However, this entails buffering frames at the mobile device, therefore adding latency.

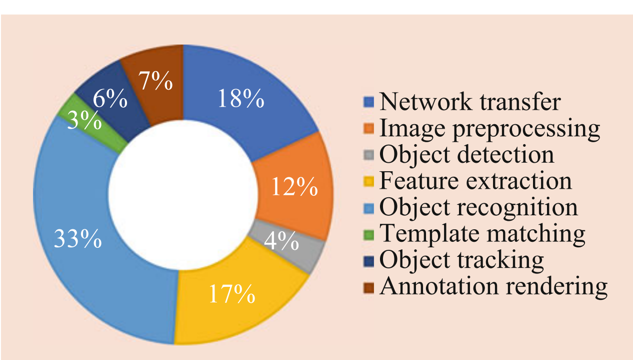


Fig. 11.9 Components of end-to-end latency as percentages with an assumption of network latency of 50 ms and offloading template matching, object recognition, feature extraction, and object detection to the cloud

For transmitting feature points, there are similarly two options: raw feature points or feature point fingerprints. Raw feature points may be much larger than the original image [19] but are readily available for transmission, whereas fingerprints (like in VisualPrint [19]) are smaller, but fingerprint generation requires computation which increases energy and latency.

The overall discussion illustrates that there are clear trade-offs between recognition accuracy, energy consumption, data usage, and latency in such mobile AR systems.

Where to Offload? Tasks can be offloaded to several different types of nodes, including the remote cloud, edge cloud [25, 107], and even proximal mobile devices. The remote and edge cloud both have significantly more compute power than the mobile device, but the edge cloud has lower latency to the mobile device. However, the edge cloud likely needs modifications to the access and core network architecture to function. Besides, edge servers do not have the same elasticity as cloud datacenters, and the amount of resources is more likely to be limited within a given location. As such, sharing edge computing resources fairly with other users while guaranteeing some degree QoS is going to be one of the major challenges in the future. Offloading to proximal mobile devices is also feasible but raises many questions regarding computation latency, security, privacy, and the general selfishness of users in peer-to-peer systems.

Push the Limit of AR Latency The overall analysis suggests that offloading object recognition to the cloud is a worthwhile idea. Unfortunately, the tasks are still computationally expensive on commercial servers with total task time over 150 ms. This time is the most significant component of the end-to-end latency when offloading AR to the cloud. A major next step would be to leverage GPU execution on cloud or edge servers to accelerate the tasks and significantly lower this task latency.

11.6 Conclusion

In this chapter, we aimed at bridging the gap between the constrained hardware of mobile AR devices and the ever-increasing demands of AR applications. With motion-to-photon latency being one of the most critical parameters of the quality of experience, we showed that cyber-foraging, the action of offloading computation-heavy tasks to remote devices and machines, can significantly improve the performance of MAR applications, but can also negatively impact Quality of Experience if performed haphazardly. After discussing the latest advances in wireless networks, we proposed to tackle the stringent requirements of offloading-enabled MAR applications through examinations at almost every layer of the OSI model. We learned the following lessons:

11.6.1 Access Link

Our review of current wireless networking technologies showed that their capabilities are too limited to provide a truly seamless mobile AR experience. However, some systems can partly address the QoS requirements. For instance, Wi-Fi is characterized by high bandwidth and low latency, which can enable real-time MAR computation offloading. On the other hand, though, Wi-Fi access points are scattered, with Internet connectivity rarely available outside of controlled environments. LTE has very high availability in urban areas, thanks to dense deployments of base stations with fast handovers. However, LTE is reaching the end of its life cycle and has bandwidth and latency constraints that are inappropriate for real-time AR. Therefore, 5G may represent a breakthrough for applications requiring high bandwidth with strict latency constraints such as MAR. According to the currently available technical reports, 5G should provide bandwidth, latency, and coverage, enabling a complete seamless MAR experience. However, we noted that similarly to LTE, the actual implementation of 5G might differ from white papers. Besides, we predict that usage will eventually catch up to 5G capabilities, and thereafter performance may suffer. Specifically, we showed that MAR can theoretically require much more bandwidth than 5G can offer to unleash its full potential. Based on these observations, we concluded that *cyber-foraging for MAR cannot rely on an increase in performance of the physical layer alone and must combine more efficient wireless networks with clever protocols at the network, transport, and application layers*.

11.6.2 Network and Transport Layer

We then defined guidelines for designing a MAR-compliant transport layer protocol. We based these guidelines on the idea that *a seamless, yet degraded, experience is preferable to a choppy experience.* We thus proposed to introduce graceful degradation to the protocol. Therefore, the protocol could address the diversity and constraints of MAR data flows while preserving fairness for other connections. To develop the protocol, we first defined internal QoS classes to differentiate the different sub-flows within MAR traffic. Such classes define both the priority and the loss recovery strategy for a given flow. Given these classes, we designed a congestion control algorithm and several loss recovery approaches for low-latency MAR communication. We also discussed how multiple paths and distributed cyber-foraging can be used in conjunction with such an algorithm for enhanced performance in modern networks. Finally, we gave recommendations on queuing policies to avoid common congestion issues that can considerably increase latency. *By integrating graceful degradation at the network and transport layers, we provide a framework for low-latency transmission of the most critical information, even in degraded conditions.*

11.6.3 Application Layer

When designing a transport protocol for MAR applications, we realized that optimal performance can only be achieved through constant communication between the application layer and the lower layers. Indeed, many decisions need to be taken by the application. Some decisions directly impact the defined QoS parameters, such as the video transmission quality, and must be the result of common agreement between the transport and application layers. Additionally, more generic decisions may also significantly impact the overall latency. We analyzed two standard SDKs and showed that different design choices lead to significantly different performance. Specifically, some tasks are better suited for execution on mobile devices to avoid costly network transmissions. For instance, feature extraction can be executed on-device in a reasonable amount of time and significantly reduces the subsequent network payload. On the other hand, some other tasks can only be performed within a virtually limitless resource cloud environment. Object recognition, for instance, takes more than 2 s on our experimental mobile device, but less than 100 ms in the cloud. Based on this knowledge, we discussed what tasks should be offloaded, and where, in order to considerably shrink the latency in MAR applications. *With the increasing computing power of devices, many smaller tasks can be executed on-device, including tracking, mapping, and positioning. Only the most computation-intensive operations such as object recognition require offloading, in the cloud or at the edge of the network.*

11.6.4 Future Challenges

Cloud and edge offloading for MAR is still a very active field. The stringent motion-to-photon latency requirements should be addressed at every layer, from faster and more reliable communication technologies to latency-compensation techniques at the application level to mask the effect of latency to the end user. In particular, we identify the following challenges and research directions:

Exploring Inconspicuous Markers Object recognition is a computationally intensive process. On the other hand, it is possible to develop marker-based solutions that seamlessly integrate in the environment. Visible light communication (VLC), for instance, allows integrating the market in the status LED of a device, or in the lighting of an object. However, VLC is sensitive to interference [108], and medium access strategies need to be developed for dense deployment [109]. Similarly, wireless strategies can also be used to locate objects. For instance, near-field communication allows batteryless communication at a short distance, allowing to detect objects in the immediate vicinity of the user [110]. Newer technologies such as backscatter communication can be used for more direct communication between objects and user devices [111].

Fluid Computing Resource Usage With the advent of D2D communications and fog computing, multiple computing resources presenting various characteristics are scattered between the user device and the more traditional computing resources in the cloud [112]. These resources are available with much lower latency, but also more limited computing power and energy. It is thus necessary to strike a trade-off between computing latency and network latency. Another aspect is leveraging the power of nearby idle devices operated by other users through D2D communication. Such operation raises concerns related to privacy, user compensation, and fair use of the user resources [113].

Collaborative AR Offloading object recognition tasks to the edge increases the efficiency of mobile AR applications, but edge servers still face serious challenges in serving large amounts of concurrent user requests. Furthering the idea of leveraging D2D communication, it is possible to envision collaborative AR applications [114]. In such applications, active users in close proximity share their computing power in order to spread the task of detecting or mapping a space among multiple devices, reducing the latency of the task [115]. Once the environment has been mapped, and objects have been recognized, devices only need to perform the space transformation to adapt to their own point of view. Such a strategy has been applied to share vehicular vision in safety applications [116], but latency remains high, and scaling to many users may result in information overload.

Large-Scale Applications To this day, AR solutions are developed for small-scale applications, involving indoor environments, and a finite number of objects to detect. However, visions for AR include larger-scale environments, up to city-wide AR. As such, one of the biggest challenges of future AR applications will be to leverage spatiotemporal context awareness in order to minimize the number of objects to detect and focus on content retrieval and rendering. Few studies have considered such a large-scale environment [117], and much work remains to be done in adapting the algorithms for large-scale environments.

Latency Compensation Whether the result of high computation times, low available bandwidth, or even transient increase in latency over a wireless link, high latency is unavoidable. In such cases, it is critical to minimize its impact on the user. AR is a relatively novel technology, and the research community is only starting to study and understand the effects of latency on users [118, 119]. Figuring out compensation strategies to account for the different sources of latency in AR application is one of the next significant steps for providing a seamless experience to users.

With the recent developments in communication and computation technologies, MAR is progressively moving from the proof-of-concept stage to a fully enabled technology, deployed at a larger scale. With latency being at the core of the user experience, it is critical to address the challenges mentioned above to support this evolution.

Acknowledgments This research has been supported in part by project 16214817 from the Hong Kong Research Grants Council, the 5GEAR project, and the FIT project from the Academy of Finland.

References

1. Chatzopoulos, D., Bermejo, C., Huang, Z., Hui, P.: Mobile augmented reality survey: From where we are to where we go. IEEE Access **5**, 6917–6950 (2017)
2. Höllerer, T., Feiner, S.: Mobile augmented reality. In: Telegeoinformatics: Location-Based Computing and Services, vol. 21. Taylor and Francis Books Ltd., London (2004)
3. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graph. Appl. **21**(6), 34–47 (2001)
4. Huang, Z., Hui, P., Peylo, C., Chatzopoulos, D.: Mobile augmented reality survey: a bottom-up approach. arXiv preprint arXiv:1309.4413 (2013)
5. Chen, T.Y.-H., Ravindranath, L., Deng, S., Bahl, P., Balakrishnan, H.: Glimpse: Continuous, real-time object recognition on mobile devices. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, ser. SenSys '15, pp. 155–168. ACM, New York (2015). http://doi.acm.org/10.1145/2809695.2809711
6. Jain, P., Manweiler, J., Roy Choudhury, R.: Overlay: Practical mobile augmented reality. In: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, pp. 331–344. ACM, New York (2015)
7. Shi, B., Yang, J., Huang, Z., Hui, P.: Offloading guidelines for augmented reality applications on wearable devices. In: Proceedings of the 23rd ACM International Conference on Multimedia, ser. MM '15, pp. 1271–1274. ACM, New York (2015)
8. Abrash, M.: Latency—the sine qua non of ar and vr, accessed 23-02-2017. http://blogs.valvesoftware.com/abrash/latency-the-sine-qua-non-of-ar-and-vr/
9. Zhou, P., Zhang, W., Braud, T., Hui, P., Kangasharju, J.: Enhanced augmented reality applications in vehicle-to-edge networks. In: Proceedings of the 2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), pp. 167–174 (2019)

10. Kosta, S., Aucinas, A., Hui, P., Mortier, R., Zhang, X.: Thinkair: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading. In: Infocom, 2012 Proceedings IEEE, pp. 945–953. IEEE, New York (2012)
11. Cuervo, E., Balasubramanian, A., Cho, D.-k., Wolman, A., Saroiu, S., Chandra, R., Bahl, P.: Maui: Making smartphones last longer with code offload. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '10, p. 49–62. Association for Computing Machinery, New York (2010)
12. Patel, M., Naughton, B., Chan, C., Sprecher, N., Abeta, S., Neal, A. et al.: Mobile-edge computing introductory technical white paper. In: White Paper, Mobile-edge Computing (MEC) Industry Initiative (2014), accessed 28-08-2020. https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf
13. Beck, M.T., Werner, M., Feld, S., Schimper, T.: Mobile edge computing: a taxonomy. In: Proceedings of the Sixth International Conference on Advances in Future Internet, pp. 48–54 (2014)
14. The mobile economy 2019, GSMA Intelligence, Tech. Rep. (2019), accessed 28-08-2020. https://www.gsma.com/r/mobileeconomy/3/
15. Chun, B.-G., Ihm, S., Maniatis, P., Naik, M., Patti, A.: Clonecloud: elastic execution between mobile device and cloud. In: Proceedings of the Sixth Conference on Computer Systems, pp. 301–314. ACM, New York (2011)
16. Bermejo, C., Huang, Z., Braud, T., Hui, P.: When augmented reality meets big data. In: Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 169–174 (2017)
17. Satyanarayanan, M., Bahl, P., Caceres, R., Davies, N.: The case for vm-based cloudlets in mobile computing. IEEE Pervasive Comput. **8**(4), 14–23 (2009)
18. AT&T Edge Cloud (AEC)—White Paper (2017), accessed 05-04-2018. https://about.att.com/content/dam/innovationdocs/Edge_Compute_White_Paper%20FINAL2.pdf
19. Jain, P., Manweiler, J., Roy Choudhury, R.: Low Bandwidth Offload for Mobile AR. In: Proceedings of the 12th International on Conference on Emerging Networking EXperiments and Technologies, pp. 237–251. ACM, New York (2016)
20. Hagbi, N., Bergig, O., El-Sana, J., Billinghurst, M.: Shape recognition and pose estimation for mobile augmented reality. In: Proceedings of 8th IEEE International Symposium Mixed and Augmented Reality (ISMAR), pp. 65–71. IEEE, New York (2009)
21. Qiu, H., Ahmad, F., Govindan, R., Gruteser, M., Bai, F., Kar, G.: Augmented vehicular reality: enabling extended vision for future vehicles. In: Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications, pp. 67–72. ACM, New York (2017)
22. Zhou, P., Zhang, W., Braud, T., Hui, P., Kangasharju, J.: Arve: Augmented reality applications in vehicle to edge networks. In: Proceedings of the 2018 Workshop on Mobile Edge Communications, ser. MECOMM'18, p. 25–30. Association for Computing Machinery, New York (2018)
23. Yan, T., Kumar, V., Ganesan, D.: CrowdSearch: exploiting crowds for accurate real-time image search on mobile phones. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp. 77–90. ACM, New York (2010)
24. Rallapalli, S., Ganesan, A., Chintalapudi, K., Padmanabhan, V.N., Qiu, L.: Enabling physical analytics in retail stores using smart glasses. In: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, pp. 115–126. ACM, New York (2014)
25. Ha, K., Chen, Z., Hu, W., Richter, W., Pillai, P., Satyanarayanan, M.: Towards wearable cognitive assistance. In: Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '14, pp. 68–81. ACM, New York (2014). http://doi.acm.org/10.1145/2594368.2594383
26. Yamamoto, H., Ohshima, T., Satoh, K., Tamura, H.: Ar2 hockey: A case study of collaborative augmented reality. In: Virtual Reality Annual International Symposium, p. 268. IEEE Computer Society, Los Alamitos (1998). https://doi.ieeecomputersociety.org/10.1109/VRAIS.1998.658505
27. Szalavári, Z., Schmalstieg, D., Fuhrmann, A.L., Gervautz, M.: "Studierstube": An environment for collaboration in augmented reality. Virtual Reality **3**(1), 37–48 (1998)
28. Reitmayr, G., Schmalstieg, D.: Mobile collaborative augmented reality. In: Proceedings IEEE and ACM International Symposium on Augmented Reality, pp. 114–123 (2001)
29. Regenbrecht, H.T., Wagner, M.T., Baratoff, G.: MagicMeeting: a collaborative tangible augmented reality system. Virtual Reality **6**(3), 151–166 (2002)
30. Henrysson, A., Billinghurst, M., Ollila, M.: Face to face collaborative ar on mobile phones. In: Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05), pp. 80–89 (2005)
31. Habak, K., Ammar, M., Harras, K.A., Zegura, E.: Femto clouds: Leveraging mobile devices to provide cloud service at the edge. In: 2015 IEEE 8th International Conference on Cloud Computing (CLOUD), pp. 9–16. IEEE, New York (2015).
32. Orsini, G., Bade, D., Lamersdorf, W.: Computing at the mobile edge: designing elastic android applications for computation offloading. In: IFIP Wireless and Mobile Networking Conference (WMNC), 2015 8th, pp. 112–119. IEEE, New York (2015)
33. LiKamWa, R., Priyantha, B., Philipose, M., Zhong, L., Bahl, P.: Energy characterization and optimization of image sensing toward continuous mobile vision. In: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '13, p. 69–82. Association for Computing Machinery, New York (2013)
34. Naderiparizi, S., Zhang, P., Philipose, M., Priyantha, B., Liu, J., Ganesan, D.: Glimpse: a programmable early-discard camera architecture for continuous mobile vision. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '17, p. 292–305. Association for Computing Machinery, New York (2017)
35. Mathur, A., Lane, N.D., Bhattacharya, S., Boran, A., Forlivesi, C., Kawsar, F.: Deepeye: resource efficient local execution of multiple deep vision models using wearable commodity hardware. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '17, p. 68–81. Association for Computing Machinery, New York (2017)
36. Huynh, L.N., Lee, Y., Balan, R.K.: Deepmon: Mobile gpu-based deep learning framework for continuous vision applications. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '17, p. 82–95. Association for Computing Machinery, New York (2017)
37. Braud, T., Zhou, P., Kangasharju, J., Hui, P.: Multipath computation offloading for mobile augmented reality. In: Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 1–10 (2020)
38. Speedtest®market report united states, accessed 23-02-2017. http://www.speedtest.net/reports/united-states/
39. Sumits, A.: The history and future of internet traffic, accessed 23-02-2017. https://blogs.cisco.com/sp/the-history-and-future-of-internet-traffic
40. Braden, B., Zhang, L., Berson, S., Herzog, S., Jamin, S.: Resource reservation protocol (RSVP)—version 1 functional specification, Internet Requests for Comments, RFC Editor, RFC 2205 (1997), accessed 28-08-2020. http://www.rfc-editor.org/rfc/rfc2205.txt
41. Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V.: RFC 3550: RTP: A Transport Protocol for Real-Time Applications.

IETF, New York. Technical Report (2003) accessed 28-08-2020. www.ietf.org/rfc/rfc3550.txt
42. Schulzrinne, H., Rao, A., Lanphier, R.: Real time streaming protocol (RTSP), Internet Requests for Comments, RFC Editor, RFC 2326 (1998), accessed 28-08-2020. https://www.rfc-editor.org/rfc/rfc2326.txt
43. H222—Information technology—Generic coding of moving pictures and associated audio information: systems, ITU, Technical Report (2014), accessed 2020-28-08. https://www.itu.int/rec/T-REC-H.222.0-201808-I/en
44. Chatzopoulos, D., Bermejo, C., Haq, E. u., Li, Y., Hui, P.: D2d task offloading: a dataset-based q a. IEEE Commun. Mag. **57**(2), 102–107 (2019)
45. Fu, Z., Meng, X., Lu, S.: A transport protocol for supporting multimedia streaming in mobile ad hoc networks. IEEE J. Sel. Areas Commun. **21**(10), 1615–1626 (2003)
46. Luo, H., Wu, D., Ci, S., Sharif, H., Tang, H.: TFRC-based rate control for real-time video streaming over wireless multi-hop mesh networks. In: 2009 IEEE International Conference on Communications, pp. 1–5 (2009)
47. Ha, S., Rhee, I., Xu, L.: Cubic: a new tcp-friendly high-speed TCP variant. SIGOPS Oper. Syst. Rev. **42**(5), 64–74 (2008)
48. Mascolo, S., Casetti, C., Gerla, M., Sanadidi, M.Y., Wang, R.: TCP westwood: bandwidth estimation for enhanced transport over wireless links. In: Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, ser. MobiCom '01, pp. 287–297. Association for Computing Machinery, New York (2001)
49. Caini, C., Firrincieli, R.: TCP HYBLA: a tcp enhancement for heterogeneous networks. Int. J. Satell. Commun. Netw. **22**(5), 547–566 (2004)
50. Abbasloo, S., Xu, Y., Chao, H.J., Shi, H., Kozat, U.C., Ye, Y.: Toward optimal performance with network assisted TCP at mobile edge. In: 2nd USENIX Workshop on Hot Topics in Edge Computing (HotEdge 19). USENIX Association, Renton, WA (2019)
51. Alrshah, M.A., Al-Maqri, M.A., Othman, M.: Elastic-tcp: Flexible congestion control algorithm to adapt for high-bdp networks. IEEE Syst. J. **13**(2), 1336–1346 (2019)
52. Ford, A., Raiciu, C., Handley, M., Bonaventure, O.: TCP extensions for multipath operation with multiple addresses, Internet Requests for Comments, RFC Editor, RFC 6824 (2013), accessed 28-08-2020. http://www.rfc-editor.org/rfc/rfc6824.txt
53. Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L., Paxson, V.: Stream control transmission protocol, Internet Requests for Comments, RFC Editor, RFC 2960 (2000), accessed 28-08-2020. https://www.rfc-editor.org/rfc/rfc2960.txt
54. Paasch, C., Detal, G., Duchene, F., Raiciu, C., Bonaventure, O.: Exploring mobile/wifi handover with multipath tcp. In: ACM SIGCOMM Workshop on Cellular Networks (Cellnet'12) (2012)
55. Jana and Swett, I.: QUIC: A UDP-based secure and reliable transport for http/2, Working Draft, IETF Secretariat, Internet-Draft draft-tsvwg-quic-protocol-00 (2015). http://www.ietf.org/internet-drafts/draft-tsvwg-quic-protocol-00.txt
56. Kohler, E., Handley, M., Floyd, S.: Datagram congestion control protocol (DCCP), Internet Requests for Comments, RFC Editor, RFC 4340 (2006), accessed 28-08-2020. http://www.rfc-editor.org/rfc/rfc4340.txt
57. Braud, T., Bijarbooneh, F.H., Chatzopoulos, D., Hui, P.: Future networking challenges: the case of mobile augmented reality. In: Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 1796–1807 (2017)
58. Speedtest by ookla—the global broadband speed test, accessed 29-08-2020. http://speedtest.net/
59. Opensignal combines real-world measurements with scientific analysis to provide independent insights on mobile connectivity globally, accessed 29-08-2020. https://www.opensignal.com/
60. State of mobile networks: Usa (2016), accessed 23-02-2017. https://opensignal.com/reports/2016/02/usa/state-of-the-mobile-network/
61. Xu, Y., Wang, Z., Leong, W.K., Leong, B.: An end-to-end measurement study of modern cellular data networks. In: International Conference on Passive and Active Network Measurement, pp. 34–45. Springer, Berlin (2014)
62. Braud, T., Kämäräinen, T., Siekkinen, M., Hui, P.: Multi-carrier measurement study of mobile network latency: the tale of hong kong and helsinki. In: 2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), pp. 1–6 (2019)
63. Mobile broadband explosion, Rysavy, Tech. Rep. (2011), accessed 28-08-2020. https://rysavyresearch.files.wordpress.com/2017/08/2013-08-4g-americas-mobile-broadband-explosion.pdf
64. Realistic lte performance, Motorola, Tech. Rep. (2009), accessed 28-08-2020. https://www.apwpt.org/downloads/realistic_lte_experience_wp_motorola_aug2009.pdf
65. Heisler, Y.: A huge 4g milestone: Lte is now available for 98% of Americans, accessed 23-02-2017. https://bgr.com/2015/03/23/lte-coverage-map-united-states/
66. Overview of 3gpp release 12 v0.2.0, 3GPP, Tech. Rep. (2015), accessed 28-08-2020. https://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/
67. Mumtaz, S., Huq, K.M.S., Rodriguez, J.: Direct mobile-to-mobile communication: Paradigm for 5g. IEEE Wirel. Commun. **21**(5), 14–23 (2014)
68. 802.11ac: It's still hard to find, but it's fast, accessed 23-02-2017. https://opensignal.com/blog/2016/05/05/802-11ac-its-still-hard-to-find-but-its-fast/
69. LTE latency: How does it compare to other technologies? accessed 23-02-2017. https://opensignal.com/blog/2014/03/10/lte-latency-how-does-it-compare-to-other-technologies/
70. Castignani, G., Lampropulos, A., Blanc, A., Montavont, N.: Wi2me: a mobile sensing platform for wireless heterogeneous networks. In: Proceedings of the 2012 32nd International Conference on Distributed Computing Systems Workshops, pp.108–113 (2012)
71. Heusse, M., Rousseau, F., Berger-Sabbatel, G., Duda, A.: Performance anomaly of 802.11 b. In: INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, vol. 2, pp. 836–843. IEEE, New York (2003)
72. Shrivastava, V., Rayanchu, S., Yoonj, J., Banerjee, S.: 802.11n under the microscope. In: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '08, pp. 105–110. ACM, New York (2008)
73. Wifi peer-to-peer (p2p) technical specification version 1.7, Wifi Alliance, Tech. Rep. (2016), accessed 28-08-2020. https://www.wi-fi.org/download.php?file=/sites/default/files/private/Wi-Fi%20P2P%20Technical%20Specification%20v1.7.pdf
74. Chatzopoulos, D., Sucipto, K., Kosta, S., Hui, P.: Video compression in the neighborhood: an opportunistic approach. In: 2016 IEEE International Conference on Communications (ICC), pp. 1–6 (2016)
75. Condoluci, M., Militano, L., Orsino, A., Alonso-Zarate, J., Araniti, G.: Lte-direct vs. wifi-direct for machine-type communications over lte-a systems. In: 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 2298–2302. IEEE, New York (2015)
76. Alliance, N.: 5g white paper. In: Next generation mobile networks, white paper (2015)
77. Global infrastructure (2018), accessed 29-08-2020. http://aws.amazon.com/about-aws/global-infrastructure
78. Alibaba cloud global locations: Deploy around the world including mainland China, accessed 29-08-2020. https://www.alibabacloud.com/global-locations

79. Global locations—regions and zones | google cloud, accessed 29-08-2020. https://cloud.google.com/about/locations/
80. Broadband vs. dial-up adoption over time (2019), accessed 28-08-2020. https://www.pewresearch.org/internet/chart/broadband-vs-dial-up-adoption-over-time/
81. Broadband vs. dial-up adoption over time, accessed 23-02-2017. http://www.pewinternet.org/chart/broadband-vs-dial-up-adoption-over-time/
82. Fiber offers, accessed 02-01-2020. https://boutique.orange.fr/internet/offres-fibre
83. Meller, H., Cohen, P.: The state of mobile web us 2015, SimilarWeb, Tech. Rep. (2015), accessed 28-08-2020. https://www.similarweb.com/corp/reports/the-state-of-mobile-web-in-the-us-2015/
84. Desktop vs mobile vs tablet market share africa, accessed 28-08-2020. https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/africa
85. van der Vorst, T., Brennenraedts, R., Driesse, M., Bekkers, R.: Beyond fast—how the speed of residential internet access will develop between now and 2022, Eindhoven University of Technology, Tech. Rep. (2016), accessed 28-08-2020. http://www.cable-europe.eu/wp-content/uploads/2010/09/NLkabel-Cable-Europe-Dialogic-TU-Eindhoven-Beyond-fast-How-the-speed-of-the-internet-will-develop.pdf
86. Pesovic, A.: Is symmetrical bandwidth a myth or a must? accessed 23-02-2017. https://insight.nokia.com/symmetrical-bandwidth-myth-or-must
87. Cloonan, T., Emmendorfer, M., Ulm, J., Al-Banna, A., Chari, S.: Predictions on the evolution of access networks to the year 2030 and beyond, ARRIS, Tech. Rep. (2015), accessed 28-08-2020. https://www.nctatechnicalpapers.com/Paper/2014/2014-predictions-on-the-evolution-of-access-networks-to-the-year-2030-beyond
88. Digi-Capital, For ar/vr 2.0 to live, ar/vr 1.0 must die, accessed 17-01-2019. https://www.digi-capital.com/news/2019/01/for-ar-vr-2-0-to-live-ar-vr-1-0-must-die/
89. Braud, T., Heusse, M., Duda, A.: Dynamics of two antiparallel TCP connections on an asymmetric link. In: Proceedings of the 2016 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, New York (2016)
90. Heusse, M., Merritt, S.A., Brown, T.X., Duda, A.: Two-way TCP connections: Old problem, new insight. SIGCOMM Comput. Commun. Rev. **41**(2), 5–15 (2011)
91. Hoeiland-Joergensen, T., McKenney, P., Gettys, J., Dumazet, E.: The flowqueue-codel packet scheduler and active queue management algorithm, Working Draft, IETF Secretariat, Internet-Draft draft-ietf-aqm-fq-codel-06 (2016), accessed 28-08-2020. http://www.ietf.org/internet-drafts/draft-ietf-aqm-fq-codel-06.txt
92. Flach, T., Dukkipati, N., Cheng, Y., Raghavan, B.: TCP instant recovery: incorporating forward error correction in TCP, Working Draft, IETF Secretariat, Internet-Draft draft-flach-tcpm-fec-00 (2013), accessed 28-08-2020. http://www.ietf.org/internet-drafts/draft-flach-tcpm-fec-00.txt
93. Sundararajan, J.K., Shah, D., Medard, M., Jakubczak, S., Mitzenmacher, M., Barros, J.: Network coding meets TCP: Theory and implementation. Proc. IEEE **99**(3), 490–512 (2011)
94. Hunger, A., Klein, P.A.: Equalizing latency peaks using a redundant multipath-tcp scheme. In: 2016 International Conference on Information Networking (ICOIN), pp. 184–189 (2016)
95. Frommgen, A., Erbshäußer, T., Buchmann, A., Zimmermann, T., Wehrle, K.: Remp TCP: Low latency multipath TCP. In: 2016 IEEE International Conference on Communications (ICC), pp. 1–7 (2016)
96. Paasch, C., Detal, G., Duchene, F., Raiciu, C., Bonaventure, O.: Exploring mobile/wifi handover with multipath TCP. In: Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design, ser. CellNet '12, pp. 31–36. ACM, New York (2012)
97. Raval, N., Srivastava, A., Razeen, A., Lebeck, K., Machanavajjhala, A., Cox, L.P.: What you mark is what apps see. In: Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '16, pp. 249–261. ACM: New York (2016). http://doi.acm.org/10.1145/2906388.2906405
98. Bo, C., Shen, G., Liu, J., Li, X.-Y., Zhang, Y., Zhao, F.: Privacy.tag: privacy concern expressed and respected. In: Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, ser. SenSys '14, pp. 163–176. ACM, New York (2014)
99. Shu, J., Zheng, R., Hui, P.: Cardea: Context-aware visual privacy protection for photo taking and sharing. In: Proceedings of the 9th ACM Multimedia Systems Conference, ser. MMSys '18, pp. 304–315. Association for Computing Machinery, New York (2018)
100. Aditya, P., Sen, R., Druschel, P., Joon Oh, S., Benenson, R., Fritz, M., Schiele, B., Bhattacharjee, B., Wu, T.T.: I-pic: A platform for privacy-compliant image capture. In: Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '16, pp. 235–248. ACM, New York (2016).
101. Geambasu, R., Kohno, T., Levy, A.A., Levy, H.M.: Vanish: increasing data privacy with self-destructing data. In: Proceedings of the 18th Conference on USENIX Security Symposium, ser. SSYM'09, pp. 299–316. USENIX Association, New York (2009)
102. (2015), accessed 28-08-2020. http://www.linux-magazine.com/Issues/2015/175/FQ-CoDel-and-MPTCP
103. Fiddler—web debugging proxy—telerik, accessed 29-08-2020. https://www.telerik.com/fiddler
104. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: Computer Vision (ICCV), 2011 IEEE International Conference on Computer Vision, pp. 2564–2571. IEEE, New York (2011)
105. Horn, B.K., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**(1–3), 185–203 (1981)
106. Shoaib, M., Venkataramani, S., Hua, X.-S., Liu, J., Li, J.: Exploiting on-device image classification for energy efficiency in ambient-aware systems. In: Mobile Cloud Visual Media Computing, pp. 167–199. Springer, Berlin (2015)
107. Cho, J., Sundaresan, K., Mahindra, R., Van der Merwe, J., Rangarajan, S.: ACACIA: Context-aware Edge Computing for Continuous Interactive Applications over Mobile Networks. In Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies, pp. 375–389. ACM, New York (2016)
108. Guler, A.U., Braud, T., Hui, P.: Spatial interference detection for mobile visible light communication. In: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 1–10 (2018)
109. Ndjiongue, A.R., Ngatched, T.M.N., Dobre, O.A., Armada, A.G.: Vlc-based networking: feasibility and challenges. IEEE Netw. **34**(4), 158–165 (2020)
110. Ceipidor, U.B., Medaglia, C.M., Volpi, V., Moroni, A., Sposato, S., Carboni, M., Caridi, A.: NFC technology applied to touristic-cultural field: a case study on an italian museum, In: 2013 5th International Workshop on Near Field Communication (NFC), pp. 1–6 (2013)
111. Liu, V., Parks, A., Talla, V., Gollakota, S., Wetherall, D., Smith, J.R.: Ambient backscatter: wireless communication out of thin air. ACM SIGCOMM Comput. Commun. Rev. **43**(4), 39–50 (2013)
112. Yi, S., Li, C., Li, Q.: A survey of fog computing: concepts, applications and issues. In: Proceedings of the 2015 Workshop on Mobile Big Data, ser. Mobidata '15, pp. 37–42. Association for Computing Machinery, New York (2015) https://doi.org/10.1145/2757384.2757397
113. Chatzopoulos, D., Ahmadi, M., Kosta, S., Hui, P.: Openrp: a reputation middleware for opportunistic crowd computing. IEEE Commun. Mag. **54**(7), 115–121 (2016)

114. Billinghurst, M., Kato, H.: Collaborative augmented reality. Commun. ACM **45**(7), 64–70 (2002)
115. Zhang, W., Han, B., Hui, P., Gopalakrishnan, V., Zavesky, E., Qian, F.: Cars: collaborative augmented reality for socialization. In: Proceedings of the 19th International Workshop on Mobile Computing Systems and Applications, ser. HotMobile '18, pp. 25–30. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3177102.3177107
116. Zhou, P., Braud, T., Zavodovski, A., Liu, Z., Chen, X., Hui, P., Kangasharju, J.: Edge-facilitated augmented vision in vehicle-to-everything networks. IEEE Trans. Veh. Technol. **69**(10), 12187–12201 (2020)
117. Arth, C., Schmalstieg, D.: Challenges of large-scale augmented reality on smartphones, pp. 1–4. Graz University of Technology, Graz (2011)
118. Louis, T., Troccaz, J., Rochet-Capellan, A., Bérard, F.: Is it real? measuring the effect of resolution, latency, frame rate and jitter on the presence of virtual entities. In: Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces, ser. ISS '19, pp. 5–16. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3343055.3359710
119. Brady, K., Wu, B., Sim, S.H., Enquobahrie, A., Ortiz, R., Arikatla, S.: Modeling reduced user experience caused by visual latency. In: Advances in Ergonomics Modeling, Usability and Special Populations, pp. 267–277. Springer, Berlin (2017)

Tristan Braud got his PhD degree from Université Grenoble Alpes, France, in 2016. He is currently Assistant Professor at the Division of Integrative Systems and Design at the Hong Kong University of Science and Technology, Hong Kong. His research interests revolve around human-centered pervasive system design, including mobile computing, cyber-foraging, and AR.

Wenxiao Zhang received the bachelor degree in computer science and technology from Zhejiang University in 2014 and the PhD degree in computer science and engineering from the Hong Kong University of Science and Technology in 2019. He is currently a senior engineer in the Edge Cloud Innovation Lab at Huawei Technologies Co., Ltd. His research has included work on augmented reality, virtual reality, cloud gaming, and mobile computing.

Benjamin Finley received his D.Sc. degree with distinction in telecommunication engineering from Aalto University, Helsinki, Finland, in 2017. He is currently a postdoctoral researcher in the Department of Computer Science at the University of Helsinki, Finland. His current research areas include telecom data analysis including empirical IoT data and machine learning in user quality of experience.

Pan Hui is a Professor of Computational Media and Arts and Director of the Centre for Metaverse and Computational Creativity at the Hong Kong University of Science and Technology. He is also the Nokia Chair in Data Science at the University of Helsinki, Finland. He is a world-leading expert in Augmented Reality and has been recognised as an International Fellow of the Royal Academy of Engineering, a member of Academia Europaea, an IEEE Fellow, and an ACM Distinguished Scientist.

Part IV

Applications in Arts, Education, and Culture

Augmented Reality in Arts Education

12

Chiara Panciroli, Manuela Fabbri, Alessandro Luigini, Anita Macauda, Laura Corazza, and Veronica Russo

Contents

12.1 Basic Concepts 305
12.1.1 Augmented Reality and Arts Education: What Intersection Point? 305
12.1.2 Arts Education Strategies in the Twenty-First Century 308
12.1.3 Augmented Reality in Art Teaching at a European Level: What Developments? 310

12.2 Augmented Reality for Meeting and Making Art 314
12.2.1 Augmented Reality for Knowledge of Arts 314
12.2.2 Augmented Reality-Based Learning and Creative Expression 316

12.3 Digital Environments and Augmented Reality 317
12.3.1 Augmented Reality as Third Space 317
12.3.2 Augmented Reality for Creating Digital Artefacts 319

12.4 The Language of Images and Augmented Reality 320
12.4.1 The Audio-Visual Language 320
12.4.2 Body Language 321
12.4.3 Educational Experiences 321
12.4.4 Augmented Reality and Educational Experiences with Arts 323
12.4.5 Augmented Reality: Art for the Sake of Art 324

12.5 Digital Competences and Augmented Reality 324
12.5.1 Augmented Reality and Digital Innovation 324
12.5.2 Augmented Reality and Digital Competences of Museum Educators 326

References 328

Abstract

This chapter focuses attention on the potential of augmented spatial experience technologies in the pedagogical use of art and how these can significantly enhance the role of art education. The reasons why we feel it is necessary to dedicate deeper analysis to the theme of augmented reality in arts education are fueled by the acknowledgement that the theme is developing rapidly but lacks a systematization of the field experiences, coming from different research fields. An assiduous interdisciplinary discussion, with special reference on the one hand by the scholars of the Digital Heritage, is forever committed to the documentation and valorization of the tangible and intangible historical-artistic heritage and, on the other by the scholars of arts pedagogy and educational technology, seems more than ever necessary in order to contribute to minting the same coin, that is, the one relating to the safeguarding of the value of the arts and the heritage for the development of individual and, therefore, a society capable of evolving starting from the memory of its own expressive capacities.

Keywords

Arts education · Augmented reality · Digital environments · Learning augmented · Digital competences · Language of images · Third space · Creative expression · Digital artefacts

C. Panciroli (✉) · M. Fabbri · A. Macauda · L. Corazza · V. Russo
Department of Education Studies "Giovanni Maria Bertin", University of Bologna, Bologna, Italy
e-mail: chiara.panciroli@unibo.it; m.fabbri@unibo.it; anita.macauda@unibo.it; laura.corazza@unibo.it; veronica.russo6@unibo.it

A. Luigini
Faculty of Education, Free University of Bozen-Bolzano, Brixen-Bressanone, Italy
e-mail: alessandro.luigini@unibz.it

12.1 Basic Concepts

12.1.1 Augmented Reality and Arts Education: What Intersection Point?

The reasons why we feel it is necessary to dedicate deeper analysis to the theme of augmented reality in arts education are fueled by the acknowledgement that the theme is

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_12

developing rapidly but lacks a systematization of the field experiences, coming from different research fields. An assiduous interdisciplinary discussion, with special reference on the one hand by the scholars of the Digital Heritage, is forever committed to the documentation and valorization of the tangible and intangible historical-artistic heritage and, on the other by the scholars of arts pedagogy and educational technology, seems more than ever necessary in order to contribute to minting the same coin, that is, the one relating to the safeguarding of the value of the arts and the heritage for the development of individual and, therefore, a society capable of evolving starting from the memory of its own expressive capacities. In this innovation process, the applied technology acquires a proprietary role. It presents itself as the "mirror" within which scholars reflect, look at each other, observe each other, and in some cases "resemble" each other; specifically, it refers to digital technology applied to arts education. In fact, the environments that are described are digital, as are the experiences that a large part of society carries out every day. We obtain information, build relationships, buy services or products, spend free time, and leave traces of our existence in digital environments, whether they are contained in a mobile device or in the physical space in which we find ourselves.

In this context, where the individual is formed and acts, representation has the primacy, thanks to digital technologies that have increased its possible declinations. We speak through the representation of our voice on the phone, we watch entertainment and on-demand contents through digital devices, and we interact via social networks; this list could be extended to include a significant number of daily activities. Ever since the reproducibility of significant artefacts has been extensively investigated, it is clear how the advent of digital processing and communication has transformed artefacts: from simple "faithful copies" of the original to identical originals. In the educational experience of art, both in the school environment and in the exhibition-museum environment, it is now possible to experience all the features of augmented reality that digital technologies have promised for decades: from immersive VR through head-mounted display (HMD) stereoscopic viewers to the use of smartphones or tablets for the enjoyment of augmentative content. The most relevant aspect of these technologies is precisely that of expanding the real experience of the acted-out space, both physical and digital, in different contexts. For various reasons, our interest is focused on the potential of augmented spatial experience technologies in the pedagogical use of art and how these can significantly enhance the role of art education.

When Myron Krueger coined the term artificial reality (AR) [1], his goal was to define a type of digital experience so immersive as to be perceived as real. He used the concept of AR as a tool to examine human/machine relationships by analyzing their possible exchange interfaces and examining the socio/cultural relationships connected to it. At the early 1990s, the idea of AR was superseded by the concept of the reality-virtuality continuum [2], graphically synthesized in a horizontal segment, where the real and the virtual were identified at the two extremes and, in the interval between the two, a type of mixed reality, which fades into augmented reality or augmented virtuality. Milgram used this definition for the construction of a taxonomy of visualization systems for mixed reality environments in relation to the degree of immersion required and the immersion device. These intermediate levels therefore belonged to a mixed reality in which the relationship between figure (the user who experiences reality) and background (the environment in which the user is immersed) determines the point of the Continuum in which one finds oneself.

Therefore, if the user experiences a reality in which structured digital information is added to what really exists around them, the field of augmented reality is created, that is, the field of computer graphics that studies the possibility of superimposing digital processing on the perceived reality. Conversely, if the user experiences a completely digital artificial reality, in which digital information is structured in such a way as to conform the perceived world, we are in what Jaron Lanier in the early 1980s had called virtual reality [3].

Today, virtual reality, augmented reality, mixed reality, haptic feedback, and gesture recognition represent technologies that increasingly tend to be merged together and that can be grouped right under the general term of artificial reality, thereby fulfilling Krueger's predictions [1].

In this regard, it is with the launch of Google Cardboard and with the crowdfunding campaign to finance the Oculus Rift project, at the beginning of the 2010s that commercial interest has reawakened in a sector that has only remained of interest for some research centers. Google with its Google Glass (2013) has tried out pioneering ideas in the field of augmented reality, as has Microsoft, which with the expected global launch of HoloLens first of all and then with HoloLens 2 has tried to fulfil the expectations generated, and not fully satisfied, by Google Glass. Finally, the Bridge market was launched, a 3D scanner for small environments which, combined with a mobile headset, allows real-time scanning of the environment in which one finds oneself, allowing for the geometrically coherent integration of 3D models.

As for the possibility of user interaction within an artificial reality, the enabling technologies are those that derive from the field of *physical computing*, that is, the science that studies the creation of physical interactive systems through the integrated use of software, hardware, and sensor and actuator systems, capable of "recording" the user's movements, rendering them in the form of a digital signal. Microsoft's Kinect 3D sensor (2010) was the first to enter the video game market on a large scale and, thanks to a moderately "open" policy, offers manipulation actions for real-time tracking of body movements or for home motion capture sessions (2016).

Regarding gesture control, a solution is represented by the devices, developed starting from 2010 by Leap Motion: a sensor that recognizes and tracks the movement of the fingers of the hands translating it into 3D coordinates, which can be used to interact virtually in mixed reality environments. Similarly, the most recent Google Project Soli promises the realization of an electromagnetic sensor capable of interpreting the movements of the fingers as well as some material or morphological characteristics of inorganic objects. Similarly, the MYO, an intelligent bracelet, can read the electrical activity of the forearm muscles to control other associated technologies via arm movements. The most recent developments in mobile devices see the integration of AR technology and, starting from 2020, the use of LiDAR technology, i.e., the possibility of measuring the distance of objects within a radius of a few meters, allowing to contextualize the two-dimensional and three-dimensional graphic elements with extreme accuracy and speed. The reasons that motivate the development of AR applications with respect to VR are different, but two are of considerable impact. The first is the possibility to implement many AR applications on mobile devices (hand-held devices) that are particularly easy to use and that do not fully engage the user's visual field, offering a natural perception of the environment in which they is situated. The second is the ease with which digital content can be distributed on paper (using marker activators or even markerless) such as books and magazines, postcards, or advertisements.

A decisive development for the near future will be the integration of AR technologies with artificial intelligence (AI), to allow better and better performance integration between digital content and real environments. The support of neural networks, like GAN (Generative Adversarial Network) and CNN (Convolutional Neural Network), will allow the application to recognize the elements that make up the surrounding environment and the consequent stabilization of the three-dimensional contents and their adaptability to context.

These, and other technologies, which are being released, will contribute to the development of new forms of learning and experiential art teaching. In fact, compared to the overall picture described, it can be observed that the advent of digital and social media offers a new change of perspective, especially in the ways in which we communicate, produce cultural content, and "express our identities" [4].

In particular, if communities interact with places of art through research and observation, similar approaches are essential to start a cognitive process with art objects. In this sense, the artistic object becomes an exceptional text capable of combining and narrating extraordinary stories and contexts. In fact, one of the fundamental tasks of art education is that of questioning reality through technologies, when the latter allow us to ask new questions. Among these technologies, the reference is to the new immersive and augmented digital environments that allow not only to access information but also to experiment and create original meanings. It is thus necessary to rethink media in a systemic perspective, in which art and heritage ought to be analyzed as part of a complex context and not just as isolated entities; to do this, an articulated design centered on the procedural nature of the experience is required. The ecosystemic approach to art and heritage leads to an increasingly natural interweaving between physical-virtual spaces and different media, in which the context plays a decisive role to achieve the goals. Many of the technologies today guarantee multiple parties can collaborate and build artefacts in a shared system. With the evolution of products conveyed by one or more media, even the experience is transformed, becoming movement, and transit between the various elements of the system itself [5]. Most of our experiences with art and heritage in digital environments, in fact, are configured as a process of actions, facts, and behaviors linked together: from initial intentions to research and comparison, to construction, and so on. This means that, for example, when the output of an activity is the production of a tangible work, this brings with it a series of visible and invisible connections with other elements of the experience that constitute the actual added value of that product. It is precisely this process that allows the artistic and cultural heritage both to convey information, knowledge, and emotions and to transform over time and resemantize itself. In this regard, an impressive number of objects are transformed, acquiring different meanings even long after their creation, thus bringing signs of more recent events [6].

It is by now a given that defines heritage education "as a formative activity, formal and informal, which while it educates towards the knowledge and respect of the assets, by means of the adoption of responsible behaviors, makes of the heritage the concrete subject of research and interpretation, adopting the perspective of the recurrent and permanent development of everyone's active and responsible education" [7]. The different learning models, together with those of didactic mediation, have a key role in defining the concept of enjoyment of the artistic item. This aspect underscores how experience of the heritage also in digital environments cannot coincide solely with one or more active methodologies of knowledge but must be situated within a problematizing pedagogy that is capable of defining the cultural project by developing suitable methods and strategies, in which there is specific attention to the needs of people and society. Jenkins [8] in this regard speaks of "participative cultures" that develop in the open space of the Internet, thanks to the Web and social networking instruments with the aim of fostering artistic expression and civic commitment promoting a sense of belonging. Didactics centered increasingly on the relationship between new technologies and forms of learning/entertainment, orienting itself towards an approach between amusement and edutainment, through which it is possible to learn and educate. Techniques of digital mediation, graphic models and three-dimensional models, 360° panoramas and dynamic interfaces have redefined the spaces and times of leaning. There is indeed no doubt that today digital media are the protagonists of a shift towards renewed models of communication that aim at an expansion, according to ever more rapid and immediate forms, of the didactic and cultural provision. The communicative actions and the new forms of representation aim to facilitate understanding, to clarify aspects of complexity, to present concepts in a clear and synthetic way, and to make the information more explicit and accessible guaranteeing at the same time an elevated scientific level of the contents proposed.

In this regard, we cannot end this synthetic review without citing Google once again and, in particular, the Expeditions project that expands the set of tools to support education in school contexts with particular reference to its virtual educational environment Classroom. Google Expeditions allows teachers and students to undertake immersive "expeditions" by exploring natural environments or museums or art galleries affiliated with the Google Arts and Culture project, virtually redefining the spatial boundaries of classrooms and laboratories.

12.1.2 Arts Education Strategies in the Twenty-First Century

Growing globalization is posing significant problems to the world of education with particular reference to migration and multiculturalism, on the one hand, and technological advances and the development of the knowledge economy, on the other hand. The educational system can contribute in a decisive way to prepare young people for the different roles they are called to play in contemporary society. In particular, schools have the task of helping young people to develop self-confidence, as individuals and members of community groups, supporting them in the acquisition of a wide range of skills and interests, as well as identifying and expanding their creative potential.

These elements also pose several challenges for art and digital education. In fact, the growing interest shown by international organizations has led, in recent years, to a fundamental awareness on the subject and to the proposal of research on artistic and cultural education in formal and nonformal contexts in Europe, also through the contribution of digital.

The specific reference is to the actions proposed by UNESCO regarding the development of initiatives that promote art education, emphasizing its role in everyone's education. In this regard, the first World Conference on Arts Education: Building Creative Capacities for the Twenty-First Century, held in Lisbon in 2006, affirmed the need to define and increase the importance of art education in all societies, through the carrying out research aimed at establishing guidelines for the enhancement of artistic education. "Art education is a universal human right and the arts play a key role allowing the full development of the individual."

In the decade 2003–2012, UNESCO specifically published three strategic documents for culture, art, and art education: the UNESCO Convention for the Protection of the Intangible Cultural Heritage (ICH-UNESCO, 2003), the Road Map for Arts Education (UNESCO, 2006), and Seoul Agenda for the Arts Education (UNESCO, 2011). These three documents intersect with the broader conventions and policies of the United Nations, seeking to promote humanism, cultural pluralism, and equality and therefore represent guidelines at the basis of regional and national cultural policies. In particular, the Intangible Cultural Heritage recognizes performing arts practices as forms of immaterial cultural knowledge and emphasizes the rights of minorities to support these practices by withstanding the hegemony culture. Ratified in 2003, ICH extends UNESCO's previous political directives on human rights (UN, 1948); economic, social, and cultural rights (UN, 1966a); civil and political rights (UN, 1966b); cultural and natural heritage (UNESCO, 1972); traditional culture and folklore (UNESCO, 1989); and cultural diversity (UNESCO, 2001). ICH has been widely applied within formal education [9] and supports the values of programs dedicated to research, higher education, and the practice of intangible cultural heritage. Specifically, the convention promotes the following:

1. Educational programs, of awareness and information, aimed at a general public, particularly young people
2. Specific educational and training programs within communities and interest groups

These two points represent an indispensable social track especially for those who are not aware of the value of intangible culture and who, therefore, need an education in the continuous practices of immaterial culture. These two objectives also raise questions for art educators who work within dynamic and multicultural communities, in particular, in urban schools and places of informal education [10–12]. Teachers who implement the ICH objectives within the curriculum have to deal with uncertain paths, to integrate the cultural performances of minorities, without trying to represent them rigidly or statically.

In fact, the Road Map for Arts Education (UNESCO, 2006) focuses on role of art educators by emphasizing the importance of cultural pluralism as essential for equality, with reference to the United Nations Convention on Human Rights (1948) and the United Nations Convention on the Rights of the Child (ONU, 1989). Indeed, this document does not promote artistic education as an end unto itself but identifies the function of the arts in the development of every individual's potential. Based on the theoretical assumptions of arts education [13–15], the Road Map recognizes how the arts can enhance young people's creativity (not just a small elite considered "talented") and support learning in interdisciplinary areas within formal education by emphasizing the relationship between art education and the economies of creative knowledge of the twenty-first century [16].

Emphasizing the importance of artistic education in formal curricula, the Road Map seems to assume that arts education around the world takes place predominantly within institutional contexts but neglects the complex diversity of global arts pedagogies that resist institutionalization.

The Seoul Agenda for Arts Education, adopted in 2011 (UNESCO, 2011), sought to extend the Road Map, through specific objectives, to be achieved at national and regional level [17]. The importance of informal education, of the involvement of students of different age groups and with different learning styles, in order to achieve personal and social well-being is highlighted. In fact, the third objective of the Seoul Agenda refers to the application of principles and practices of art education to help solve the social problems and cultural challenges that today's world must face (UNESCO, 2011). The Seoul Agenda can thus be seen aligned with university education and specifically with degree courses that emphasize arts educators' social responsibility.

The European Council has also proposed significant reflections and actions towards the development of artistic education at school. Already in 1995, the Council of Europe launched a project called Culture, Creativity, and Youth, aimed at dealing with art education in the schools of the Member States, involving professional artists, and creating collaborations for the implementation of extracurricular activities.

Ten years later, in 2005, a Council of Europe Framework Convention on the value of cultural heritage for society was defined, underlining the need for European countries to preserve their cultural resources, to promote the cultural identity, to respect diversity, and to encourage intercultural dialogue. In particular, the Article 13 of the Framework Convention recognized the important role of cultural heritage in the field of art education but also recommended developing links between the disciplines taught in different fields of study. In 2006, an international conference was organized on the theme Promoting Cultural Education on Europe, preceded by a meeting of the European network of officials working in the field of artistic and cultural education. During this conference, the idea of a glossary was launched to lay the common foundations for the definition of cultural education and other related terms. The 2007 Council resolution also introduced a new Open Method of Coordination in the field of culture. Within the framework of the OMC, a working group on synergies between culture and education has been set up to promote key competence for "cultural sensitivity and expression." The task of this work group was to validate the best practices and formulate recommendations for initiatives aimed at promoting cooperation between culture and education, including arts education. In May 2007, the European Commission adopted a communication on A European Agenda for Culture in a World undergoing Globalization. In November of the same year, this communication was transformed into a Council Resolution on a European Agenda for Culture that recommended "encouraging arts education and participation in cultural activities in order to develop creativity and innovation." This resolution was in turn followed by a 2008–2010 Work Plan that recognized the importance of culture and creativity. 2008 is thus designated as the European Year of Intercultural Dialogue and 2009 as the Year of Creativity and Innovation. In fact, in 2008 the Council of Europe published a White Paper on intercultural dialogue proposing an intercultural approach to manage the diversity of cultures. This document highlighted how educational institutions, including museums, heritage sites, and schools of all types and levels, must be able to support intercultural exchanges, study, and dialogue through arts and cultural activities.

In March 2009, the European Parliament passed a Resolution on Art Studies in the European Union, which included the following recommendations:

- Artistic and intercultural teaching must be compulsory at all levels of education.
- The teaching of the arts must use the latest information and communication technologies.
- The teaching of art history must include meetings with artists and visits to places of culture.
- To make progress in these areas, the resolution calls for better supervision and coordination of artistic teaching at European level, with monitoring of the impact of artistic teaching on the skills of students in the European Union.

In addition to these main developments, linked above all to international and European cooperation, various conferences and initiatives have been held, some of which have led to changes in the policy of the various countries in the field of artistic and cultural education. At the same time, three international organizations representing educators in the artistic field – dramatic arts/theatre, plastic arts, and music – came together to form a global alliance (International Society for Art Education 2006–2017) inviting UNESCO to place art education on the global agenda for human development and sustainable social transformation. Creative Europe is the most recent framework program of the European Commission dedicated to culture and the creative sectors for the period 2014–2020 which aims to operate beyond national borders by actively promoting cultural and linguistic diversity.

12.1.3 Augmented Reality in Art Teaching at a European Level: What Developments?

Comparisons among the various initiatives, organized at European level over the last 20 years, have led to the identification of relevant questions for the areas of art education, particularly important in the field of school teaching. The studies confirm the existence of a hierarchy of school programs, in which the skills related to reading, writing, and learning logical-mathematical skills are favored to the detriment of arts teaching. In addition, among the arts, some forms of expression such as the visual and musical arts tend to be favored over others such as drama and dance. By way of confirmation of this different recognition of the arts in the educational system, the survey on Culture, Creativity, and Youth, carried out by the European Council, highlights the importance of the cultural dimension in educational policies and the need to nurture the artistic and creative attitudes of young people within a system of educational provision.

The arts in formal educational contexts not only have little recognition compared to other disciplines, but in some European countries, there are attempts to reduce the arts supply in the curricula in favor of subjects considered more relevant in terms of economic or academic success. In most national systems, the teaching of visual arts is compulsory both in primary education and in lower secondary school and only in some courses of upper secondary school. By analyzing the importance attributed to the arts in official documents, two main approaches were identified: art as a discipline and art learned through other disciplines. In particular, drama and dance are often integrated into other disciplines such as dance within physical education. In this context, it is difficult to promote the expressive qualities of dance in a discipline focused on physical exercise and sport. In the case of the arts as an autonomous discipline, Taggart and others [18] observed that the plastic arts and music are studied as compulsory courses in all 21 countries involved in the survey. The minor status recognized to artistic disciplines is reflected in the lack of interest in the evaluation and monitoring of standards in art teaching. The research has also highlighted the problems related to the fact that the time officially dedicated to teaching arts and the time actually made available within schools is insufficient to offer a broad and balanced program. Lack of time, space, and resources are thus identified as key factors limiting the success of arts teaching. We expect more and more that artistic teaching will fulfil a series of objectives, besides offering knowledge relating to this field. If the education systems increasingly recognize the importance of developing children's creativity and of acting in favor of their cultural education, it is not clear in what way the arts can contribute to each student's educational curriculum. Taggart and others [18] also observed that almost all countries have similar purposes, summarized as follows: the development of artistic skills, the knowledge and understanding of different artistic forms, the perception of cultural realities, the sharing of artistic experiences, as well as the possibility of becoming enjoyers of art and attentive users in this area. Furthermore, in most countries, arts education is also aimed at ensuring personal and sociocultural outcomes, such as self-esteem and self-confidence, individual expression, team spirit, understanding, and intercultural participation. Among the purposes of artistic teaching, particular attention to creativity emerges, also in relation to the importance attributed to innovation. As Bamford [19] remarks, many education systems rely on generalist teachers for arts disciplines, particularly in preschools and primary schools. Teaching the arts at a high level is not easy and that is why teachers do not feel confident with this task. We can thus deduce that it would be necessary to envisage the initial preparation of the teachers for the artistic subjects via measures of continuous professional development that allow for their knowledge to be brought up to date and their competencies improved. The research has dedicated little attention to the modalities of evaluation of the quality of artistic teaching, even if concerns often emerge regarding the variability in the standards and the need to offer a high-level learning experience inside the schools. In particular, Robinson [15] highlighted a structural problem that hinders the development of art in schools. Most of the time the governmental responsibilities in the field of art, cultural heritage, and education are shared between two or more Ministries, such as those of education and culture and sometimes of youth and sport. This can be a source of difficulty in the common understanding of needs and priorities. A unification of the Ministries can therefore produce benefits in terms of better understanding and greater effectiveness and efficiency.

While it is essential to monitor pupils' progress throughout the course of all studies, for artistic disciplines assessment this is even harder. A recent international study, carried out by Bamford [19] on the evaluation of arts and culture teaching in a European context, states that the main purpose of evaluation in arts teaching should be to clarify and make more concrete the objectives that students must achieve in the arts program. Assessment can be used formally (during learning) and summative (at the end of one or more learning sequences) to establish the pupils' results. Challenges for evaluation in this sector include the trend towards more integrated approaches to arts and cultural teaching and the fact that the evaluation responsibility is often divided between different subjects that have to collaborate and carry out planning jointly. Bamford [19] also states that the evaluation itself must be a creative act, arguing that the evaluation methods must capture the different types of learning of the student, be they actor, critical observer, or creator. The previous research has shown that, where requested, the evaluation of the artistic subjects is generally entrusted to teachers who, depending on the cases, may or may not have benefited from training and an orientation suited to that purpose. Furthermore, Taggart and others [18] have discovered that the main methods of evaluation used by the teachers consist in the request for a representation and/or work by the student on the given theme and in the awarding of a mark for the work done. Hence, three main approaches have been distinguished. The first one requires that the teachers should formulate an individual professional judgement as a function of the objectives and the contents of the syllabus. The second requires that the teacher should evaluate the students' performance on the grounds of a common standard expected of a given age group/level of studies. Lastly, the third sees the teacher attribute a level of progress to each student on the grounds of a graduated scale, irrespective of the age or the level of study. Most countries use the first two approaches for the evaluation. These systems allow to identify which students are making important progress and which ones are instead unsatisfactory, but the questions concerning the validity, the reliability, and the consequences of the different teaching systems, besides the methods of arts learning, have not been examined in great depth. The involvement of professional artists in art teaching has been recommended in different studies. This should allow for an improvement in the quality of the teaching and the art learning to favor, generally speaking, a greater creativity and more specifically to develop competencies and confidence in the teachers, giving access to a greater range of cultural resources. Bamford [19] has observed the existence of a significant link between the quality of artistic teaching and the involvement of professional artists as quality art teaching tends to be characterized by a strong partnership between schools and artistic organizations inside the community. In this sense, museums can also play a significant role in projects in which the museum becomes an experimental didactic room on the territory in connection with the school spaces or in projects that see the museum enters the schools with its professional figures of reference to enrich the teaching activities.

Several studies have underlined that the art education program of the twenty-first century will have to include in an ever clearer way the study of the new media (cinema, photography, and digital arts) allowing some students to use the technologies within the scope of the creative process. Also highlighted is a tendency towards greater transcurricular activities, which involve integrations between artistic subjects and other fields relating to the creative and/or cultural themes. These developments pose new demands on the teachers and the schools, besides the need for a strong sense of responsibility and support at the political level.

From the research into artistic education in Europe, it has emerged that the objectives relating to this type of education are very similar between all the European countries involved. The programs indicate, among their main objectives, "competence, knowledge and understanding tied to the arts," "critical evaluation," "cultural heritage," "cultural diversity," "development of personal expression," and "creativity." However, the goal concerning the development of "a permanent interest in the arts," which definitely represents one of the most important goals of art education, is recalled only in a few curricula. Also, if on the one hand many art education programs identify as general aims the "development of social competencies and those of communication" and as a specific objective that of "encouraging the connections between the artistic subjects and the other disciplines," on the other hand, in some countries, the acquisition of cultural and artistic competencies remains a generic aim of compulsory education. The conception of the program of artistic education is very heterogeneous among the European countries: in around half of these, the artistic subjects are proposed as individual courses in the school curriculum (e.g., musical education or visual arts), while for the other half of the countries involved, artistic education is conceived as a field of integrated studies (e.g., under the heading "Arts"). The range of the artistic fields varies greatly, even if the programs analyzed comprise Music and the Visual Arts in all the countries and nearly everywhere also Theatre, Dance, and Crafts; the Communication Arts are offered by 12 countries; Architecture is part of the compulsory artistic program in just five countries. At the primary level, artistic education is compulsory for all the pupils. The same is true for nearly all the countries at the level of lower secondary school education. At this educational level, when the artistic subjects are not compulsory, they are offered as optional subjects. The minimum compulsory number of hours of teaching to dedicate each year to artistic education at the primary level is between 50 and 100 h in half of the countries. This number is slightly lower to the secondary level, where the programs of around half of the countries recommend dedicating to this subject between 25 and 75 h a year. However, almost all of the countries encourage the schools to offer extracurricular activities in the artistic field. Even if the forms of art proposed in this field are multiple, music is well represented [20].

In Europe, there are numerous scholastic experiments to reinforce the encounter of the pupils with the world of the Arts and Culture. Thus, in almost every country, visits are organized to places of art and cultural interest, as well as the creation of partnerships with artists. We can also cite some examples of art festivals, celebrations, and competitions which the pupils of schools are encouraged to take part in. In some countries, this effort to develop the arts, culture, and education has been institutionalized with the creation of bodies and networks addressed to promoting artistic and cultural education. In many countries, reforms of the curricula are in progress, and in many cases, the planned changes also concern artistic education. The criteria for the evaluation of the pupils in the artistic subjects are usually defined at the school level by the teachers themselves. These criteria, which are established on the grounds of the learning objectives defined in the syllabus or by the guidelines provided by the further education authorities, allow the teachers to identify the pupils' various levels of achievement. Only in a small number of countries (seven) are the evaluation criteria defined by the central educational authorities. Most of the countries recommend using one or more types of evaluation scale, mainly at the secondary level, where the numerical scales are the most diffuse. At the primary level, the most frequent situation, cited in around 12 countries, is the use of synthetic grades, especially in the early years of schooling. In most of the countries, a final mark of unsatisfactory in artistic subjects is not taken into account for class changes and does not have a direct consequence on the pupils' education. In nearly all the countries, artistic education is entrusted to unspecialized teachers at the primary level, that is, to teachers qualified to teach all or most of the curricular subjects. In most of the cases, these teachers have received a general training in one or more artistic subjects and a targeted pedagogical training. The subjects in question are often Music and Visual Arts, which are part of the compulsory subjects in the study syllabus. Instead at the secondary level, specialized teachers are responsible for the artistic education. Before starting teacher training, they usually need to have acquired skills in one or more artistic subject.

The professional artists are seldom involved in art teaching in the schools is they do not have adequate qualifications. If they are authorized, this generally occurs for short periods. Furthermore, the participation of artists in the initial and continuing education of the teachers is seldom encouraged in the state projects. The realization of collaborative project between the different actors in the artistic field at school, also by means of political agreements and understandings, is a strategic choice that can valorize artistic education. At the political level, in some countries, collaborations have already been defined between the different ministries that support the projects or the creation of networks and bodies for the promotion of artistic education. However, it is recognized how artistic education in schools can draw greater benefits from the experience of professionals and institutions specialized in this sector, to ensure that art does not just represent a subject of study but, above all, a real-life experience [20]. The use of information and communication technologies is explicitly cited in the program of artistic education in many European countries where specific initiatives are organized to encourage the use of technologies. This is a priority goal for the future development prospects and thus also in the perspective of augmented reality.

The new technological augmented reality instruments can indeed offer concrete potential if they are put at the service of critical models of didactic mediation and developed within a precise educational project, relevant, and coherent. In this sense, the most recent developments of the augmented reality technologies pave the way to unexpected scenarios for learning in formal and informal educational contexts, in particular, in schools and museums. Indeed, the applicability and the application of augmented reality to the fields of didactics represents an innovative research field, whose growing experimentations provide significant elements for a pedagogical reflection. The efficacy of the educational pathways realized through augmented reality represents one of the most interesting themes in the current debate, especially as concerns the design, implementation, and evaluation of an enhanced didactics to be traced back to within an adequate theoretical-methodological framework [21–23].

Several studies have evidenced the educational potential of the digital environments (virtual, augmented, and hybrid), in which each student can try out concrete learning models of a cognitive and socio-relational nature [21, 24–27], with particular reference to education to the arts and to the cultural heritage. In this sense, the design and the construction of augmented reality digital environments for education to the heritage falls within the actions of the NOP (National Operative Program) "For school, competencies and environments for learning 2014–2020" that support the need to sensitize the students to the cultural, artistic, and landscape heritage to construct a full citizenship. In this perspective, the schools are trying out new experiences, oriented to the spread of a digital culture, for the knowledge and valorization of the cultural heritage through forms of digital artistic creation (performing arts) and digital communication (digital media, e-learning).

Hence, the design and realization of didactic experience through augmented reality represents one of the continuously evolving innovative themes [28–32]. The application of augmented reality to the fields of didactics relates, in particular, to mobile learning, literally understood as learning that avails itself of mobile devices [31, 32]. This form of learning draws on the affordances proper to the mobile devices: portability and flexibility, multifunctionality, ubiquity and ease of access, multimediality, multitouch, and personal possession [25, 33]. These characteristics determine the development of "educational experiences based on learning systems hinged upon mixed and/or augmented learning systems in the light of the construction of meanings by the student, allowing them to take part in a rich media milieu, distinguished by the combination of real and virtual objects, by the use of sensorial inputs and by the possibility to place virtual learning objects in the real world and to interact virtually with a hybridized world" [34], p. 134. We are thus witnessing the "transition from Mobile Learning to Augmented Reality Mobile Learning" [35], where upon "Augmented Reality Learning is a further development and extension of Mobile Learning" [35]. The use of augmented reality is part of the most recent developments in Learning with Technology that looks at new technological devices as a means that can facilitate the learning process [36]. In fact, augmented reality is a dynamic and interactive teaching tool that helps transform spaces, times, and ways of learning, thanks also to the fact that laboratories and classrooms are beginning to be widely equipped with suitable technological infrastructures. The Wi-Fi networks, for example, allow for the connection of the devices managed directly by the students for research and work activities done in class, according to the "Bring Your Own Device" philosophy (BYOD) [37]. The way of managing the classroom changes as a consequence, and new didactic methodologies and strategies are tested such as simulations, gamification, augmented reality, virtual reality and immersive learning, wearable devices, mobile learning, and shifting the attention to the new emerging models of learning (immersive learning, via simulation, learning in mixed realities . . .). In addition, the potential of augmented reality is significantly manifested also in the context of situated learning: "in a broader context of education, augmented reality is appealing because it aligns with situated learning. Students find connections between their lives and their education through the addition of a contextual layer" [38]. In fact, augmented reality, as a new frontier of digital communication based on the combination of augmented content and geo-referencing, allows just-in-time and just-in-place access to digital content with respect to the real perceptual experience. Digital objects and real objects coexist in another space that does not replace the physical world but overlaps it through a process of digital addition, in synchronicity and interactively, in order to provide an experience to high content that involves all our senses and reaches gradually rising levels of concreteness. In fact, the total "transparency" of the devices (from the monitor to the display, from the helmet to the glasses) moves in the direction of the immediacy of the experience on a perceptual, sensory, and motor level. By offering the possibility of experimenting in real-time new and creative ways of interaction, contextual to the experience, augmented reality is an active technology that offers opportunities for "immersion" and involvement, also on a cognitive, emotional, and relational level. Therefore, augmented reality is significantly inserted in the relationship between technologies and didactic mediation [24]. Its applications are placed within that category of mediators, analogues, based on simulation and which Damiano [39] places between iconic and symbolic mediators. With reference to the model of education architecture proposed by Clark [40] and subsequently integrated [41], learning centered on the use of augmented reality is placed in simulation architecture and recognized precisely in the symbolic simulation the its main didactic strategy that "is based on reproducing, in a protected and controllable context, experiences similar to those of the real world, to provide students with the possibility to act and learn from the consequences of their actions (. . .) through a both intellectual and emotional involvement" [34], p. 68, leading "to reconceptualize teaching in a more participatory and interactive form" [34], p. 133 (Fig. 12.1).

Fig. 12.1 Screenshot of AR application of interactive postcards, committed by Seoul Metropolitan Government in 2017

12.2 Augmented Reality for Meeting and Making Art

The latest developments in augmented reality technologies offer unexpected opportunities for learning in formal and non-formal educational contexts. In particular, the international scientific literature focuses on the relationship between augmented reality and arts education and on its potential [42, 44–46]. In particular, the use of augmented reality in educational contexts refers to the acquisition of the key competence recommended by the European Council for lifelong learning: the "cultural awareness and expression." The main feature of the this competence is to be found in the "awareness of the importance of the creative expression of ideas, experiences and emotions in a range of media, including music, performing arts, literature and the visual arts," and includes "an awareness of local national and European cultural heritage and its place in the world (. . .) and a basic knowledge of major cultural works, including popular contemporary culture" [47]. Such competence refers to skills related to "the appreciation and enjoyment of works of art and performances as well as self-expression through a variety of media" [47]. In this sense, access, exploration, knowledge, and enhancement of the artistic heritage through technological experiments orient towards "a new cultural approach that requires a rethinking of the identity construction of the self, the vision of urban spaces and the ways of using cultural and artistic products" [48]. Specifically, the use of augmented reality in the field of art education can be declined via two different modalities: augmented reality for the knowledge of art, at different levels, through the ability to access personalized and customizable paths and reality increases to support creative expression, promoting forms of participation and re-elaboration. The first modality refers to the field of reception/use of the artistic heritage. The second modality is positioned within the scope of production. In both cases, AR can offer involving, creative, and participative forms of learning [49].

12.2.1 Augmented Reality for Knowledge of Arts

Augmented reality technologies transform objects and places of art into digitally enriched environments through the preparation of paths that provide information that is not immediately usable. In fact, augmented use renews the methods of mediation of art and the related forms of learning. In this sense, augmented reality is an environment in which the acquisition of basic knowledge (didactic information) takes place, as well as the possibility of further study (contextualization, logical connections, . . .) [45].

The use of augmented reality is in fact correlated with an increase in the accessibility of information able to offer the subject who learns an experience of a predominantly visual nature, based on the possibility of exploring the object on the surface, through zooming later, but also from the inside, showing the invisible, thus improving and facilitating understanding also through a more realistic and engaging approach. The subject can query the works exhibited within a museum exhibition space and explore them without actually needing to touch them, giving rise to an enriched visit at the perceptive-sensorial level that extends the range of possibilities tied to the user experience. Indeed, the superimposing of layers of data of a contextual nature upon the artistic object impacts the process of acquisition and the deepening of the knowledge. Augmented knowledge is provided in a stratified way to allow progressively for the search of further meanings relating to the provenance of an object, to the historical and economic context of production, to the technique of manufacture, and to the curiosities connected to it. The languages and the media used are multiple: short textual contents, images, graphical representations, video reconstructions, and audio contents. Some augmented reality experiences are based on the rendering of a digital version of the artistic object in high reproduction, in which the elements of virtual reconstruction are superimposed upon the real ones, in an immersive experience that renders the impressions of finding oneself before a unique object (painting, monument, sculpture, etc.). This aspect turns out to be particularly important where it is necessary to show what the monuments were like in the past comparing them with the present. In these cases, the mobile device works like a real GPS navigator that accompanies the visitor along the thematic routes. Specifically, there are three aspects that the AR technologies have in common: 1. motion tracking, to identify the subject's position; 2. environmental understanding, to understand the surrounding space; and 3. light estimation, to make the illumination of the virtual objects and the light sources of the scene coherent with one another. Specifically with regard to the artistic assets, the augmented reality technologies can be used in a dual modality [50]: indoors, with particular reference to the pathways developed inside the museums and the galleries, and outdoors, in relation to the external pathways realized both in the place of art present diffusely in the urban territory and in the archaeological parks and the contemporary art parks [51–53].

Augmented Reality-Based Art Learning in Museums

Historically, the museums as places of conservation and exhibition valorize the materiality of the objects and stimulate the visitor to have a direct experience to reflect on the singularities of the craft objects. Indeed, the museums and more in general the cultural goods are recognized as large dataspaces [54], places of transmission of knowledge in which the memory of a people is collected [53].

Each object conserved in a museum is not a simple datum, but it inscribed inside it a story that the technologies and, in particular, the augmented reality devices succeed in telling. Augmented reality indeed allows us to represent a cultural object without isolating it from the context it belongs to. Its nature as a datum in this way is enriched and becomes a story in itself. The museum experience is thus transformed into performative terms, and the museum becomes a sensible organism that finds in the interaction between artwork and visitor its most evident rationale [55, 56]. When the spectator is in a position to articulate their relationship with the work, they are predisposed to a better knowledge of it and cooperates creatively to its evolution. Thus, the digital technologies start to make practicable far more radical forms of interactivity that can enrich the visitor's experience [57]. Hereafter, we present AR pathways realized in some museum contexts both with regard to their own permanent collections and with regard to temporary exhibition events, according to an explorative approach that allows the invisible to be seen. Since 2017, the Detroit Institute of Arts has offered its visitors, via systems of augmented reality activated on mobile devices, the chance, after having framed one of the mummies conserved in the exhibition spaces, to know the ancient burial rituals and to discover what is not visible and hidden beneath the bandages. The visualization of the relics through the AR system shows another layer of reality, otherwise invisible to the naked eye (the skeleton of the mummy, the mummification process, and the way in which the curators have taken care of the object; the Ishtar Gate in scale, the figurative scenes inside the Mesopotamian seals, . . .), as well as the possibility to access a musical layer that renders sounds and noises correlated to the object or to the scene of reference, conducting the subject to other contexts, beyond the physical walls of the museum.

The director of the DIA Salvador Salort-Pons has praised these educational values, stating that "augmented reality allows the user to see the invisible, to imagine art in its original situation and to understand how the objects have been used and experiences by people in their everyday lives" [58].

With reference to this type of use of AR, *GO! Muse* is also used. GO! Muse is the application for AR experiences developed for the MUSE – Sciences Museum of Trento (Italy). It offers the visitor the opportunity to see how the prehistoric animals housed in the museum looked and moved, whose physiognomy has been reconstructed, thanks to the collaboration between the museum's researchers and the paleoartists Davide Bonadonna and Fabio Manucci. The application allows to virtually place virtual 3D models of dinosaurs, prehistoric reptiles, and whales, largely extinct, in the real spaces of the museum, after framing their skeletons with the mobile device that supports the application. Among the applications that transform paintings into augmented installations, the Fourth Estate: Beyond the Visible represents a significant example of the use of augmented reality applied to a painting by Giuseppe Pellizza da Volpedo *The Fourth Estate* (1901). This is an initiative created by SMART – Augmented Culture in collaboration with the Museo del Novecento in Milan (Italy), which exhibits the work. Visitors are given the opportunity to know and interact with the painting in an unprecedented way. Through the AR, the picture becomes an access interface to learn about its history and the history of the characters represented. In fact, by framing the work through a mobile device, it is possible to activate an original form of interaction to access different content. At a first level of perception in which the image of the painting is placed, a second level given by the information on the painting is superimposed. The characters depicted seem to "come forward" to meet the observer who can listen to stories and curiosities told through their own voices, in an experience that is both visual and auditory. The experience ends when the observer of the work "meets" the character represented in the foreground, in the middle of the portrait, the man with the hat, and next to whom one can pose for a photo to be posted on the main social media [45]. Rembrandt Reality is the AR application released by the Mauritshuis museum (in the Hague). Through this application, *The Anatomy Lesson of Doctor Tulp* (1632) by Rembrandt van Rijn, exhibited at the Mauritshuis museum, it is possible to move inside the painting, closely observing the scene portrayed depicting the public dissection on the corpse of a just executed criminal. Once the app has been launched on their device, looking through the screen the viewer is faced with a portal. Inside is a faithful reproduction of one of the towers of the public weighing in Amsterdam, which at the time of Rembrandt housed the headquarters of the surgeons' guild. A voice guides you during the visit, explaining the painting, its history, and its details. In this specific case, the application allows to enrich the world with virtual holographic projections projected and into the physical world, where users can interact with them. With reference to the exhibition events, the Pérez Art Museum in Miami has proposed itineraries in augmented reality. As stated by its director, Franklin Sirmans, the PAMM is a place for experimentation and laboratory of ideas, to allow visitors of all ages and backgrounds and to interact with the most innovative visual arts of our times. Specifically, the reference is to the exhibition by the artist Felice Godrin, *Ivasive Species* (2017). It is a digital exhibition accessible to visitors using iOS devices in PAMM's outdoor areas and in the Padma and Raj Vattikuti Learning Theater on the museum's first floor. The exhibition offers a reflection on the relationship between physical and mental territories, with particular reference to the transformative and unstable state of our ecosystem, influenced by climate change. To this end, the works in AR overlap with the physical spaces of the museum, interacting with its architecture, evolving, and transforming its environment.

Augmented Reality-Based Art Learning in Archaeological and Art Sites

Augmented reality applies to outdoor mobile cultural access which provides visitors with the chance to move around the archaeological and art sites, viewing detailed contents and 3D images overlapping monuments, sculptures, and contemporary buildings and places [59]. In particular, for works of ancient art and archaeological sites, real vision is enriched with the picture of the place as it was in the past (the way it was like compared to the way it is now) [60]. In this regard, some AR applications provide customized routes, tailored to the user's profile, automatically providing information based on the location (reconstruction of sites in ruins; simulation of ancient streets, . . .). "Now that there are open-access, low-cost tools available to make virtual and augmented reality models, more archaeologists are creating and using such technologies to reach out to the public through experiential learning. Projects range from static models of individual objects to entire landscapes that translate aspects of memory, phenomenology, and materiality into virtual or augmented reality experiences" [61], p. 305. A significant example is offered by the *i-MareCulture project* (2016–2020) implemented in the submerged archeological park of Baia, a protected marine area located to the north of the Gulf of Naples. The project, funded by the European Union, thanks to the Lab4DIVE project (Mobile Smart Lab for augmented archeological dives), aims to document and make known the submerged archaeological heritage through an experience in virtual and augmented reality that does not require the visitors to take a dive underwater. I-MareCulture consists of a 3D navigation that allows the users to view the current underwater heritage and compare it with a hypothesis of site reconstruction (prior to being submerged) through information and digital images created, thanks to the collected material by divers. The application of dedicated underwater AR thus allows users to dive down to the seabed where the archaeological relics are located. Two modalities are proposed:

- With water, through a realistic representation of the environment and its characteristics
- Without water, through a decontextualized vision of the cultural heritage in such a way that the images are clearer and sharper

The project with which an AR application (2014) was honed, realized on the occasion of an archaeological excavation conducted inside Palazzo Baldini, in the historical center of Florence, is positioned inside an equally meaningful field. During the excavation, layers of different construction phases of the building emerged. The application offers the possibility to see the 3D model of the building, providing multiple levels of information, enriched with images, on the finds and on the original morphology of the building [60]. With reference to the use of AR in art parks, the Beyeler foundation introduced in 2020 a specific application "ART in the PARK" to discover the hidden features and history of five works in the Berower Park (Riehen, Switzerland): Thomas Schütte's sculptures *Hase* (2013), Jenny Holzer's *Living Series: You should limit the number of times . . .* (1989–1989), Ellsworth Kelly's *White Curves* (2001), Alexander Calder's *The Tree* (1966), and Philippe Parreno's *Water Lilies* (2012). The application also allows people to interact with virtual works of art through a series of proposals (Swing "The Tree"; Open "White Curves"; Play "Water Lilies", . . .). The experiences proposed here show how one of the most significant dimensions developed by augmented reality applications is represented by interactivity: virtual objects that increase real objects are not static but can perform movements and animations in response to user actions. Dünser states that the "[i]nteractions in AR engage learners with the content and allow for knowledge to be acquired through their [the students] own manipulation of content [. . .], as supported by constructivist learning theory" [22], p. 113. This dimension is also linked to the improvement of memory skills that leads to the preservation of the knowledge acquired through augmented reality for long term. In this regard, Chang points out how "[the AR application] facilitates the development of art appreciation [. . .], supporting the coupling between the visitors, the guide system, and the artwork by using AR technology, and helping visitors keep their memories of the artwork vivid" [62], p. 193. In fact, many projects aimed at enhancing art and promoting art education paths with the use of augmented reality propose learning based on the discovery that, unlike "static" or "one-dimensional" learning, occurs as a multi-perceptive, immersive, and engaging learning, with regard to multiple intelligences [43].

12.2.2 Augmented Reality-Based Learning and Creative Expression

Several sector studies highlight how augmented reality is understood not only as a technology for the use and knowledge of artworks but also as a possibility of expression, reflection, and critical thinking, as well as designing and testing significant and original paths of knowledge. In this regard, several artists have recognized these cognitive values of augmented reality technologies, which are of particular relevance in educational contexts.

An example of this is the work developed by the Danish-Icelandic artist Olafur Eliasson who created *AR Wunderkammer*, that is, a cabinet of curiosities in augmented reality that viewers can collect and experience: extra-terrestrial rocks, insects and rare birds like the puffin, and various objects including the floating compass that always returns to true north, an insect, and a Little Sun, Eliasson's solar-powered lantern, that can be charged by the AR sun. "The audience is invited to 'bring the outside in' by creating their own environment as they add AR objects, atmospheres and 'imaginary friends' to their own, personal space" [63].

Eliasson himself explains the significance of his work: "Today, where physical distancing guides our lives, it's as crucial as ever that we surround ourselves with things and atmospheres that really matter to us. [. . .] The artwork is about challenging our perception of the everyday and actively welcoming that which lies on the boundary between the known and the unknown. It is about creating spaces that meld the everyday and the extraordinary spaces that evoke vivid perceptions and embodied engagement" [64]. Similarly, some young artists working in the digital field use augmented reality within the museum field in an original way and in a playful-collaborative dimension that stimulates users to deal with the works through a creative reinterpretation of some masterpieces in art history, developed in a modern way. A significant example is *ReBlink* – project hosted by the Art Gallery of Ontario (Toronto) in 2017 – developed by the artist Alex Mayhew, thanks to this project the artworks displayed in the museum can be animated by means of a personalized application for smartphones and tablets. The development of the interactive dimension is fundamental to fully understand the "re-created" meaning of the exhibited works. The characters portrayed in some paintings, animated and brought into the contemporary world, begin to interact with the visitor (they leave the frame, take selfies . . .) [45]

In this way, reality and imagination become intertwined, and as a result, past art is put in relation with new generations and paves the way to creative experimentations. In fact, augmented reality as a tool used by young people to express themselves and re-elaborate their acquired knowledge, especially in education in art and cultural heritage, is becoming increasingly popular in school contexts. An exploratory survey carried out on national territory in 2018 by the research group of the Department of Education of the University of Bologna highlighted that the designing of didactic pathways through augmented reality is positioned within a constructivist and problematicist perspective [65, 66]. By means of augmented reality, students learn to explore, search, and construct new knowledge using creative methods; they structure and re-elaborate the information and the data collected and construct digital pathways/narratives, particularly within exhibition environments. This learning process is developed through successive phases: (1) the research, collection, selection, and organization of information content about artistic objects, with particular attention to the reliability of the sources and content of network; (2) the search for images on sharing sites in creative commons and/or the creation of photographic shots; (3) the construction of videos through the combination of text, images, and audio recordings; and (4) the construction and sharing of augmented content, capable of providing new readings of the artworks. Specifically, the students design didactic art projects with their teachers to be carried out in class within the field of augmented learning and learn how to observe, interact, and find out about the artistic heritage actively and interactively. Augmented reality indeed allows them to know, discover, and reinterpret each one according to his or her own vision, a cultural asset, or a place, transforming the physical spaces into scenarios of a person imaginary to be shared both with the school community and with the community of the territory of belonging. In this way, historically relevant spaces from an artistic point of view offer creative and participatory ways of learning. In so doing, augmented reality produces an involving kind of learning on the motivational level, with a significant impact on attention, concentration, satisfaction, cognitive maturity, and imagination. These aspects emerge, in particular, when augmented reality is used to re-elaborate the knowledge acquired about the artistic heritage, through the construction and sharing of personal content that allows them to provide new readings and interpretations. In this perspective, augmented reality enhances a kind of student-centered learning, in the direction of active learning and self-learning that allows the student to create study materials, the objects of learning process, in order to gradually reach the more complex stages of thinking [28]. Students thus become aware of the reality that surrounds them; they are no longer passive users but creators of innovative content. In this regard, Liu, Tan, and Chu [67], p. 173 underline how augmented reality "improves the ability to explore, absorb new knowledge and solve problems" with a consequent impact also on an emotional level. "Augmented reality arouses these emotions due to its potential to connect the power of the network, the power of technology and the power of communication in the transmission of content" [68], p. 26. A further aspect that emerges from the school projects taken into consideration in the exploratory survey [65, 66] is the relationship between augmented reality and the development of a creative mind capable of reworking and reinterpreting art objects. In this sense, some educational experiences tend to forge a bond between creative expression and enhancement of the cultural heritage, both tangible and intangible, that students feel as their own.

12.3 Digital Environments and Augmented Reality

12.3.1 Augmented Reality as Third Space

The scientific literature highlights how augmented reality in the context of art education acquires particular relevance when it represents a privileged space not only for the enjoyment of art but also for manipulation and experimentation [25, 36, 69, 70]. These aspects recall the concept of multimodality [71] used in reference to the multiple modalities used to communicate and express, through the creation and production of digital artefacts, bearers of original meanings within an eco-systemic perspective. Augmented reality thus places the user in a new space, a "third space" [23, 72, 73], where virtual and real objects coexist through a process of digital addition: a hybridized space in which digital artefacts are superimposed upon reality perceived through the senses.

According to Flessner [73], in this third space, the formal and the informal, the presence and the distance, and the "real" and the digital are recombined, building new meanings: it is possible to work to aggregate and re-elaborate materials and experiences from the first and second spaces, to reflect on them, to understand the experiences lived in the informal with the lens of theory, and to rethink theories based on experiences. In this sense, digital environments as a third space are also augmented reality environments that in the contexts of art provide tools to personalize, build, enjoy, and share new artefacts. This third space also looks like a space of flows as augmented reality technologies connect different places through the subjective practices of individuals who relate to each other; but it also looks like "distributed intelligent space in which it becomes increasingly difficult to distinguish between real and virtual and where mobile technologies mediate the experience of a new sense of space, which we can call augmented" [24], p. 22. Thus, emerges the possibility of designing innovative art education activities based on augmented reality in relation to new approaches for accessing and building knowledge. In fact, increasing reality also means providing places for cognitive growth in which forms of collaborative construction of knowledge and skills that can be implemented gradually which, in the theoretical field, refer to connectivism. It attributes to the condition of always being connected a key value for the development of knowledge in the digital age [34], p. 179. Hence today, with the spread of the visual richness provided by technology and the development of the possibilities of interaction, digital environments become the place where both the instances of creativity and expressiveness can be connected in the educational sphere, as well as those of design and experimentation, to arrive at a concept of renewed and expanded knowledge [74]. The third space of augmented reality is provided by the combination of several elements: the application, the content, the interaction, and the physical environment and subjects. The application is the program that allows us to organize and control the different aspects of the augmented reality experience, including the recognition of the physical world in reference to digital content and the synchronization between the physical and virtual world, in order to add digital elements to the user's vision. Augmented content, the digital layer, includes all objects, ideas, stories, and sensory stimuli. The interaction allows the user to observe the digital layer from different points of view or perspectives. Each augmented reality experience is closely linked to the physical environment in which it is made and to the real object that is augmented with virtual information. The actions and movements of the subjects influence the entire system of creation and reception of digital artefacts in AR. The experience with these artefacts is configured as a process or an evolution in space and time of actions, facts, and behaviors linked to each other. Thus, the perspective of a third space between real and digital is realized, which offers different training opportunities: cognitive and affective purposes, historical adherence and scientific rigor, and realization of expressive-creative experiences with high coefficients of imagination. With reference to this meaning of AR as a third space that enhances the dimensions of knowledge, creative re-elaboration, and participation in the artistic and cultural heritage, two AR experiences are presented here: MoMAR and Snapchat augmented reality. The first was created in a specific museum context, the MOMA in New York, and the second created in collaboration with Jeff Koons on Snapchat Art.

MoMAR stands as a particularly significant experience that through AR transforms a traditional museum setting into a third space. In 2018, a collective of eight artists digitally transformed Jackson Pollock's gallery at the Museum of Modern Art in New York, with the aim of democratizing public spaces for art. The performance entitled *Hello, we're from the Internet* – preceded in 2010 by an analogue intervention by the title WeARinMoMA – acted on two perceptual levels: the first is visible to the naked eye and the second is visible only through the app. With reference to the first level, visitors were able to observe Pollock's canvases, as they appear in physical reality; in relation to the second level, visitors were able to enjoy and interact with the original reinterpretations made by the artists of the collective. The gallery space thus became a third space, and Pollock's works were the interfaces for access to new forms of visual interpretation. One of the AR installations, created by artist Gabriel Barcia-Colombo, transformed the White Light painting – one of Pollock's last paintings before he died in 1956 – into an interactive game. In the game, small skeletons quickly climbed up all sides of the painting. In this regard, the artist claims that "I wanted to make an experience that played with the existing form but also commented on the painting itself" [75]. The installation created by Damjanski, called *One: Number 12811912112811950* (2018), presented visitors with a gif that merges the faces of Pollock and Ed Harris, the man who played the artist in the film Pollock: "I started questioning the value of representation and manifested these thoughts in a hybrid character of Jackson Pollock and Ed Harris playing him (. . .). This new character interrogates the lines between fact and fiction and what's 'real' and 'fake' [75]. Artists have made their MOMAR app open source as a means of encouraging participation, through release of an instructional PDF that allow anyone to make changes, even without specific skills in coding. The artist highlights: “it’s the first iteration of a set of instructions to give people the power to show their work in any physical exhibition space around the world” [75].

In 2017, Snapchat launched an innovative project of artistic enjoyment which consists in proposing the works of Jeff Koons with augmented reality, showing them as 3D sculptures on the screen of one’s mobile device. In fact, Snapchat allows us to admire artworks that do not exist in the physical reality, and for this purpose, it has published a page dedicated to artists, called Snapchat ART [76].

To launch this initiative, Koons has made available some of his sculptures which, elaborated in three-dimensional graphics, are virtually placed in various parts of the world. Through Snapchat’s World Lenses function, users can admire and photograph Koons’ works in augmented reality, also thanks to a notification message they receive when they are near the places where the virtual works of art have been placed; or they can consult the SnapMap to go directly to the places indicated for use (Central Park in New York, Hyde Park in London . . .). The sculptures thus become 3D stickers that users see only on their device’s screen that they can also be added to their photographs or videos [77]**.** In terms of participation, other artists can participate by adding their work to the virtual collection. A Chilean artist Sebastian Errazuriz even reworked a Koons augmented reality sculpture, Balloon Dog, situated in Central Park, covering it with tags, by superimposing a duplicate of the modified work upon the original. The same collective act, this time on the level of participatory enjoyment, determines the ‘visibility’ of Koons’ augmented reality sculptures. In this specific case, it is the users/viewers who recognize the ‘presence’ of the artist’s works in well-defined places and ‘materialize’ them on the display of their mobile device.

12.3.2 Augmented Reality for Creating Digital Artefacts

In educational contexts, the application of augmented reality is declined in two different ways:

1. As a *support* to students/users who, within an “augmented” environment, discover and frame the graphic elements that give access to the training content useful for describing and narrating a specific experiential context; in this context, augmented books are also included “which aim to enhance printed communication through the placement of paper markers on the covers and pages (. . .): photographs become films, paragraphs come alive and the columns give way to graphic and multimedia objects” [53], p. 43.
2. As a *product* of the students/users who, after having learned the basic principles of designing through augmented reality, create the augmented content with their smartphone/tablet, starting from the research and selection of study materials considered interesting and relevant. In this way “students can construct the contents and place them in context using their mobile phone or tablet, share them with other students, who in turn can add further contents” [34], p. 135. In this sense, the digital artefact [78, 79] represents a form of knowledge processing supported by multiple languages and multiple ways of connection, starting from the subject’s experience in formal and non-formal contexts. The graphic-visual dimension represents a fundamental characteristic of this typology of artefacts, whose design and construction refers to different intelligences, including the graphic-visual one. In fact, with reference to the theory of multiple intelligences [43], visual intelligence is in a complementary relationship with graphic intelligence and defines the fundamental cognitive abilities of cognitive processes. In this regard, it is recognized that the digital graphic-visual artefact has a strong impact on the motivational context of the learning subject (pleasure, need, duty). In this sense, the visual digital artefact is part of a global learning process that integrates cognitive, socio-relational, and emotional dimensions [79].

The design of these digital artefacts in AR is based on some fundamental actions: search and select information; build the texts, images, and videos that make up the overlays; associate the augmented content with objects/art images with a trigger function; and share augmented content [80]. The production of digital artefacts goes beyond the pure theoretical dimension, promoting “different thinking styles, preparing for creative and divergent solutions to the problems of contemporary life” [81], favoring the implementation of an operational methodology of discovery, exploration, and research. With regard to the digital artefacts made and accessible via AR, here we present the experience of the MAUA – *Museo di Arte Urbana Aumentata* and that of SketchAR, a specific AR application based on artificial intelligence.

The MAUA – *Museo di Arte Urbana Aumentata* (Milan), founded in 2017, is a sort of open-air gallery dedicated to animated street art that proposes novel cultural itineraries in the five most degraded neighborhoods in Milan, through the involvement of the residents themselves. Indeed, in the initial design phase, the citizenry was asked to select the street art works most representative of their own area, which were then visually documented by students and neighborhood associations, along with teachers of the school CFP Bauer. Among the documented street art works, 50 were selected that were then elaborated by just as many young animation designers. They transformed the street art works into augmented reality producing 50 digital artefacts. Today, the museum is made up of these 50 animated street art works in augmented reality. Specifically, the access to each of these works distributed in the peripheral neighborhoods of the city consists in the fact that each work can be framed with a mobile device (smartphone or tablet), generating a new work of digital art. On the level of the participation of the citizens, the cognitive dimension was activated with respect to those works of street art deemed to be most representative through a careful selection and documentation. This first level of acquaintance was followed by a phase of creative re-elaboration performed by young animation designers. They elaborated the images in augmented reality producing digital contents that animate the selected street art works, recreating and transforming them into digital artworks. In this way, the augmented street artworks, whose location is marked from one time to the next by one's own device, have become the chance to explore and know the city's peripheral neighborhoods. The *Museo di Arte Urbana Aumentata* has thus proposed a diffuse and participated model of museum to attract a "particular public," made up of the inhabitants of the most degraded urban outskirts. The involvement of the inhabitants beginning from the phase of the selection of the street artworks has been translated into "an advanced experiment of diffuse curatorship" which has envisaged the collective identification of the works and the sharing of their meanings. The production of augmented reality digital artefacts has marked the shift from the condition of curatorship to a singular and novel experience of participated visit [77].

SketchAR is a mobile app based on artificial intelligence – from artificial vision to automatic learning, to neural networks – finalized to developing the dimension of creativity in contexts of art education, putting forward an interactive approach to drawing in augmented reality. Any AR product is indeed based on artificial vision. In this specific case, SketchAR uses a tracking system of markerless artificial vision. The application uses a technology based on the computerized vision of augmented reality, which allows for the space framed by the smartphone to be scanned and "set" the virtual image on a real surface. This application with the feature of personalized virtual assistant refers to a machine learning system (automatic learning) that can help both the users to understand how to draw a particular object and to analyze the habits of the users themselves, with respect to the drawing methods adopted. An algorithm collects this information and supports the users so that they can achieve their objectives in the graphic activity. Developing this core of artificial vision, different methods have been created: "progressive markers" and "predictable markers." This means that the lines traced by the user transform into an anchor and improve the retention of the virtual object. An increasingly relevant relationship thus unfolds between action and knowledge and between action and perception [82–85]. The augmented reality environments show characterizations relating to that specific milieu of study that analyses how the life spaces are structured in relation to the digital media, whose characteristics and functions need to be understood. Augmented reality thus represents a third space, syncretic and multimodal in nature, that brings into play a plurality of languages and heterogenous expressive forms in a unitary strategy of communication, finalized to the production of digital artefacts. Within a pedagogical-didactic reading, these characteristics of augmented reality thus allow one to create a context particularly suited to the valorization of the cultural heritage [86].

12.4 The Language of Images and Augmented Reality

12.4.1 The Audio-Visual Language

Augmented reality is a visual method of presenting digital information that corresponds to the need to find ever more engaging and realistic ways of expression. Augmented reality can be used in a wide variety of application domains [87–90]. It can be defined as a new form of audio-visual, together with other types of digital video, which are both texts and experiences [91–93].

As such it does conceal some pitfalls. It seems to be an immediate form of communication, but in actual fact, albeit fascinating, it is a complex system. This is shown by the fact that if to write a book, you simply need to be a good writer; and to produce an audio-visual, you need different skills that correspond to different professions. The question "who is the author of a film" always leads one to reflect on the complexity of language and obliges one to reply in the plural form: scriptwriter, director of photography, editor, musical composer, and director are just some of the skills involved in cinema production [94–100].

In an interview given to me for the Media Education LAboratory (MELA) of the Department of Education Sciences of the University of Bologna, the renowned Italian screenwriter, author, and television host Carlo Lucarelli stated: "The audio-visual works have a true force, which we take for granted. They belong to a type of imaginary according to which something that is seen is simpler and more immediate that one that is read: but that is just not true. It is, however, always true that if you suggest to someone, above all a young person, that they should read a book or see a film, he or she will think they will understand things much faster and more easily by seeing a film. This makes the audio-visual language a good vehicle, but not a simple language. It is a good vehicle because it draws the attention and curiosity that can then be shifted onto the written text, essential for getting into the complexity of the message".

Audio-visual language is not easy to understand because it is a sort of macrolanguage, a system of different languages that together form something more complex than the sum of the individual parts, each of which has its own ancient history and comes from preexisting worlds to the film/television one.

The Language of the Texts

Every audio-visual product has a story to tell, a narration that is based on a written text, whether it is a simple schedule, a story, or a full-fledged screenplay.

The Language of Filming

The language of filming derives from the pictorial language, going via photography, with the addition of the peculiarities given by the movement and the specific aspects relating to the special effects.

The Language of Lighting

From painting derives the art of illuminating that comprises the choice of the quantity of light, its distribution in space, and its color.

The Language of the Setting

The choice and the reconstruction of the milieu and the objects to have interact with the characters instead derives from the architecture, the design, and the scenic theatrical language.

The Language of Characterization

This deals with the external image of the characters that must be coherent with the history and the settings and derives mostly from theater.

12.4.2 Body Language

Movements, gestures, and mimicry are as important as the dialogues and derive from the theater.

The Language of Editing

This is the most innovative and specific language since it does not derive from the previous arts.

The Language of Sound

Ambient sounds help to create the overall atmosphere and affect the general impression that the viewer gets from the work. It is a specific variant of the sound language used in radio programs.

The Language of Music

Music contributes to giving emotion to the story and transforming a simple narration into experience.

The Language of Graphics

This is brought into play to produce graphic interventions and titles.

The use of the above-said languages in the forms of augmented reality, at least in the more evolved ones, conditions the successful outcome of the final product, characterized by many narrative facets. The text from which to start for the narration, any recorded images, the quality of the light, the forms of the setting, the images of the characters and their movements in space, and the montages of the clips with sounds, music, and graphics correspond to just as many design choices and make manifest the complexity of the narration with the audio-visual language.

12.4.3 Educational Experiences

Sectoral studies show how visual perception is closely connected to the mental processes of exploration and selection: these are two activities that come into play when we visually perceive the elements of a context. The gaze is positioned on one object rather than another, thanks to the attention mechanism [101, 102]. The visual elements thus perceived then enrich the subject's cognitive structure, conditioning the successive perceptive act, in a continuous process that involves the eyes and the mind together: perceiving and thinking are two interconnected moments that call for an active engagement of the mind, so that it can be stated that visually perceiving is thinking visually. Visual intelligence grasps in the context single visual elements that are associated to mental categories, proceeding by way of resemblances and associations [103, 104]. The objects and concepts are associated with patterns of reference, becoming full-fledged mental images. The process is influenced by emotive involvement: the more meaningful the experience, the more efficacious the perceptive experience is and the more long-term memorization is facilitated. As in impressionist painting, the images are not the photograph of the real but the result of impressions that is of a selective and perceptive process conducted through the emotions [105, 106]. In that sense, such a process enriches the cognitive structure, thus configuring itself as a full-fledged form of thinking and intelligence. Visual thinking has in its elaborative-constructive function the most interesting aspect for the individual's cognitive growth, as it allows for the creation of mental images of the facts. Indeed, the visual instruments enable us to communicate and at the same time to structure ideas; an example of this are the mental and conceptual maps, graphs, tables and diagrams, animations, simulations, and virtual realities [107].

Moreover, visual thinking not only elaborates perceptive data but also creates new structures that, in turn, have a generative power vis-à-vis ideas. Indeed, the image is an interpretative model and not a faithful representation; for this reason, visual thinking is closely connected with the creative processes.

With reference to the theory of multiple intelligences articulated by Gardner [43], visual intelligence is positioned in a relationship of complementariness with graphic intelligence: it defines the cognitive abilities connected to the imagination and the capacity of "thinking by images" that is to mentally portray the concepts even before verbalizing them, allowing one to make an immediate experience of the world; graphic experience concerns the capacity to integrate perception, thinking, and representation of reality to create artefacts finalized to the acquisition and the construction of new knowledge [103, 108, 109]. The vision/production of an image allows the student to activate cognitive explorative processes, those of categorization, memory, prediction, comprehension, emotion, and empathy [41]. In this regard, Clark and Lyons [110], in identifying some functions of the images concerning attention, the activation of knowledge, the minimization of the cognitive load, the support to motivation and, in particular, the images, can exert a function of mediation, anticipation, and modeling with respect to knowledge [4].

Images and Learning

Images play a precise role in the formation and development of learning. Historically, illustrated fairy tales for childhood have represented a language endowed with an autonomous force, capable of constructing their own discourse on the text of which they become an alternative narrating voice. The illustrations often represent the child's gaze where everything appears huge and boundless. At other times, they strip the fairy tale of its metaphorical coating and reveal the underlying message with a precise pedagogical intentionality. That master of creativity and fantasy, Gianni Rodari, enjoyed reading comics, images in succession inserted within environments that have a narrative function as much as words did: in the passage from one image to the next, the child must carry out operations of recognition and connection, making an effort to fill with meanings the blank spaces between one vignette and the other [111]. It is the game between Antoine de Saint-Exupéry's *Little Prince*, who by drawing his famous lines asks his aviator friend, and thus the reader, to recognize the drawing as it gradually evolves and to imagine the "implicit" sense in the passage from one to the next.

The birth of cinema and moving images has caused the use of a language, that of the audio-visual, capable of fostering particular psychological and emotive conditions, thanks to their undoubted seductive power. For this reason, the moving images accompanied by sound have often helped the adult in their most authentic educational task, that of teaching, contributing at the same time to developing the need to train children towards a critical sense and the conscious reading of the visual stimulus [112].

The "educational value" of the film has been recognized at various times and with different valences through the twentieth century. The pedagogical debate, among others, has given rise to opposing positions between those who believed in the utility of film for the transmission of contents and those who, focusing on the cinematographic experience in general, identified the educational function in the emotive strength and in the undeniable psychological stimulus of audio-visual education [113]. The two different visions have led to different uses in school: on the one hand, documentaries were searched for, in which the cinematographic language was conducive to the educational purpose; on the other hand, there was the tendency to use films chosen on the grounds of ideological or moral criteria, with the aim of showing models to be imitated or from which to draw some teachings. Only at the end of the twentieth century was there an overturning of the reflection on the didactic use of film, reaching the conclusion that it is neither necessary nor useful to look for a film with a clearly educational purpose: attention ought to be placed in the didactic project, thus in the process of teaching and in the strategies deployed to insert the film everyday school life [112].

Digital Images and Media Education

The culture of information today is mainly visual culture, and digital media are the way in which information is presented in various areas. The digital gives image a new dimension and unexpected perspectives. The computer monitor revives the cinema screen in unpredictable contexts and gives audio-visual communication a central role [54, 114].

Digital media objects, including films, are the new cultural products and affect the kind of experience through relations and the environment. They should be read by means of a pragmatic paradigm, as for the Internet in general, according to the rationale of the practices of knowledge-building and exchange of meanings that are rooted in the contexts of belonging of the individuals, albeit in the presence of digital mediation [115]. The educational experience also requires a spatio-environmental space in which to give rise to relations and to learning, thanks to a precise pedagogical intention. The digital spatial environment has its own characteristics that lead to new meanings and to different modalities in knowledge-building [116]. In this scenario, the artistic dimension of digital audio-visual language requires a greater effort in the direction of a media education understood as education towards reading and the interpretation of such a language, to start distinguishing, albeit in the multiplication of the spatial environments of interaction (from the cinema to places of public projection, from the single device to the monitors distributed in the indoor and outdoor public spaces), the styles and the functions, the genres and the perspectives, the message and the meanings, and the art and the propaganda.

Today media education must find a way to renew itself, and it can do so by adhering to a pedagogical design that, in line with the resources of the territory, can integrate different cultural and social perspectives, communicative multimodality, active didactic strategies, and those for the social knowledge-building, adherence to the everyday. What we feel we need is not so much a "technical" education (relating to the procedures and the resources) as a broader education relating to the horizons of meaning and to cultural and value elements. For this reason, we suggest a declination of the concept of media education that corresponds to the following paradigms of reference: "political" education, which forms the instruments for an autonomous knowledge-building, interpretative capacities, project-building, and the choice of a definite cultural direction; "critical" education for the reading of the image and for the recognition of the fundamental principles of the audio-visual grammar and the rules of cinema; and "artistic-creative" education for the development of minds sensitive to art and capable of recognizing the high-quality artistic productions and to embrace a creative attitude [8, 117, 118, 119, 120, 121].

12.4.4 Augmented Reality and Educational Experiences with Arts

Designing and developing augmented digital environments is an opportunity to enhance digital communication skills in the production and dissemination of images and knowledge relating to art and heritage.

The digital environment can be used for a simple transmission of contents but also to favor cognitive and perceptive immersion in spaces and objects, thanks to particular techniques of processing moving images and sounds that allow us to live real-life, sensory, cultural, and artistic experiences. In any case, we can also speak of learning experiences, as the subject develops knowledge and skills based on the relationship with objects, materials, and intellectuals and on sensory and emotional actions and perceptions.

If we allow ourselves to be oriented by the pragmatist principle of John Dewey, the aesthetic experience can overcome the merely contemplative function to foster full-fledged knowledge processes. For this to come about, the experience has to be the result of the interaction between body and environment and must develop aesthetic qualities and perceptions that have a very intense emotional character. It is above all emotion that leads to the accomplishment of the knowledge process by generating meanings: if the situation, in this case created by the artist but that can also be produced by a teacher during educational moments, generates emotions, then it develops new perceptive and cognitive modalities. In Dewey's conception, experience can also be just intellectual and not necessarily tied to objects or materials; it must however be a right balance between actions and passions to generate knowledge. The teacher is thus compared to the artist as they create situations in which both the object of the learning and the subject that is learning play a role, through action and experimentation of situations, tasks, and roles. So, the quality of the experience and its emotional valency determine the quality of the learning itself [122–124].

The Deweyan reflection leads us today to rethink the relationship between art and technique. In particular, in the artefacts of augmented reality, which use techniques of the figurative arts and multimedia languages, the interactive and virtual media acquire a role of redesigning reality that goes beyond the mere contemplative or transmissive function to acquire, in the relationship with art, an educational role as learning mediators. Being a question of media of an iconographic type and strongly characterized by moving images, they use the audio-visual language in a predominant way. As for cinema, for these products as well, we can wonder what relationship they have with art, on the one hand, and education, on the other hand.

The artistic and education worlds have a space of intersection, as we have also been taught by Dewey, which can be inhabited by many expressive realities among which also cinema, with its languages, its forms, and its knowledge mediators. This is an active space, which garners structures and stimuli from both worlds ending up resembling a little of one and a little of the other. The fact that cinema is an art has been proven in the twentieth century by the "Philosophy of Cinema." Film is considered to be an art form in that it is the means with which to arouse thoughts and emotions and to arouse questions, reflections, and new world views [125]. The concept of experience returns, in that film, besides being a cultural artefact, offer the spectator the fullness of a living experience, since it includes temporality and movement, a multiperspectival vision, and the emotional impact. Its extraordinary capacity to offer multiple points of view is ensured by the possibility to put together what the character sees and what the camera captures, the first and the third person, to tell with the intermediation of different techniques of shooting, photography, montage, and sound [126]. Viewing becomes an experience and an artistic experience even before being technological. As an artform, it survives in time before every expansion of contemporaneity and thus before every change of the means and of the technique [127, 128].

The relationship with technique is portrayed in the image as a separate bubble yet immersed in the same fabric of experience. Together with it, there is the other element characterizing art, education, and cinema: dream. Cinema is the expression of the soul and of the sentiments: it paints the soul of things and at the same time the artist paints their own soul in things [129].

The primary vocation of cinema is poetry and the dream is a predominant part of it. The "cinema of poetry" is the artistic and metaphorical experience of the experiences and the emotions [130]. For Pier Paolo Pasolini, cinema is the artistic language, in that it is an arc and never a direct or philosophical conceptual expression; it belongs to poetry and not to the novel or to theatrical writing. Cinema contains an irrational element that cannot be eliminated. The effort to transform cinema into pure technique has produced the effect of pushing back the unconscious and oneirical element to the background, concealing it from a superficial vision. This has allowed for a manipulative use of cinema, the search that is of a form of rationality through the adhesion to pre-established formats and the productions of standardized films.

On the contrary, the director who proposes an art-house movie is aware of using a non-conventional vocabulary because it is their vocabulary, the one relating to their own ideological and poetic vision of reality. True cinema, as a consequence, can only be metaphorical, that is the one in which the dream has a predominant and ineluctable part [131]. Pasolini often reiterates the audio-visual nature of cinema, in which the image, words, and sound have the same importance and contribute together towards the final product. Their connection, which is above all implemented in the montage, creates endless expressive and stylistic possibilities and requires of the spectator an elevated competence in the reading and interpretation of the audio-visual product that modifies, as narrative, the relationship between man and the reality represented [132].

The artefacts of augmented reality contain many of the technical and narrative elements of cinema. Their production and enjoyment recall the experience of production and enjoyment of the audio-visual products; their strongly multimedia connotation requires the same knowledge of the audio-visual and multimedia language that also a film puts into play. The Web is one of these spaces, as are the mobile applications and the augmented reality software, which read and reconstruct narrations by means of images and sounds. We believe that to make these experiences really educational, there also has to be the contribution of the characteristics of the aesthetic and cinematographic experience, with the languages of poetry and the emotions of art.

12.4.5 Augmented Reality: Art for the Sake of Art

Communication is one of the primary functions of art, by virtue of its potential universality and its being intrinsically expressive [123]. Bruno Munari [133, 134] starts from this assumption to theorize a new concept of the artwork, highlighting the communicative function and the visual language it uses. The artist, or in any case the author of a visual work, albeit with their own personal and intimate vision of the world, must worry about transmitting messages that are as objective as possible, so as not to risk entering the world with personal codes, so that certain messages are only understood by a few people.

Munari today would probably look carefully at the forms of augmented reality applied to art, as they offer the opportunity to build stories in a digital visual format, like the project "Tap the artwork" (ARTAP) funded by Heritage Srl, a start-up founded in Turin in 2013 that operates in the field of Smart Cultural Heritage. The aim of this project is to create a guided museum tour by trying to integrate the new technologies with the narration of a story. The user is put at the heart of their experience in that the application aims to make them responsible in the active learning of knowledge and is not limited to the mere transmission of knowledge; the individual is called upon to take part actively in the structuring and the personalization of their museum tour.

Specifically, ARTAP is proposed as an instrument with which to utilize a museum service whose objective is the involvement and satisfaction of the user, thanks to the use of the digital storytelling technique. Indeed, the narrative element and the meaning of the works that must be clear and accessible are important for the understanding of the visit. The narration technique not only provided contents but poses new questions, novelties, and stimuli for the user who participates in the construction of meaning.

The visitor follows a narrative that guides them in the exploration of the works and their meanings through insights, settings, audio or video supports, visual elements, and a geo-localization of space; these elements overlap the physical work, thanks to the mediation of a device: to access these ways, just "tap" on the work via your smartphone. The app provides the chance to view a narration on the screen overlapping the framed works, increasing the reality that the user experiences. The user can go back, visualize the information several times, and interact with it. ARTAP integrates the narrative dimension with the new technologies such as Beacon, touch mobile, Virtual Reality, Content Management System, and the new forms of communication: storytelling.

12.5 Digital Competences and Augmented Reality

12.5.1 Augmented Reality and Digital Innovation

Following the Covid-19 pandemic, which has led to rapid and profound cultural and social transformations, museums have found themselves having to face a series of important challenges. In fact, if before then the museums had been using augmented reality apps, audio-guides, visors for virtual tours, stations for virtual reality, and immersive installations, only upon completion or in support of the experience of the visit, with the health emergency have they increased their online presence by delivering contents and offering new cultural and educational provisions [135–137]. In this scenario, there has been a renewed awareness that questions the future of the museums in two directions: the development of a digital strategy of audience development in the mid-long term, for an access to knowledge that is increasingly open and democratic and a necessary investment in the training of museum personnel that will have to work to make the assets accessible, also on the digital platform. In this sense, the museum is an experience an organizational and cultural change that coincides with the recent definition of the "phygital" museum [138] in which the physical space and the digital space are part of an ecosystem, and the availability of instruments of mediation in the museums, on the territory, and on the digital platform translates into new possibilities of learning and participation. The investment that derives from it is thus based on the development of digital competencies for all the museum personnel [139, 140], in a sector in which these competencies are fragmented [141, 142, 136], and the internal planning of the training activities is not very apparent [143]. The museum professionals will thus have to be involved in dealing with this challenge by asking themselves about the methods and the instruments that can progressively be adopted, while also paying particular attention to sustainability. In this sense, rethinking the use of easily accessible digital devices may result to be relevant not only for the public but also for the personnel of the museum who, inside the workgroup, will be involved in the various phases of analysis, design, and development of the applications. In this regard, in view of the potential of the use of augmented reality (AR) in the museums, it emerges from the many studies [144–146] how this technology allows the experience to be more involving and interactive, providing new learning opportunities and a personalized access to the contents. If then AR can be considered as a technology that can bring an added value to the learning experience, less evident is instead the reflection in the accessibility of the use of the instrument by the museum personnel. It is indeed presumed that in the upcoming future also the museum professionals will deal with designing and developing new applications in which AR will be ever more interconnected with artificial intelligence systems (AI). Indeed, it is very likely, as demonstrated by the new developments on the subject of AI, that the demand for the skills required will change as the supply will change, new and different ways of leaning will emerge, and this will lead to a reorganization of the activities and a rethinking of the educational systems of the professions that will have to adapt to the changes in society [147]. In the specific case of AR, it is evidenced how this technology is already amply used in many contexts and will become, in the near future, ever more indispensable in the educational environments [148].

Considering, therefore, the design and the development of the AR applications for didactic purposes, it is believed that the museum educator can become, along with other professionals, a principal reference for the museum. In this regard, the project Mu.SA [149] has worked to define at the European level four new professional profiles indicating, for each of these, the digital and transferable competences. Specifically, the workgroup has identified among the professions of the future the Digital Interactive Experience Developer who specifically deals with "designing, developing and implementing innovative and interactive experiences providing a meaningful experience for all types of visitors" [150], p. 48. This figure, who has among their key responsibilities those of "carrying out audience research and observation analysis; developing accessibility tools for all types of visitors; facilitating communication flows between various different museum teams and external high-tech companies" [151], p. 7, seems very closely related to the profile of the museum educator who has among their tasks that of promoting the assets to a heterogeneous public, designing, and managing recreational activities and cultural animation, performing activities of monitoring and evaluation [152]. The museum educator, adequately trained in the digital field, can put at the disposal of the museum their pedagogical and didactic competencies to adapt the AR applications to the demands of the users [153]. The museum educator can also design the didactic pathways on the digital platform, contribute to the creation of augmented educational contents, and propose different modalities of navigation of the contents to the visitors, depending on their learning goals.

Before the exponential increase in digital resources available on the Web, the choice of reliable and authoritative sources for the creation of augmented contents turns out to be a meaningful practice to be pursued. In this regard, the joint work with the curators, archivists, art critics, and communications managers is relevant in order to evaluate the reliability of the sources that will have to give value to the heritage and at the same time be accessible to the public with involving and multimodal languages [71]. The knowledge of didactic methodologies and strategies of mediation with the heritage allow the educator to design the education pathway according to precise pedagogical goals, not oriented to entertainment for its own sake but to a broadening of the knowledge in which the cognitive dimension is added to the emotional and social one. Another important collaboration for the museum educator is that with the curators, artists, and public who are involved in the production of the contents. In this sense, creativity is not born from the technology in itself;

this must be fed by an intellectual activity that avails itself of technological instruments to give rise to novel creations.

If, as has been said, AR technology allows the visitor to enrich and deepen the visit by overlapping additional information to what is already there, on the design side it is a matter of deploying targeted choices, antecedent and successive to the pathway, on what the visitor will find themselves observing, on the time of use of the device, on the behavioral responses, and on the movement in space. For the sake of example, the locator system (GPS) allows the user to be more easily guided to find points of interest in the museum space, but, at the same time, the technology can also outline their itinerary during the visit by providing the museum personnel with important spatial data. The user's experience, which is put at the forefront, thus becomes the object of observation for the educator who, in relation to the feedback he or she receives and the interactions in the contexts, can study his or her public and work for the overall improvement of the educational experience.

Again, regarding the design, the museum educator can choose how to make meaningful a specific didactic itinerary, including AR devices with more traditional routes, along the lines of the guided tour. In this way, the educator remains a fundamental figure who accompanies the visitor through the museum collections, and technology becomes functional to the visit in that it enriches it. Furthermore, it can happen, as demonstrated by some devices already available in the museums, that the tour is conducted by a digital guide, in which case the museum educator can study and choose the most suitable and enticing modalities (visual, tactile, auditory) to interest and involve the users. By way of example, an AR application can be used in the course of a guided tour to allow the user to deepen the augmented contents individually at first; subsequently, the educator can return to the traditional tour by having emerge questions and observations in the wider group. The development of associations between the information added by the educator and those yielded by the technology, together with the personal experiences of the visitors, can enrich the learning by fostering the development of new interpretations. The choice to proceed in this way cannot be random but is connoted by a precise conceptual positioning that inevitably acts as a guide for the design of the application. In the specific case outlined here, this conceptual choice, always more in line with the developments of the contemporary museum, refers to a constructivist learning model in which known is situated and distributed and in which the visitor is not considered as an acritical consumer of knowledge but as an active user, stimulated to search for new meanings, starting from their own prior experiences.

The AR devices can be designed to become flexible instruments of self-learning, allowing the user to construct their own pathway in complete freedom; the prior knowledge is thus combined with the new experience that can be enriched by way of the interaction with other subjects. In the design of AR devices, the social dimension of learning is thus taken into account by the educator who constructs didactic pathways in which the user can share their own experience and compare with other users to reflect on the experience itself.

To conclude, the AR technology, if properly valorized, can increase the value of the heritage and at the same time give a meaning to the individual and social experience of the users. It thus follows that museum education can become a resource for the development of innovative AR approaches. The museums will be called upon to promote routes of continuous training for their professionals. It will also be necessary to revise the role of museum educator that, as we have tried to show in this contribution, is the bearer of a corpus of methodological and didactic competencies that still need to be valorized [154, 155].

12.5.2 Augmented Reality and Digital Competences of Museum Educators

The current modernization, digitally speaking, of the museum and the skills of the different profiles connected to it further valorize the role of the museum institution as an intentionally educational agency, both physical and virtual, capable of existing, alongside and in a synergic relationship with other educational realities, for the raising of the educational and cultural quality of society, valorizing the cultural assets, and offering to the whole citizenry educational opportunities for the construction of the transversal competencies of identity, civil, social, digital, values, entrepreneurial, and citizenship [156] that can contribute to making them critical and responsible actors.

The professionally trained museum educator should embrace such demands and consider the acquisition, on the part of the different publics, of digital, media, and data literacy skills as an objective and a strategic opportunity to answer the challenges of the twenty-first century, in view of an active, informed, reflexive, and critical participation creating a sense of responsibility towards the process of construction and co-construction of knowledge within an everydayness increasingly connoted by the dimension of the so-called onlife [157]. The concept proposed by Floridi marries and interweaves with that of augmented reality: the AR experiences are placed *in continuum* vis-à-vis the real experiences inside the museum, positioning the publics in a hybrid space in which the body and the experience are at the center of the learning processes and in which the reality and the physical environment are enhanced, broadened, integrated, and blended in their different components, dilating vision and perception and stimulating the learning processes in a novel way. We can consider AR to be an ingredient of creative complexity, thanks to which the spatiotemporal limitations of the present, of the "here and now," are implemented by proposing to us "new where's" and "new when's," shifting the experience of the subject to a "where" and to a "when" that are, indeed, "augmented."

On the grounds of what has been stated, it appears opportune to train professional capable of promoting in the users' reflexivity, metacognition, critical skills, participation, and active knowledge-building. Given the critical aspects at the international level regarding the recognition of a specific professional role of the museum educator [154], the present contribution aims to outline some of their ineluctable competencies.

Starting from some of the peculiarities of the profile of the Digital Interactive Experience Developer [140, 150, 151], we propose a conceptual and operational framework [144] which considers the museum educator as a professional who relates in a systemic way with the publics and with other museum professionals. Such a professional is capable of designing, realizing, managing, and evaluating *diffuse* didactic-educational pathways, accessible and sustainable, both inside and outside the museum, via the use and the personalization of interactive and innovative installations based on the needs of the different publics who exploit the potential of the physical, digital, and augmented museum environments for the promotion of the museum heritage and the digital and key competences for lifelong learning in the different publics. We consider first of all as the pedagogical scenario of reference the *distributed* TPACK framework [158, 159], a model that outlines the knowledge domains underpinning teaching/learning processes in which technology plays a substantial role [160]. As a consequence, the digital competencies of the museum educator will have to interweave, on one side, with those relating to the museum heritage, on the other, with transversal pedagogic-didactic competencies relating to a design that takes into consideration, among its resources, the potential offered by AR as a further *third space* [23] to be explored by the different publics, on the grounds of the specific needs – also special ones – and in consideration of different interpretations of learning. While from the contents point of view, the museum educator will be supported by several different figures, such as by the collections curator, archivists, art critics, and the artists themselves, from the digital point of view, they will have to possess competencies relating to the five areas identified by the European frameworks of reference [161, 162], that is, *information and data literacy*, *communication and collaboration*, *digital content creation*, *safety*, *problem-solving*, shifting from levels of "foundation" mastery, and "intermediate" to the levels of a mastery "advanced" and "highly specialized." From the pedagogical standpoint, it is a matter of integrating in the design of inclusive and sustainable museum pathways that exploit the potential of AR, the different methodologies, and didactic approaches that hark back, in particular, to three specific dimensions of learning [163] interweaving them with the eight key competences for lifelong learning [156], in order to determine goals of cognitive, social, and identity competencies to be achieved.

In particular, we suggest here the following three different exemplifications.

(a) *Learning by Doing and Cultural Re-elaboration*

Learning by doing and *cultural re-elaboration* involves cognitive domain of *remembering*, *understanding*, *applying*, and *evaluating*, together with digital competencies of *information*, *data literacy*, and *safety*, with particular attention to knowledge dimension as cultural reproduction of the museum heritage *in continuum* between real environment and AR. The museum educator must develop that critical thinking essential for researching, selecting, analyzing, comparing, and evaluating the credibility and the reliability of the data and the digital contents to create their own strategy of critical and responsible navigation. It is a question of acquiring specific competencies for organizing, managing, archiving, recovering digital museum data and contents, re-elaborating them, and structuring them considering the needs – also special ones – of the different users, to propose educational pathways in real, virtual, and *augmented* environments that are safe both in terms of reliability of the contents and in terms of safety of the augmented reality and the protection of personal data. The ultimate objective is to propose to the different public experiences of enhancement and broadening of the vision of reality and the physical environment, through a perfect integration between real context and virtual objects; "entering" the art work, flanking the artist is their different phases of realization; knowing a physical object also through specific localized digital information; knowing landscapes, scenarios, and architectures through their augmented historical reconstruction; and broadening the knowledge about a given phenomenon, event, etc.

(b) *Learning by Construction*

Learning by construction involves cognitive domain of *metacognition*, *creating*, and *"knowledge building,"* together with digital competencies of *communication*, *collaboration*, and *problem solving*, inside of which we find dimensions of subjective cultural re-elaboration of the museum heritage together with the sphere of participation with other users, aiming at building up a fruitful confrontation, an exchange of different points of view to find out similarities and differences in reality, and augmented reality perception. The museum educator must be capable of evaluating resources, instruments, and competencies present within an educational situation, together with the needs of the single users, to match such needs with possible solutions and the digital applications of

AR. Furthermore, they must develop transversal competencies with regard to active listening, critical thinking, problem-solving, communication strategies, negotiations and mediation, team working, the management of situations of individual and collaborative learning, the spirit of initiative and entrepreneurship to transform their own ideas, and those of others into educational opportunities. Lastly, they will have to acquire the capacity to work in collaborative mode inside their own professional team, in order to design, manage, and evaluate situated didactic pathways, meaningful and personalized, motivating the users and valorizing their experiences and their cognitive and emotional development. The ultimate aim is to develop in the public the skills not tied to learning for its own sake but that allow one to move and act with awareness, competence, and creativity, using effectively the content, digital, and transversal knowledge learned by recombining it and producing new acts for the resolving of "situational" problems, in the individual and group context also, thanks to the contribution of the others in a constructivist and co-constructivist vision of learning [164] and in a dimension of lifelong learning.

(c) *Learning by Discovery*

Learning by discovery involves the cognitive domains of *insight* and *invention*, together with digital competencies of *content creation*, conceived as possibilities of exploring adventure educational dimension of adventure both on an individual and social level through AR, where finding out new cultural paths becomes the purpose, always open, and never definable beforehand. The museum educator must, supported by the ICT teams, develop *augmented* digital contents, integrating them and re-elaborating them so as to make them personalized and developed around the needs – also special ones – of the different users. The aim is to valorize the personal history of each user stimulating and giving space to the creative thinking and the emotional processes, proposing an "immersion" in the art work, analyzing it as a whole or dwelling on a detail, decomposing and recomposing it several times over, making it one's own, and reinventing it by choosing completely novel pathways for themselves, thereby formulating new interpretations of the real. The user can, based on personal suggestions produced by the AR, create physical and digital artefacts in different formats and using materials, to express one's own ideas, opinions, and emotions in an original way.

Supplementary Video Material *Examples of the applications described in this chapter can be seen in the following video(s):* https://www.youtube.com/watch?v=uuWdGb5yBZw&t=4s. This video, titled "aRtelier. Educational Service and Video Research", edited by Roberto Farnè and Laura Corazza and directed by Enrico Masi, shows the aRtelier, an Educational Service of the District "S.Vitale" (Bologna, Italy). It consists of three laboratories for children 1 to 6 years. The children esperience the pleasure and sensations of color, objects, sound: Colorì (color and light laboratory), Dado (laboratory to build with recicled materials), Sonido (sound and silence laboratory).

Acknowledgment The contributions of this chapter are as follows: **1. Basic concepts** of Chiara Panciroli and Alessandro Luigini (1.1 Alessandro Luigini; 1.2–1.3. Chiara Panciroli); **2. augmented reality for meeting and making art** of Anita Macauda; **3. digital environments and augmented reality** of Chiara Panciroli and Anita Macauda (3.1. Chiara Panciroli; 3.2. Anita Macauda); **4. the language of images and augmented reality** of Laura Corazza; and **5. digital competences and augmented reality** of Manuela Fabbri and Veronica Russo (5.1. Veronica Russo; 5.2. Manuela Fabbri).

References

1. Krueger, M.W.: Artificial Reality. Addison-Wesley, New York (1983)
2. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **12**(12), 1321–1329 (1994)
3. Lanier, J.: Virtual reality. The promise of the future. Interact. Learn. Int. **8**(4), 275–279 (1992)
4. Rivoltella, P.C.: La previsione. Neuroscienze, apprendimento, didattica. La Scuola, Brescia (2014)
5. Rossi, P.G.: Tecnologia e costruzione di mondi: post-costruttivismo, linguaggi e ambienti di apprendimento. Armando, Rome (2010)
6. MacGregor, N.: A History of the World in 100 Objects. Penguin, London (2012)
7. Bortolotti, A., Calidoni, M., Mascheroni, S., Mattozzi, I.: Per l'educazione al patrimonio culturale. 22 tesi. FrancoAngeli, Milan (2018)
8. Jenkins, H.: Confronting the Challenges of Participatory Culture: Media Education for the 21st Century. MIT Press, Cambridge (2009)
9. Kasapoğlu-Akyol, P.: Somut olmayan kültürel mirasın örgün eğitime uygulanması: ağaraştırması (webquest) örneği. Millî Folklor. **28**, 125–146 (2016)
10. Ashley, L.: Encountering challenges in teacher education: developing culturally pluralist pedagogy when teaching dance from contextual perspectives in New Zealand. Res. Dance Educ. **15**(3), 254–270 (2014)
11. Mabingo, A.: Decolonising dance pedagogy: application of pedagogies of Ugandan traditional dances in formal dance education. J. Dance Educ. **15**(4), 131–141 (2015)
12. Morunga, T.: Classroom Korero: An Exploration of Relationships and the Delivery of Maori Artsin the Mainstream Dance Class. Doctoral Dissertation, University of Auckland (2013)
13. Craft, A., Jeffrey, B., Leibling, M. (eds.): Creativity in Education. Continuum, London (2001)
14. Eisner, E.: The Arts and the Creation of Mind. Yale University Press, New Haven (2002)
15. Robinson, K.: The Arts in Schools: Principles, Practice and Provision. Calouste Gulbenkian Foundation, London (1982)
16. Rowe, N., Martin, R., Buck, R., Anttila, E.: Researching dance education post-2016: the global implications of Brexit and trump on dance education. Res. Dance Educ. **19**(1), 91–109 (2018)

17. Buck, R.: Stepping back to step forward: reflections on future directions for dance education. In: Dance Education around the World: Perspectives on Dance, Young People and Change, pp. 161–167. Routledge, New York (2015)
18. Taggart, G., Whitby, K., Sharp, C.: Curriculum and Progression in the Arts: an International Study. Final Report. Qualifications and Curriculum Authority, London (2004)
19. Bamford, A.: The Wow Factor: Global Research Compendium on the Impact of the Arts in Education. Waxmann Verlag, Berlin (2006)
20. Panciroli, C.: Le professionalità educative tra scuola e musei: esperienze e metodi nell'arte. Guerini, Milan (2016)
21. Rivoltella, P.C.: Oltre il virtuale: la nostra è una 'realtà aumentata'. Vita e Pensiero. **5**, 102–108 (2010)
22. Billinghurst, M., Dünser, A.: Augmented reality in the classroom. Computer. **45**(7), 56–63 (2012)
23. Rossi, P.G.: Realtà aumentata e mediazione didattica. In: Persico, D., Midoro, V. (eds.) Pedagogia nell'era digitale, pp. 73–76. Edizioni Menabò, Ortona (2013)
24. Ranieri, M., Pieri, M.: Mobile learning. Dimensioni teoriche, modelli didattici, scenari applicativi. Unicopli, Milan (2014)
25. Diegmann, P., Schmidt-Kraepelin, M., Van Den Eynden, S., Basten, D.: Benefits of augmented reality in educational environments. A systematic literature review. In: Thomas, O., Teuteberg, F. (eds.) Internationale Tagung Wirtschaftsinformatik, pp. 1542–1556. Universität Osnabrück (2015)
26. Hills-Duty, R.: Technology. In Education-The Effect of VR and AR. VR/Focus (2017). Retrieved from https://www.vrfocus.com/2017/08/technology-in-education-the-effect-of-vr-and-ar/
27. Bacca, J., Baldiris, S., Fabregat, R., Graf, S.: Kinshuk: augmented reality trends in education: a systematic review of research and applications. J. Educ. Technol. Soc. **17**(4), 133–149 (2014)
28. Bower, M.: Augmented reality in education – cases, places, and potentials. In: 2013 IEEE 63rd Annual Conference International Council for Educational Media (ICEM) (2014). Retrieved from https://quality4digitallearning.org/wpcontent/uploads/2016/02/ar_overview_bower2014.pdf
29. Brown, P.: How to transform your classroom with augmented reality. EdSurge News (2015). Retrieved from: https://www.edsurge.com/news/2015-11-02-how-to-transform-your-classroom-with-augmented-reality
30. Chen, P., Liu, X., Cheng, W., Huang, R.: A review of using augmented reality in education from 2011 to 2016. In: Popescu, E., et al. (eds.) Innovations in Smart Learning. Springer Science+Business Media, Singapore (2017)
31. Dunleavy, M., Dede, C.: Augmented reality teaching and learning. In: Spector, J.M., Merrill, M.D., Elen, J., Bishop, M.J. (eds.) The Handbook of Research on Educational Communications and Technology, pp. 735–745. Springer, New York (2014)
32. Echeverría, A., Gil, F., Nussbaum, M.: Classroom Augmented Reality Games: a model for the creation of immersive collaborative games in the classroom (2016). Retrieved from https://pdfs.semanticscholar.org/1413/85717d0c961f79a136a9c8b3388b516104a6.pdf
33. FitzGerald, E., Ferguson, R., Adams, A., Gaved, M., Mor, Y., Rhodri, T.: Augmented reality and mobile learning: the state of the art. Int. J. Mobile Blended Learn. **5**(4), 43–58 (2013)
34. Bonaiuti, G., Calvani, A., Ranieri, M.: Fondamenti di didattica. Teoria e prassi dei dispositivi formativi. Carocci, Rome (2016)
35. Petrucco, C., Agostini, D.: Teaching cultural heritage using mobile augmented reality. J. E-Learn. Knowl. Soc. **12**(3), 115–128 (2016)
36. Salmi, H., Thuneberg, H., Vainikainen, M.P.: Making the invisible observable by augmented reality in informal science education context. Int. J. Sci. Educ. (2016). Retrieved from http://www.tandfonline.com/doi/pdf/10.1080/21548455.2016.1254358?needAccess=true&
37. Grant, M.M., Tamim, S., Sweeney, J.P., Ferguson, F.K.: Teaching and learning with mobile computing devices: case study in K-12 classrooms. TechTrends. **59**(4), 32–45 (2015)
38. Johnson, L., Smith, R., Willis, H., Levine, A., Haywood, K.: The 2011 Horizon Report. The New Media Consortium, Austin (2011). Retrieved from http://www.nmc.org/sites/default/files/pubs/1316814265/2011-Horizon-Report(2).pdf
39. Damiano, E.: La mediazione didattica. Per una teoria dell'insegnamento. FrancoAngeli, Milan (2013)
40. Clark, R.C.: Four architectures of instruction. Perform. Improv. **39**(10), 31–77 (2000)
41. Calvani, A. (ed.): Principi di comunicazione visiva e multimediale. Fare didattica con le immagini. Carocci, Rome (2011)
42. Di Serio, Á., Ibáñez, B.M., Kloos, C.D.: Impact of an augmented reality system on students' motivation for a visual art course. Comput. Educ. **68**, 586–596 (2013)
43. Gardner, H.: Frames of Mind: The Theory of Multiple Intelligences. Basic Books, New York (1983)
44. Smith, D.: Augmented Reality in Art Education. The University of Arizona Libraries (2017). Retrieved from http://arizona.openrepository.com/arizona/handle/10150/621860
45. Macauda, A.: Arte e innovazione tecnologica per una didattica immersiva. FrancoAngeli, Milan (2018)
46. Lichty, P.: Making inside the augment: augmented reality and art/design education. In: Augmented Reality in Education. A New Technology for Teaching and Learning. Springer, Cham (2020)
47. CoE: Raccomandazione del Parlamento europeo e del Consiglio del 18 dicembre 2006 relativa a competenze chiave per l'apprendimento permanente (2006). Retrieved from http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32006H0962&from=IT
48. Lughi, G.: Spazi urbani, come il digitale ne cambia la cultura (2017). Retrieved from https://www.agendadigitale.eu/senza-categoria/citta-arte-tecnologie-come-cambia-la-cultura-degli-spazi-urbani-col-digitale/
49. Macauda, A.: Visual learning and education for augmented reality environments. IMG J. **2**(2), 180–199 (2020)
50. Panciroli, C., Macauda, A., Russo, V.: Educating about art by augmented reality: new didactic mediation perspectives at school and in museums. In: Proceedings of International and Interdisciplinary Conference IMMAGINI? Image and Imagination between Representation, Communication, Education and Psychology, pp. 1–11. MDPI, Basel (2018)
51. Bonacini, E.: La realtà aumentata e le app culturali in Italia: storie da un matrimonio in mobilità. Il Capitale Culturale. Studies on the Value of Cultural Heritage (2014)
52. Bonacini, E.: Nuove tecnologie per la fruizione e la valorizzazione del patrimonio culturale. Aracne Editrice, Rome (2011)
53. Simonetta, G. (ed.): Realtà aumentate. Esperienze, strategie e contenuti per l'Augmented Reality. Apogeo, Milan (2012)
54. Manovich, L.: The Language of New Media. MIT, Cambridge (2001)
55. Spallazzo, D., Spagnoli, A., Trocchianesi, R.: Il museo come organismo sensibile. Tecnologie, linguaggi, fruizione verso una trasformazione design-oriented (2010). Retrieved from https://designforculturalheritage.files.wordpress.com/2010/05/il_museo_come_organismo_sensibile_spallazzo1-spagnoli-trocchianesi.pdf
56. Martini, B. (ed.): Il museo sensibile: le tecnologie ICT al servizio della trasmissione della conoscenza. FrancoAngeli, Milan (2016)
57. Montani, P.: Tecnologie della sensibilità. Estetica e immaginazione interattiva. Raffaello Cortina, Milan (2014)
58. Detroit Institute of Arts to premiere Lumin, a mobile tour using augmented reality, Jan. 25 Developed by DIA in partnership with Google and mobile developer GuidiGO. Retrieved

from https://www.dia.org/about/press/news/detroit-institute-arts-premiere-lumin-mob
59. Graham, M., Zook, M., Boulton, A.: Augmented reality in urban places: contested content and the duplicity of code. Trans. Inst. Br. Geogr. **38**(3), 464–479 (2013)
60. Battini, C.: Augmented reality applied to the archaeological site. A case study in Palazzo Baldini, Florence. In: Proceeding of International Conference on Cultural Heritage and New Technologies. Wien (2014)
61. Ellenberger, K.: Virtual and augmented reality in public archaeology teaching. Adv. Archaeol. Pract. **5**(3), 305–309 (2017)
62. Chang, K.E., Chang, C.-T., Hou, H.-T., Sung, Y.-T.: Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. Comput. Educ. **71**, 185–197 (2014)
63. Hitti, N.: Olafur Eliasson creates augmented-reality cabinet of curiosities. De Zeen (2020). Retrieved from https://www.dezeen.com/2020/05/14/olafur-eliasson-augmented-reality-wunderkammer/
64. Kamp, J.: Olafur Eliasson released a new augmented reality art project (2020). Retrieved from https://www.artsy.net/news/artsy-editorial-olafur-eliasson-released-new-augmented-reality-aproject
65. Panciroli, C., Macauda, A.: Heritage education and augmented reality: what prospects. Ital. J. Educ. Res. **11**(20), 47–62 (2018)
66. Macauda, A.: Augmented reality environments for teaching innovation. REM. **10**, 17–25 (2018)
67. Liu, T.Y., Tan, T.-H., Chu, Y.-L.: Outdoor natural science learning with an RFID-supported immersive Upiquitous learning environment. Educ. Technol. Soc. **12**, 161–175 (2009)
68. Bidoia, C.: Erbario in movimento, la realtà aumentata. Bricks. SIe-L Soc. Ital. e-Learn. **1**, 1–27 (2016)
69. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. **6**(4), 355–385 (1997)
70. Herrington, K.S., Crompton, H.: Augmented learning with augmented reality. In: Churchill, D., et al. (eds.) Mobile Learning Design, pp. 97–108. Springer Science Business Media, Singapore (2016)
71. Kress, G.R.: Multimodality: a Social Semiotic Approach to Contemporary Communication. Taylor & Francis, Routledge/London (2009)
72. Gutiérrez, K.D.: Developing a sociocritical literacy in the third space. Read. Res. Q. **43**(2), 148–164 (2008)
73. Flessner, R.: Revisiting reflection: utilizing third spaces in teacher education. Educ. Forum. **78**(3), 231–247 (2014)
74. Luigini, C., Panciroli, C. (eds.): Ambienti digitali per l'educazione all'arte e al patrimonio. FrancoAngeli, Milan (2018)
75. DeGeurin, M.: Internet Artists Invaded the MoMA with a Guerrilla Augmented Reality Exhibit (2018). Retrieved from https://www.vice.com/en/article/8xd3mg/moma-augmented-reality-exhibit-jackson-pollock-were-from-the-internet
76. Chow, A.R.: Snapchat and Jeff Koons collaborate on augmented reality art project. The New York Times (2017). Retrieved from https://www.nytimes.com/2017/10/03/arts/design/snapchat-and-jeff-koons-collaborate-on-augmented-reality-art-project.html
77. Panciroli, C., Macauda, A.: Ambienti virtuali e aumentati per valorizzare l'arte e il patrimonio. In: Lugini, A., Panciroli, C. (eds.) Ambienti digitali per l'educazione all'arte e al patrimonio, pp. 204–220. FrancoAngeli, Milan (2018)
78. Rossi, P.G.: Gli artefatti digitali e i processi di mediazione didattica, pp. 11–26. Tecnodid, Napoli (2016)
79. Panciroli, C., Macauda, A., Corazza, L.: Costruire relazioni di conoscenza attraverso artefatti digitali. In: Panciroli, C. (ed.) Animazione digitale per la didattica, pp. 37–49. FrancoAngeli, Milan (2020)
80. Panciroli, C., Macauda, A., Corazza, L.: Digital cultural heritage: from educational experience to the artefact in augmented reality. In: Luigini, A. (ed.) Proceeding of International and Interdisciplinary Conference on Digital Environments for Education, Arts and Heritage, pp. 345–354. Springer, Cham (2019)
81. Arduini, G.: La realtà aumentata e nuove prospettive educative. Educ. Sci. Soc. **3**, 209–216 (2012)
82. Berthoz, A.: La vicariance. Le cerveau créateur de mondes. Odile Jacob, Paris (2013)
83. Rossi, P.G., Rivoltella, P.C. (eds.): L'agire didattico. Manuale per l'insegnante. La Scuola, Brescia (2012)
84. Rossi, P.G.: Didattica enattiva. Complessità, teorie dell'azione, professionalità docente. FrancoAngeli, Milan (2011)
85. Sibilio, M.: La dimensione semplessa della didattica: traiettorie non lineari della ricerca pedagogica. In: Corsi, M. (ed.) La ricerca pedagogica in Italia. Tra innovazione e internazionalizzazione, pp. 191–200. PensaMultimedia, Lecce (2014)
86. Luigini, A.: Four dichotomies on digital environments between art, heritage and education: opening address. In: Proceedings of the 1st International and Interdisciplinary Conference on Digital Environments for Education, Arts and Heritage, pp. 3–12. Springer, Cham (2019)
87. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality. Foundations and trends. Human Comput. Interact. **8**(2–3), 73–272 (2015)
88. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. Multimed. Tools Appl. **51**, 341–377 (2011)
89. Furht, B. (ed.): Handbook of Augmented Reality. Springer, New York (2011)
90. Schmalstieg, D., Hollerer, T.: Augmented Reality: Principles and Practice. Addison-Wesley, Boston (2016)
91. Arcagni, S.: Visioni digitali. Video, web e nuove tecnologie. Einaudi, Torino (2016)
92. Arcagni, S.: L'occhio della macchina. Einaudi, Torino (2018)
93. Arcagni, S.: Immersi nel futuro: la realtà virtuale, nuova frontiera del cinema e della TV. University Press, Palermo (2020)
94. Corazza, L., E. Masi: Video-ricerca.: Una metodologia di esplorazione. In: Garcia, G., Iannucci, A., Vandini, M. (eds.) Il patrimonio culturale tra conoscenza, tutela e valorizzazione. Il caso della Piazzetta degli Ariani di Ravenna. BUP, Bologna (2015)
95. Eco, U.: I limiti dell'interpretazione. Bompiani, Sonzogno (1990)
96. Foucault, M.: What is an author? In: Rabinow, P. (ed.) The Foucault reader, pp. 101–120. Pantheon Books, New York (1984)
97. Goodman, N.: Languages of Art: An Approach to a Theory of Symbols. Bobbs-Merrill, Indianapolis/New York (1968)
98. Livingston, P.: Art and Intentions: A Philosophical Study. Oxford University Press, Oxford (2005)
99. Meskin, A.: Authorship. In: The Routledge Companion to Philosophy and Film, pp. 12–28. Routledge, New York (2009)
100. Perkins, V.F.: Film as Film: Understanding and Judging Movies. De Capo Press, London (1993)
101. Shams, L., Kim, R.: Crossmodal influences on visual perception. Phys. Life Rev. **7**(3), 269–284 (2010)
102. Lumbelli, L.: Il ruolo della percezione visiva nell'apprendimento con animazioni. Form@re. Open J. formazione in rete. **12**(80), 21–26 (2012)
103. Cicalò, E.: Intelligenza grafica. Aracne, Rome (2016)
104. Panciroli, C., Corazza, L., Macauda, A.: Visual-graphic learning. In: Proceedings of the 2nd International and Interdisciplinary Conference on Image and Imagination, pp. 49–62. Springer, Cham (2020)
105. Arnheim, R.: Il pensiero visivo. La percezione visiva come attività conoscitiva. Einaudi, Torino (1974)
106. Arnheim, R.: Arte e percezione visiva. Feltrinelli, Milan (1984)
107. Floridi, L.: La logica e il pensiero visivo. Iride. **24**(2), 343–358 (1998)

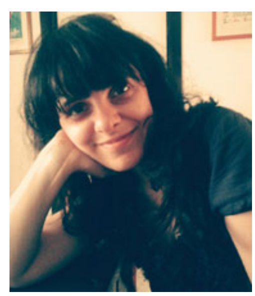
108. Robertson, I.: Intelligenza visiva. Il sesto senso che abbiamo dimenticato. Rizzoli, Milan (2003)
109. Fiorentino, E.: Il vero, il bene e il bello: le immagini come occasione di apprendimento significativo. In: Ulivieri, S., Binanti, L., Colazzo, S., Piccinno, M. (eds.) Scuola Democrazia Educazione. Formare ad una nuova società della conoscenza e della solidarietà. PensaMultimedia, Lecce (2018)
110. Clark, R.C., Lyons, C.: Graphics for Learning: Proven Guidelines for Planning, Designing, and Evaluating Visuals in Training Materials. Pfeiffer Wiley, San Francisco (2010)
111. Rodari, G.: Grammatica della fantasia. Introduzione all'arte di inventare storie. Einaudi, San Dorligo della Valle (1997)
112. Farnè, R.: Diletto e giovamento. Le immagini e l'educazione. Utet, Torino (2006)
113. Corazza, L.: Possibili scenari per un'educazione all'immagine. In: Luigini, A., Panciroli, C. (eds.) Ambienti digitali per l'educazione all'arte e al patrimonio. FrancoAngeli, Milan (2018)
114. Corazza, L.: Apprendere con i video digitali. Per una formazione online aperta a tutti. FrancoAngeli, Milan (2017)
115. Rivoltella, P.C.: Costruttivismo e pragmatica della comunicazione online. Erickson, Trento (2003)
116. Panciroli, C.: Ecosistemi digitali. In: Corazza, L. (ed.) Apprendere con i video digitali, pp. 13–30. FrancoAngeli, Milan (2017)
117. Buckingham, D.: Media Education. Alfabetizzazione, apprendimento e cultura contemporanea. Erickson, Trento (2006)
118. Buckingham, D.: Defining digital literacy. What do young people need to know about digital media? Nord. J. Digit. Lit. **2015**, 21–34 (2015)
119. Hobbs, R., Jensen, A.: The past, present and future of media literacy education. J. Media Lit. Educ. **1**(1), 1–11 (2009)
120. Rivoltella, P.C.: Media education. Modelli, esperienze, profilo disciplinare. Carocci, Rome (2001)
121. Rivoltella, P.C.: Digital Literacy: Tools and Methodologies for Information Society. IRM Press, Hershey (2008)
122. Dewey, J.: Democracy and Education. Macmillan, New York (1916)
123. Dewey, J.: Art as Experience. Minton, Balch & Company, New York (1934)
124. Dewey, J.: Il mio credo pedagogico. Antologia di scritti sull'educazione. La Nuova Italia, Florence (1954)
125. Carroll, N.: La filosofia del cinema. Dino Audino, Rome (2011)
126. Cabrera, J.: Da Aristotele a Spielberg. Capire la filosofia attraverso i film. Mondadori, Milan (2000)
127. Casetti, F.: L'occhio del Novecento. Cinema, esperienza, modernità. Bompiani, Milan (2005)
128. Casetti, F.: La Galassia Lumière. Sette parole chiave per il cinema che viene. Bompiani, Milan (2015)
129. Godard, J.L.: Il cinema è il cinema. Garzanti, Milan (1971)
130. Pasolini, P.P.: Il cinema in forma di poesia. Cinemazero, Pordenone (1979)
131. Pasolini, P.P.: Empirismo eretico. Garzanti, Milan (1972)
132. Manzoli, G.: Voce e silenzio nel cinema di Pier Paolo Pasolini. Pendragon, Bologna (2001)
133. Munari, B.: Design e comunicazione visiva. Contributo a una metodologia didattica. Laterza, Rome/Bari (1968)
134. Munari, B.: Artista e Designer. Laterza, Rome/Bari (1971)
135. International Council of Museums: Museums, museum professionals and COVID-19 (2020). Retrieved from https://icom.museum/wp-content/uploads/2020/05/report-museums-and-covid-19.pdf
136. Network of European Museum Organizations: Museums during COVID-19 (2020). Retrieved from https://www.ne-mo.org/advocacy/our-advocacy-work/museums-during-covid-19.html
137. UNESCO: Launch of UNESCO Report on Museums around the World in the Face of COVID-19. (2020). Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000373530_fre
138. Ballina, F.J., Valdes, L., Del Valle, E.: The phygital experience in the smart tourism destination. Int. J. Tour. Cities. **5**(4), 656–671 (2019)
139. Giannini, T., Bowen, J.P.: Transforming education for museum professionals in the digital age. In: Museums and Digital Culture, pp. 457–480. Springer, Cham (2019)
140. Silvaggi, A., Pesce, F.: Job profiles for museums in the digital era: research conducted in Portugal, Italy and Greece within the Mu. SA project. J. Cult. Manag. Policy. **8**(1), 56–69 (2018)
141. Antara, N., Shuvro, S.: The impact of Covid-19 on the museums and the way forward for resilience. Uluslararası Müze Eğitimi Dergisi. **2**(1), 54–61 (2020)
142. Hadley, W.: Covid-19 impact museum sector research findings (2020). Retrieved from https://www.artfund.org/assets/downloads/art-fund-covid19-research-report-final.pdf
143. Barnes, S.A., Kispeter, E., Eikhof, D.R., Parry, R.: Mapping the Museum Digital Skills Ecosystem: Phase One Report. University of Leicester, Leicester (2018)
144. Panciroli, C., Fabbri, M., Russo, V.: The professionalism of the museum educator between heritage competences and digital competences. In: International and Interdisciplinary Conference on Digital Environments for Education, Arts and Heritage, pp. 496–506. Springer, Cham (2018)
145. Wojciechowski, R., Walczak, K., White, M., Cellary, W.: Building virtual and augmented reality museum exhibitions. In: Proceedings of the International Conference on 3D Web Technology, pp. 135–144. Association for Computing Machinery, New York (2004)
146. Vlahakis, V., Ioannidis, M., Karigiannis, J., Tsotros, M., Gounaris, M., Stricker, D., … Almeida, L.: Archeoguide: an augmented reality guide for archaeological sites, IEEE Comput. Graph. Appl. **22**(5), 52–60 (2002)
147. Vincent-Lancrin, S., Van der Vlies, R.: Trustworthy Artificial Intelligence (AI) in Education: Promises and Challenges. OECD Education Working Papers. OECD (2020)
148. Gümbür, Y., Avaroğullari, M.: The effect of using augmented reality applications on social studies education. Araştırma ve deneyim dergisi. **5**(2), 72–87 (2020)
149. Mu, S.A.: Museum Sector Alliance (2016–2020). Retrieved from http://www.project-musa.eu/it/
150. Homem, P.M., Remelgado, P., Medina, S.: Museum professionals in the digital era: agents of change and innovation (2017). Retrieved from https://repositorio-aberto.up.pt/bitstream/10216/108860/2/231341.pdf
151. Borotis, S., Polymeropoulou, P., Pierrakeas, C., Kameas, A.: Training Handbook for Digital Interactive Experience Developer, Mu.SA – Museum Sector Alliance Project. Erasmus+, European Commission (2020). Retrieved from http://www.project-musa.eu/wp-content/uploads/2020/10/R3.3-Digital-Interactive-Experience-Developer_Handbook_EN.pdf
152. Panciroli, C. (ed.): Formare al Patrimonio nella scuola e nei musei. QuiEdit, Verona/Bolzano (2015)
153. Kaufmann, H.: Collaborative Augmented Reality in Education, Institute of Software Technology and Interactive Systems. University of Technology, Vienna (2003)
154. Kristinsdòttir, A.: Toward sustainable museum education practices: confronting challenges and uncertainties. Mus. Manag. Curatorsh. **32**(5), 424–439 (2017)
155. Munley, M.E., Roberts, R.: Are museum educators still necessary? J. Mus. Educ. **31**(1), 29–39 (2006)
156. European Council: Council Recommendation of 22 May 2018 on key competences for lifelong learning, Publications Office of the European Union, Luxembourg (2018). Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0604(01)&from=EN

157. Floridi, L.: The Fourth Revolution. How the Infosphere Is Reshaping Human Reality. Oxford University Press, Oxford (2014)
158. Mishra, P., Koehler, M.: Technological pedagogical content knowledge: a framework for teacher knowledge. Teach. Coll. Rec. **108**, 1017–1054 (2006)
159. Di Blas, N., Paolini, P., Sawaya, S., Mishra, P.: Distributed TPACK: going beyond knowledge in the head. In: Searson, M., Ochoa, M. (eds.) Proceedings of the Conference SITE-Society for Information Technology & Teacher Education International Conference, pp. 2464–2472. United States: Association for the Advancement of Computing in Education (AACE), Jacksonville (2014)
160. Di Blas, N., Fabbri, M., Ferrari, L.: The TPACK model and teacher training in digital skills. Italian legislation and pedagogical implications. Ital. J. Educ. Technol. **26**(1), 24–38 (2018)
161. Carretero, G.S., Vuorikari, R., Punie, Y.: DigComp 2.1. The Digital Competence Framework for Citizens with Eight Proficiency Levels and Examples of Use. Publications Office of the European Union, Luxembourg (2017). Retrieved from https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/digcomp-21-digital-competence-framework-citizens-eight-proficiency-levels-and-examples-use
162. Punie, Y., Redecker, C.: European Framework for the Digital Competence of Educators. Publications Office of the European Union, Luxembourg (2017). Retrieved from https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/european-framework-digital-competence-educators-digcompedu
163. Guerra, L. (ed.): Tecnologie dell'educazione e innovazione didattica. Junior, Parma (2010)
164. Scardamalia, M., Bereiter, C.: Knowledge building environments: extending the limits of the possible in education and knowledge work. In: DiStefano, A., Rudestam, K.E., Silverman, R. (eds.) Encyclopedia of Distributed Learning, pp. 269–272. SAGE, Thousand Oaks (2003)

Chiara Panciroli is full professor at the Department of Education of the University of Bologna. She teaches General didactics and Technologies of Education. She is referent of the scientific unit "AI and Education" of Alma Mater Research Institute for Human-Centered Artificial Intelligence. Her main research topics concern Innovative teaching in digital environments, Art education, Digital cultural heritage and Image recognition.

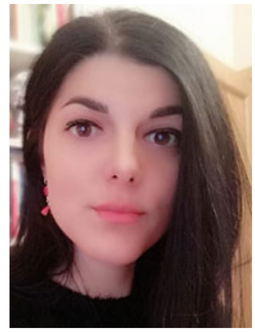
Manuela Fabbri is Associate Professor in the Department of Education Studies, University of Bologna. She's member of the scientific committee of the scientific unit "AI and Education" of Alma Mater Research Institute for Human-Centered Artificial Intelligence. Her research interests revolve around educational innovation and ICT, inclusive learning environments, socio-constructivist learning with a special emphasis on collaborative knowledge building community.

Alessandro Luigini, architect since 2003, in 2007 received his Ph.D. in "Representation and Survey of Architecture and the Environment". Since 2015 he is an Assistant Professor at the Faculty of Education of Brixen-Bressanone and since 2020 is Associate professor. He is editor-in-chief of IMG journal and head of VAR.HEE.LAB, laboratory for virtual and augmented reality for art and heritage education. He works on digital heritage, heritage education, digital representation and visual studies.



Anita Macauda, PhD in Art History. She is assistant professor at the Department of Education Studies, University of Bologna. Her main studies and research concern general teaching, museum teaching and educational technologies, with particular reference to digital environments for learning, augmented reality, robotics and artificial intelligence in educational contexts, visual intelligence, art education and digital culture heritage.

Laura Corazza, PhD in Pedagogy (University of Bologna). Her thesis was entitled “Democratic Education and the Knowledge Society” (University of Bologna). Her research deals with new media literacy, video research, visual communication of science, active learning methods and technology. She is co-author of numerous research documentaries.



Veronica Russo received her Ph.D.in Pedagogical Sciences at the University of Bologna in 2020. She is assistant professor at the Department of Educational Sciences of the University of Bologna. She works on museum audiences, digital environments for learning and teaching, active learning methodologies in schools and museums.

Augmented Reality's Application in Education and Training

13

Stephen Marshall

Contents

13.1 Development of Augmented Reality 335
13.2 Defining Augmented Reality 337
13.3 Pedagogical Framing of Augmented Reality 338
13.3.1 Location 339
13.3.2 Task 341
13.3.3 Role 341
13.4 Limitations 342
13.4.1 Technical Issues 342
13.4.2 Location Detection 343
13.4.3 Usability 344
13.4.4 Pedagogy 345
13.4.5 Health and Safety 345
13.5 Future Directions for Educational Use of Augmented Reality 346
References 348

Abstract

Augmented or mediated reality tools superimpose or composite virtual objects into the real world, supplementing reality rather than replacing it. The educational benefits include assisting learners appreciating a real-world experience more fully and developing independent thinking, creativity, and critical analysis. The intelligence amplification potential of this technology has been recognized for many years, and the development of low-cost augmented reality hardware integrated into mobile devices now means that educational implementations are plausible and can be practically implemented in a wide range of contexts. This chapter surveys the conceptualization of augmented reality and provides a detailed review of its educational applications, analyzing the underlying theoretical foundations that describe the learning and cognitive impact of its use as well as the practical benefits for learners and educators.

Keywords

Augmented reality · Education · Pedagogy · Mediated reality · Mixed reality

13.1 Development of Augmented Reality

Augmented reality technologies are a natural fit to education. The idea that human capability can be developed and complemented by tools that provide information in a useful form is closely aligned to the pedagogical processes of learning, including the use of feedback to inform learner strategies and actions. Arguably cave paintings represent an early attempt to provide an augmented experience of the world to transfer knowledge on hunting, survival, and the place of man in the environment [1].

More specifically, the modern conception of a digitally augmented world has from its earliest implementations included the recognition that it can powerfully affect the process of training and learning in many contexts. Heilig's work in the early 1960s on the earliest augmented reality system the Sensorama [2], an "apparatus to simulate the senses of an individual to simulate an actual experience realistically," includes in the patent the need for solutions for training and learning in the context of the armed forces, industry, and formal education. Similarly, the educational value of augmented reality as a tool to extend the scope of learning was seen by Ivan Sutherland in his early work on human computer interfaces, when he noted that augmentation provided users with the "chance to get acquainted with concepts that cannot be achieved in the physical world" [3].

The value of augmentation as a tool for providing information in context to workers in industry saw considerable invest-

S. Marshall (✉)
Centre for Academic Development, Victoria University of Wellington, Wellington, New Zealand
e-mail: Stephen.Marshall@vuw.ac.nz

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_13

ment in the technology. Boeing engineer and researcher Tom Caudell is attributed as having coined the term augmented reality in the early 1990s [4] to describe the use of technology to support aircraft manufacture and maintenance: "This technology is used to 'augment' the visual field of the user with information necessary in the performance of the current task, and therefore we refer to the technology as 'augmented reality.'" Similar examples of augmented reality to support maintenance and product development are now found in a range of commercial contexts, including the automotive [5, 6] and maritime [7] industries, as well as for military training [8, 9].

A recent review [7] identifies an extensive list of industry contexts that can benefit from AR, including service, manufacturing, sales, design, operations, and training (Fig. 13.1). Service aspects include provision of manuals, support for inspections and audits, and remote consultation with experts. Manufacturing aspects include quality assurance, performance monitoring, and assembly process support. Sales and marketing aspects include demonstrations of products, improvements in retail space usage and experience, and various forms of augmented advertising. Design aspects include digital prototyping, collaborative engineering, and design interface improvements. Operational aspects include the use of augmented reality to provide head-up displays for users, digital interfaces to products, operations manuals, and virtual control systems. Training aspects include the ability to provide in context coaching as well as tailored or role-specific training. In addition, all of the other uses of augmented reality can be seen as potentially tools for training staff involved in those aspects of the industry.

In the consumer space, cost has long prevented the meaningful use of augmented reality at scale. An early application was in the luxury car space, with companies such as BMW using the windshield as a display to provide augmented reality overlays for drivers [10]. The development of the smartphone with high-performance processors complemented by good-quality cameras and displays and a range of sensors means that many users now have devices capable of augmenting their reality routinely and at reasonable cost.

Modern smartphone applications seamlessly integrate a range of features that are essentially augmented reality. These are used for navigation and content creation and to interact with the world. Examples of the latter include the ability to use augmented reality to collect measurements using virtual tape measures (Fig. 13.2), real-time translation of text into other languages [11], and the ability to get real-time information on stars and other objects in the night sky [12, 13]. More elaborate augmented reality systems combine features of location awareness, collaboration, and content access to create games, such as Ingress [14] and Pokemon Go [15], which are used by millions of people. Augmented reality tools are also used more seriously to support disabled people with a range of specialist augmentations [16–18].

This explosion in applications and the diversity of possible contexts have also seen the growth in tools for those developing augmented realities. More than 20 software environments and over a dozen hardware devices are in use currently [7, 19], illustrating the diversity of interest and

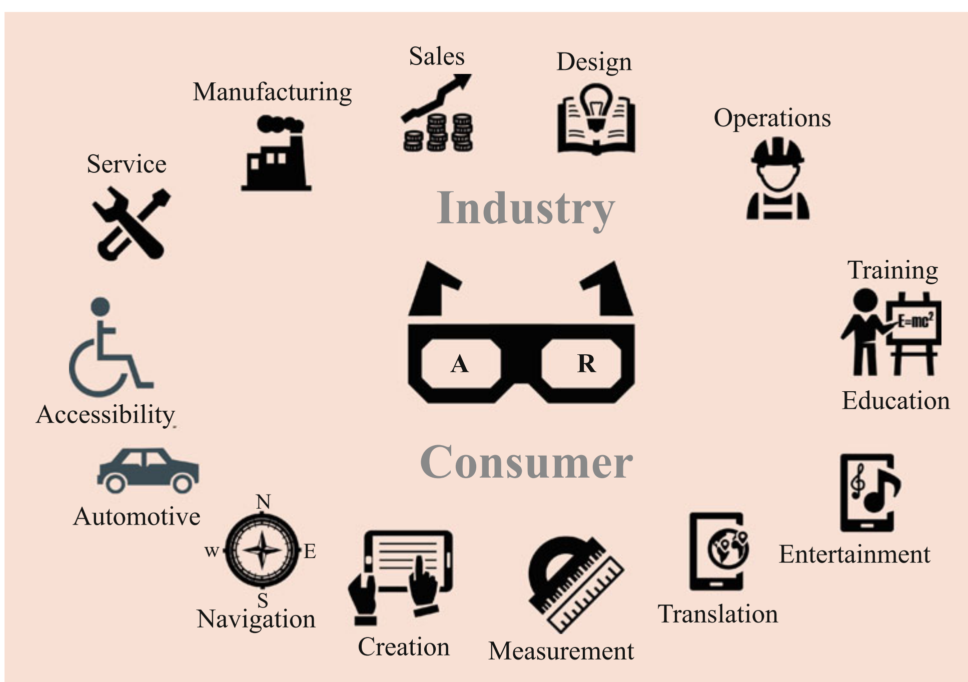


Fig. 13.1 Augmented reality industry and consumer contexts

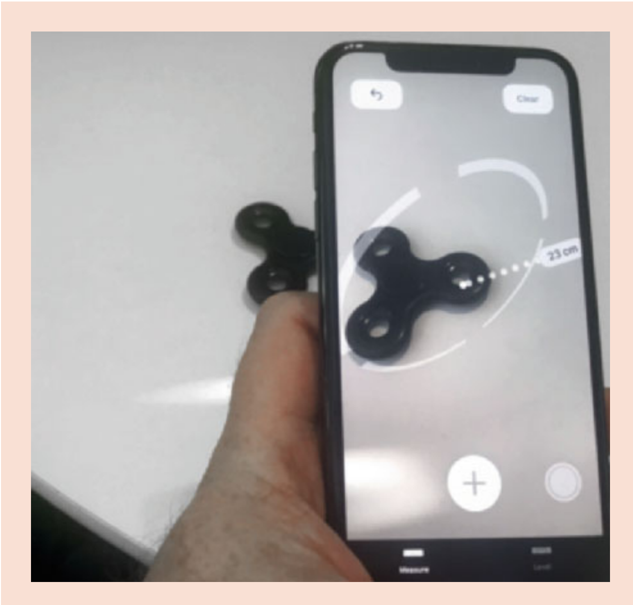


Fig. 13.2 Augmented reality measurement system

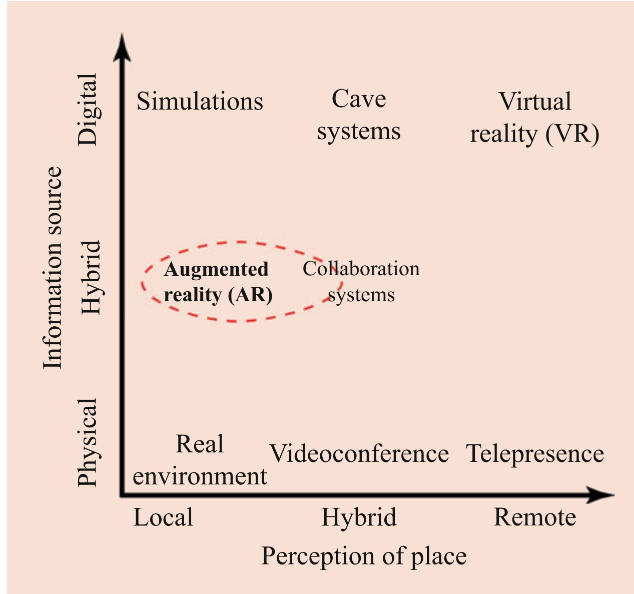


Fig. 13.3 Augmented reality placed within a broader mixed reality classification. (Inspired by [39, p. 1321] and [40])

activity in the augmented reality space but also demonstrating the complexity of the current generation of technology.

The educational use of augmented reality has also grown dramatically. Reviews of the field over the last decade [19–34] document the rapid growth in the field and the evolving status, trends, advantages, and challenges of AR in educational contexts.

The rest of this chapter surveys the conceptualization of augmented reality and provides a detailed review of its educational applications from this growing literature, analyzing the underlying theoretical foundations that describe the learning and cognitive impact of its use as well as the practical benefits for learners and educators.

13.2 Defining Augmented Reality

Caudell's original definition of the term augmented reality was using technology "to 'augment' the visual field of the user with the information necessary in the performance of the current task" [4, p. 660]. A more generalized definition reflecting the use of information in non-visual formats is offered by Craig: "Augmented Reality is a medium in which digital information is overlaid on the physical world that is in both spatial and temporal registration with the physical world and that is interactive in real time" [35, p. 36].

A variety of similar, but slightly different, definitions are used in the literature, including the superimposition or composition of virtual objects into the real world, supplementing reality rather than replacing it [36], "a situation in which a real-world context is dynamically overlaid with coherent location or context sensitive virtual information" [37, p. 205], and "augmenting natural feedback to the operator with simulated cues" [38, p. 283]. These later definitions expand from the initial conception to draw attention to the responsiveness of augmented reality to the context of the user, not only in terms of location but also in terms of the task, focus, roles, and characteristics of the specific user.

Augmented reality exists within a diverse range of digitally mediated experiences shaped by the information sources and the user's perception of place (Fig. 13.3). This shows the focus on the user's perception of the self in their local context hybridized with information sourced from outside of that immediate context. Increasingly there is a sense of hybridisation of place also occurring as collaboration tools are mixed into the augmentation, adding a sense that the user is also cognitively placed in a third space with their collaborators.

The distinctions being made in Fig. 13.3 between augmented reality and the more complete experience provided by virtual reality systems, as well as the related concepts of simulation spaces and telepresence, are increasingly blurred as tools become more capable and the supporting software and network systems develop in complexity. The distinguishing feature of augmented reality in this conception is that the user remains aware of the physical space they inhabit and engages with that space directly. Virtual information affects their actions in that space, and their real-world actions directly affect the virtual world, but their presence remains anchored in the physical reality.

Augmented reality is defined also by the technological infrastructure that provides the augmentation. Visual display

Fig. 13.4 Augmented reality headset with user indicating field of augmented information available for use

of information is the dominant channel used. Early systems depended on optical illusions [41], and the remnants of that approach are seen in the mirrors and lens used in modern devices. Modern augmented reality systems are essentially modified computer displays, typically head-mounted in the form of a helmet or increasingly as bulky glasses or through a small handheld smartphone or tablet. Irrespective of the device used for display, augmented reality is currently limited by the field of view. Dedicated augmented reality headsets, such as Microsoft HoloLens [42], struggle with significant optical constraints in the lens systems and the resolution of the displays, greatly affecting the amount of information that can be provided (Fig. 13.4).

Visual displays are not however the only hardware modes in use. Auditory information is commonly provided, including the use of stereo sound systems that provide direction and distance information as well as specific content [43, 44]. Touch can also augment reality using wearable devices, such as gloves, watches, and wristbands, or indirectly through vibrating devices, such as smartphones or the steering wheels of cars. Haptic channels are useful in contexts such as navigation where visual distractions may be safety issues or in social contexts where the intrusion of devices can be seen as impolite or disruptive.

The experience of augmented reality depends on the following: (1) the combination of virtual and real objects in a real setting, (2) people working interactively in real time, and (3) an alignment between real and virtual objects [45]. These depend on accurate information on the location and positioning of the user and the devices augmenting their reality in order to provide information. Augmentation systems currently use two main approaches for positioning information: visual analysis for key features, such as markers, and accurate device location and orienting [30].

Visual positioning systems have used a variety of approaches over the years (Fig. 13.5). Early systems used specially formatted content containing marker images [46], and there are many contemporary systems that continue to use this approach when generating educational content specifically for augmented reality. A specific approach is the use of visual barcodes in formats such as QR codes [47], which provide a standardized approach to communicating information such as URLs for accessing web content. More recent systems, including those provided by Apple for modern iPhones, use depth sensing cameras and image recognition software to dynamically respond to the environment in real time without the need for prepared markers [48].

Location-based positioning systems use real-time information on the user's location provided either directly as an absolute location provided by the use of GPS technologies, proximity to an infrared, RFID, or Bluetooth beacon [49, 50], or through radio triangulation and position derivation from network information [51]. This latter technique has the possibility of providing greater accuracy and also position information for users indoors or in locations where GPS systems are unavailable or unreliable.

13.3 Pedagogical Framing of Augmented Reality

As noted in the first section, augmented reality has been adopted extensively in education and training settings. The number of papers reporting applications of the technology continues to grow year over year with more than 16,000 papers listed by Google Scholar since 2018 with a search for augmented reality and education as keywords. Reviews of the field over the last decade [21–34] show evidence of this work, much of which indicates a positive response to the use of the technology.

At the most basic level, the research shows that learners like the use of augmented reality materials and experiences [25, 52]. This is reflected in evidence of a positive effect on motivation and focus on learning [24, 25, 27, 32, 33, 53–57]. Unsurprisingly, motivated and focused learners using augmented reality have shown improvements in their content knowledge and conceptual understanding [24, 32, 33, 53, 55, 56, 58]. Also, studies have demonstrated a wide range of positive influences of student cognitive and skill outcomes beyond content knowledge, including the development of autonomy, independent thinking, creativity, and critical analysis [27, 59]. Affect outcomes have also been shown to be positively influenced with augmented reality, influencing learners' emotional attitudes to complex issues [58]. Despite this, much of the historical literature is

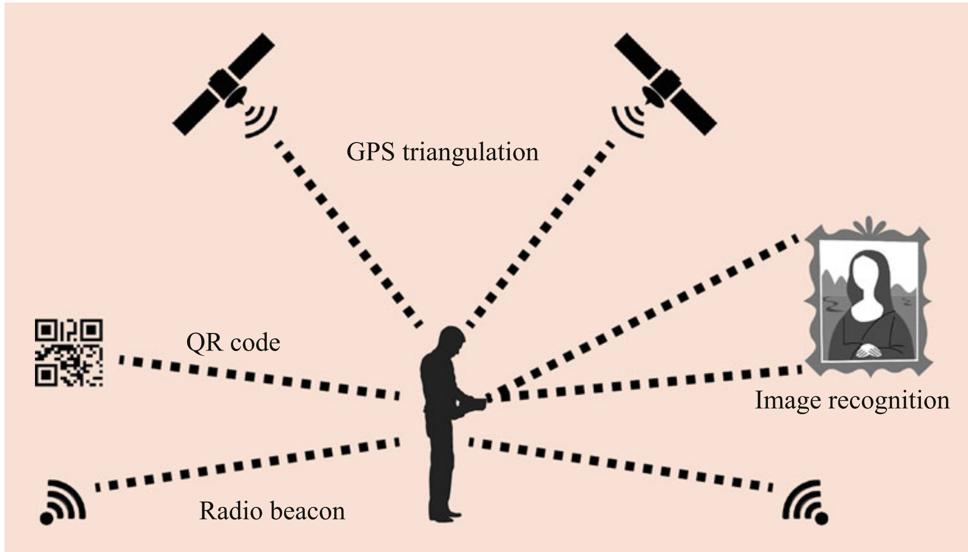


Fig. 13.5 Augmented reality location approaches

pedagogically limited. Many of the published case studies use traditional methods predominately to teach/learn, even in courses where the focus is on learning how to develop for augmented and virtual realities [29].

The impact on learning reported in the majority of studies is limited to aspects that directly result from the augmented reality activity rather than showing its influence on outcomes demonstrated in other parts of courses or programs [29]. A related observation is that most studies reported in the literature lack explicit consideration of the characteristics of the students and the learning goals of the course, important factors in designing effective instructional strategies and in aligning practice to pedagogical theory [28]. Consequently, even though use of augmented reality for knowledge acquisition dominates the published literature [27, 28, 60], empirical evidence of knowledge acquisition by learners is still relatively limited and lacking robust analysis [28].

As the field is maturing, there is clear evidence of the need to shift from a focus on purely technological aspects to practical issues affecting the implementation of the augmented reality systems to consider how its use affects learning from a pedagogical theory perspective. Two main perspectives can be seen in those studies which consider the theoretical positioning of their work, those that consider the content or resource impact of augmented reality and those that consider the learner activity impact.

The aspects of the definition of augmented reality that relate to the provision of information speak to the content models being implemented and the underlying pedagogical theories that speak to the role this plays in learning. Content orientation considers the role that augmentation plays in providing information in multiple ways, drawing on the theories of multiple media developed by Mayer [61–63]. This conception describes augmented reality as working through the use of a combination of words and pictures provided by AR annotations, including as auditory material reading out key text, and focuses the learner by providing information in context, spatially aligned and sequenced to guide learners through the experience in the most cognitively effective way [60]. Key to this conception is the recognition that information needs to be provided in complementary forms, i.e., audio annotation for a visual image, rather than in contested forms, i.e., detailed text descriptions as well as an audio narrative, in order to avoid overwhelming the user with too much complexity.

Alternatively, activity-oriented educational applications of augmented reality focus on what the learner does. These models are implicitly influenced by education theories, including constructivism (including social constructivism) [64], situated learning [65–67], and experiential learning [68–70]. In these conceptions, the focus is more on the context and on how this dynamically responds to learner activities, and the information supplied is shaped by the learner's activities and input more than by the teacher, who's role shifts that of a designer, facilitator, and enabler. Collaborative aspects of the augmented reality experience are also important under this conception, with the emphasis on the different roles learners play and how their interactions provide feedback and the development of a shared understanding towards common goals.

Educational affordances of different forms of reality under either of these conceptions can be organized according to the focus on the location of the learner, the task they are undertaking or the role they adopt [34].

13.3.1 Location

Location-defined affordances of augmented reality address the barriers to learning that are created by time, physical accessibility, safety, and ethics [71]. The emphasis in a location-defined model of augmented reality is on shifting the perception of the learner with regard to the location they are in by either adding information or by redefining the experience of that location in some more substantial way.

Two different approaches to location are apparent in augmented reality experiences, place dependent and place independent [72–74]. Place-dependent augmented reality is defined by specific locations which the learner needs to move to. These can be defined in absolute terms by the use of location information, such as GPS, or often for reasons of cost, convenience, or practicability (such as indoors) by the use of markers. This latter approach is particularly apparent in the life sciences where exploration experiences are the dominant educational design used [28].

Content can be supplied where the learner is, and in response to both their needs and the place that is represented, supporting both ubiquitous and experiential or situation learning [34, 75, 76]. This responsiveness and awareness of context supports the learner's sense of presence, immediacy, and immersion [77–79].

Presence is the illusion that arises from our engagement with an experience without conscious awareness of how that experience is generated or our engagement is sustained. It has been defined as the "illusion of nonmediation" [80], referring to the situation where the technological affordances used to provide interactive experiences become sufficiently imperceptible to the user that they essentially become "invisible" and the behavior and responses of the user are made without conscious consideration of the technology. A number of factors collaborate to create the sense of presence:

- The social richness of the experience arising from the ability to engage with others in collaborate activities and behaviors that complement language, such as facial expression, body language, and the feedback responses to these.
- The accuracy and realism of the representations provided by an experience, both in terms of social behaviors and in terms of the fidelity of the simulation in visual terms.
- The sense of place, or "place illusion" [81], which conveys the idea that the user is physically present in a location in time and space or "being there" [82]. This can involve being moved to another space, or the feeling that experiences are taking place directly in the space the user is present physically, either in response to the introduction of new affordances or the sense that other people have joined them to interact with.
- The sense of immersion, reflecting the experience of "flow" or complete engagement with an experience without external distractions. This is often conflated with the extent of technological replacement of senses but is strongly driven by cognitive and psychological factors relating to attention and enjoyment.
- The interpersonal experience and sense that interactions are generating social cues collaboratively between participants.
- The treatment of the medium as a fully contributing participant in an experience inconsistent with its physical reality, as seen when objects are personified and treated in ways analogous to interpersonal interactions with other humans.

An important element of presence is the active creation of this experience by the user's cognitive systems, which act to facilitate natural behaviors and can supplement features of the experience in order to further enable the user's activities building on even minimal cues to do so [83, 84].

A typical example of a place-dependent augmented reality experience is designed to allow students to explore an environment searching for salient features to engage with. Under the information provision model, these may involve the use of a predefined set of markers used to assist learning by identifying key features for in-depth engagement [60, 85]. Alternatively, an activity orientation model would have the students work collaboratively to construct their own learning, for example, using inquiry learning models [86, 87]. Data displayed in augmented reality can be collected by students in the real world and then integrated into the AR experience by students using environmental sensors placed in response to virtual markers [88].

A more open-ended model lets students explore environments and upload photos and commentary for other students to engage with as well [86]; here, the use of augmented reality helps focus learners by building on the situational interest that is stimulated by their environment [89]. The model also allows learners to take ownership in the use of the technology by framing their work by connections with communities local to specific places [90].

An important feature of augmented reality is the ability to overlay a different reality into the real world. This has been used for entertainment in software, such as Ingress, where real-world locations are augmented to have a different meaning and impact on a virtual landscape [14]. Educationally, a similar thing can be done to simulate otherwise impractical experiments such as virtual butterfly breeding [91] or taking learners to locations that are otherwise inaccessible by transposing features into another space that is more practically available for learning.

An example of this is that learners experience scaled down augmented reality field trips for geoscience education by moving around local campus environments and using augmented information to explore the geological features of the Grand Canyon, which are mapped to the local environment and communicated to learners through interactive activities and visual content [85].

13.3.2 Task

Task-oriented affordances of augmented reality focus the learner on a specific activity and provide information that is designed to facilitate their comprehension of the task, guide them through key processes or towards important aspects, and, in sophisticated examples, provide feedback on their work in real time [80, 92]. A particular strength of augmented reality is the ability to supplement media with supporting information. Simple versions provide content in interactive three-dimensional representations, so learners can engage with it from multiple perspectives and relationships [77, 93–95]. More complex augmentation can map or visualize invisible or abstract material in association with the physical world [76, 96–98].

The most basic content augmentation is seen in the development of "MagicBooks" [46], where specially formatted images in the books can be used to trigger an augmented overlay, such as an animation designed to support learning to read [99]. The literature now includes many examples of the augmentation of physics and mathematics course materials with simulations or dynamic visualizations [28, 100]. These are useful particularly to teach the spatial skills needed for working with complex three-dimensional spaces [101] and are also effective in helping weaker students learn spatial aspects of geometry [102].

Similarly, there are many examples in the literature showing that augmentation increases the learning impact of museums and exhibitions [60, 103, 104]. These can include interactive tasks that encourage learners to move around the space interacting with others and exploring a range of exhibits in a purposeful way that increases engagement, knowledge acquisition, and satisfaction [105]. As well as supporting improvements in knowledge acquisition generally, augmented reality use for exhibits has been shown to particularly benefit girls and weaker or less prepared learners [104].

Another significant benefit of content augmentation for task support is that it can be provided in multiple representations to support learners fluent with different languages [106] or with disabilities, for example, through the provision of text-to-speech features for visually impaired learners [107].

Task augmentation is a powerful tool for training. Here, information is provided in the context of a piece of equipment or process in a way that informs and guides the learner as they are familiarized with it. For example, learners being trained in the features of professional sewing equipment learned more in a shorter period of time using augmented reality. Augmented video instruction helped the students understand the complex spatial process of threading machines more effectively than traditional handouts [95].

More ambitiously, this can include augmentation to provide facilities that are not normally present as well. This can include the addition of new features, such as in laboratory spaces, in order to reduce the costs of providing and maintaining specialist equipment [108]. This can extend to entire spaces, such as the use of augmented reality to teach stock trading by simulating the multiple screens that would normally be available simultaneously to a trader without the expense of providing dedicated environments that replicate the professional space [109].

Augmented reality is particularly effective at mapping unobservable or invisible phenomena into the real world for educational purposes. The promise of this has been evident from the first forays into this space by Ivan Sutherland [3]:

> There is no reason why the objects displayed by a computer have to follow the ordinary rules of physical reality with which we are familiar. The kinaesthetic display might be used to simulate the motions of a negative mass. The user of one of today's visual displays can easily make solid objects transparent – he can 'see through matter!' Concepts which never before had any visual representation can be shown [. . . b]y working with such displays of mathematical phenomena we can learn to know them as well as we know our own natural world.

This can include simple concepts, such as the position of the sun or moon in the sky even indoors [110, 111], or more complex ones, such as the behavior of magnetic fields [112], kinetic energy and momentum [113], or the principles of electricity [114, 115]. A powerful example of this is the use of an augmented reality display in the form of a "magic mirror," which displays the reflection of a learner augmented by anatomical scans and information scaled to their physiognomy as a tool for teaching details of human anatomy and physiology with the learner acting as their own visual aid [116] (Fig. 13.6).

Task support can also be learner driven rather than dependent on content supplied by the teacher or collected by other devices. The augmentation of content has been used by learners to elaborate on architectural project work to present information in a variety of ways including 3D overlays and QR codes linking additional content such as commentaries and supporting analyses. The result was a significant increase in student motivation and academic achievement [117].

Finally, the task support provided by augmented reality can also be of value to educators, providing access to teaching notes while in class and as a tool for creating lesson materials from a first-person perspective [118]. Other uses being trialled include using augmented reality as a mechanism for interacting with students in classes through a virtual overlay [119].

13.3.3 Role

Role-focused affordances of augmented reality are driven primarily by the learning activities and the support for learners in adopting roles that enable their own learning, often in collaboration with other learners working cooperatively to create a shared context for learning [34, 77, 120]. Such collaboration enables engagement with others, including com-

Fig. 13.6 Magic mirror organ overlay

plete strangers, in support of learning [105], facilitates collaborative inquiry learning [87] and is generally found to improve a range of cognitive and skill learning outcomes [113].

13.4 Limitations

The extensive and growing augmented reality educational literature also describes a number of limitations with current applications. These include technical limitations with the hardware, software, and supporting infrastructure, usability issues including complexity and social acceptability, pedagogical issues including effective learning design, supporting infrastructure such as classroom design and the role of the teacher, and health and safety issues when using the technology.

13.4.1 Technical Issues

Technical issues are a frequent observation in the literature [20, 23, 27, 34, 121, 122]. Many of the issues noted in the older literature relate to the use of bespoke hardware or devices sold primarily for research and development use, but there are many significant problems arising from the limitations of the current technology that remain significant challenges for users of augmented reality experiences in educational contexts.

The immaturity of the research platforms being tested in many cases mean that it is not unreasonable that users complain of software being unreliable and thus crashing or failing to work correctly in normal use [112, 121, 123]. The limitations of current development frameworks and the different hardware environments also mean that users can experience issues when displaying certain types of media [123, 124].

These latter failures are almost certainly a consequence of the technical limitations of the hardware used to deploy augmented reality experiences with learners. Cost constraints mean that these devices often have to make compromises with system processing power and memory, which result in slow system performance [111, 125]. Many systems are compromised by the need to balance the size and quality of the learning media against the performance and cost of the available wireless networks and the storage capacity of the devices [124].

As well as preventing the display of certain media, system performance can generate issues with latency, or the delay between a change in state of either the physical or virtual aspects of the system, for example, the user moving their position or interacting with a dynamic control. Latency issues can badly affect usability and accentuate motion sickness issues [126, 127]. Many of these problems arise when the user moves their field of view more rapidly than the device is designed to support, meaning that content can jump, stutter or tear or fail to respond to user interactions when locations are not matched accurately [34, 128]. Item placement and movement needs to be done rapidly and to a high level of accuracy as the human visual system is able to detect very small inconsistencies, which generate unacceptably distracting tracking errors [129], visually disturbing movements or shaking of rendered content [54]. Such issues negatively impact on the user's sense of presence within an experience [130]. Identifying precisely where a user is looking in real time remains a significant challenge to be overcome [131, 132].

The inability to accurately locate objects for interaction arising from limited system performance is also a consequence of the technical limitations of the current generation of tools and frameworks. Detailed monitoring of people's bodies to determine their orientation, direction of view, and position of limbs, hands, and fingers in order to respond to gestures is challenging with current tools [133]. A major problem is that the information collected is typically incomplete and requires processing with tools, such as machine learning [134], in order to create a realistic model of position. The level of system processing needed currently for such tools means that a high-performance network is necessary, further limiting the usability and increasing the likelihood of failures due to disruptions in connectivity in environments outside of the laboratory.

Many studies report issues with the use of hardware developed by researchers when deployed outside of laboratories. Users report issues with hardware that is too complicated, unfamiliar or fragile for learners to use without the need for expensive additional specialist technical staff available to support the use of complex technologies [75, 120]. A major issue reported is the weight of the augmented reality hardware, typically headsets, which need to accommodate the needs of users with a wide variety of head sizes and ages [121, 123]. The biomechanical load imposed on users' necks needs to be very carefully evaluated to avoid fatigue or injury [135, 136]. Users also raise concerns about the unfashionable or distracting physical appearance of augmented reality equipment, with problems caused by the wearing of equipment particularly on the face and by the cables and other, often bulk, components needed to use the equipment [23].

Poor fitting of specialist augmented reality headsets can also exacerbate issues with the image quality [123]. Systems remain severely limited by the performance of the cameras under a range of lighting conditions [137] and by the resolution of the displays and optical systems used to present information to users [138]. Limitations with displays for augmented reality [131, 132] are fundamentally an issue of pixel density as systems attempt to present detailed information equivalent to several high-definition displays to each eye in order to achieve a wide field of view, addressing the limitations noted in Fig. 13.4. A particular issue for educational applications is that the limited resolution of displays can make text difficult to read and limits the amount of text content that can be shown at any time [123, 139].

As well as compromising the display of information, the display size and rendered area limitations make it difficult for system designers to create and sustain collaborative experiences with multiple users working directly and naturally with rendered content [140]. Finally, the limited fields of view make it harder to manage the users framing of key features, making it easy for them to miss information [137].

Sensory issues are also a problem for sound integration into augmented reality. Many systems have audio hardware that users found difficult to configure to meet their needs [123] and which is affected badly by noisy environments, particularly if the system is designed to support users recording their own audio, communicating or issuing commands by voice through microphones [75]. Achieving the normal directionality of audio requires significant investment in specialist recording and processing tools, and the area is in need of further work to reliably and efficiently reproduce the real-world experience of sounds coming from different locations [43].

The development of consumer augmented reality devices is helping address many of the practical issues, for user training and familiarization, in particular, of the diversity, complexity and fragmentation of the market, which reduces transferability and increases the cost of deploying these tools in practice [7, 19, 34, 141]. Issues of weight, portability, and performance that troubled early head-mounted augmented reality systems are increasingly being addressed through the use of consumer devices, such as smartphones and tablets, which are well accepted even by very young children in classroom tests [142]. The main benefit of the use of consumer devices is the confidence and familiarity users have with their devices, combined with the significant cost reduction that they represent when compared to specialist hardware.

There remain issues with the high costs of implementing and maintaining the software and content for augmented reality experiences [22, 143], which are not only driven by the technical issues of maintaining compatibility with the rapidly evolving software and hardware but also result from the current development model where experiences are typically developed as bespoke solutions to specific teaching needs and subjects by specialist developers rather than the teachers themselves [20, 22, 25].

Frameworks which enable non-specialists to author learning activities, such as providing training overlays for equipment [144] or for the creation of educational games [145], are starting to be developed, but these have yet to mature to the point of wider adoption. Currently, the state of the development tools and the associated hardware make the creation of new content expensive and highly complex [123].

13.4.2 Location Detection

Despite the progress in the development of consumer devices supporting augmented reality and their provision of location-responsive services, such as mapping and activity tracking, significant issues remain with the reliability and robustness of user location detection [23, 141, 143, 146], particularly indoors [147]. The simplest technique, using the GPS location reported by consumer mobile devices commonly available to students, has an error that can be several meters out, depending on location and the visibility of satellites [75], meaning that the choice of locations and objects has to be done carefully to avoid confusing overlaps that can interfere with the display of information or user interaction with experiences [148].

The inability to use GPS in many locations means that alternatives are needed (Fig. 13.5). Even when these are available, these internal locations also need high-quality maps that are easy for users to quickly comprehend in relation to their view of the environment and that can be mapped into the augmented reality experience easily as the environment is modified over time. Creating these can be very expensive given the level of accuracy needed [123], although cheaper equipment is becoming available in the form of the latest generation of smartphones with built-in LIDAR [149, 150].

The intersection of the virtual information and real-world constraints is an issue for applications that use physical waypoints to provide more accurate location information using Bluetooth or visual markers. These can be difficult to locate in a way that does not interfere with other activities and features of the context being explored [123], including disrupting the experience of other users of a space or causing offence [151]. An example of the type of system failure that poorly thought-out experiences can create is the way that poor GPS navigation systems create disruptive traffic patterns when they fail to account for the needs of the residents of different neighborhoods [152].

Location and position issues are important for many training uses of augmented reality requiring complex image processing and pattern recognition software to correctly identify specific features of equipment and annotate these correctly in the augmented overlay [123], itself often incorporating rendered material that needs to be generated, oriented, and scaled correctly in real time and then composited with the scene for the learner [7]. Increasingly, this involves offloading key processing tasks to cloud servers capable of supporting the load but at the price of increased demand on networking resources, particularly when scaling solutions to class-sized numbers of users [153].

Location inaccuracy also makes it difficult for users to engage with others through the interface in order to engage in collaborative learning experiences. Relative positions of different users may not be correctly represented or may be slow to update to match the real world [123].

13.4.3 Usability

The usability of augmented reality activities for both teachers and learners remains a challenge. Many studies report learner feedback that the technology is complex and difficult to engage with, in order to use it effectively [20, 27, 32, 78, 121, 143], and consequently requires that teachers and students be trained to use each experience, making it impractical to use these at scale [70, 111, 154]. The low usability means that many educational implementations are difficult to scale even when consumer hardware is used to overcome the expense of providing large numbers of devices to learners [20] and that, consequently, the time cost of using the technology can be excessive in proportion to the resulting learning outcomes [155].

The complexity of augmented reality for users starts with the most basic issues of identifying what can or should be done in the environment. This is a consequence of the ambiguity of visual elements as controls or markers for user interaction [123, 124]. Control usability is also compromised by the size of virtual controls and the need to position them so they are usable while avoiding obscuring important features that may not be statically positioned in relation to the user or users [123]. The wide variation in the age of users means that, even when touchscreens are used, there are problems with the ease of use of touchscreen controls designed for different hand and finger sizes [18]. An additional complication is the range of different devices, which means that users can perceive objects differently in terms of size and location [156].

The mechanism for engaging in augmented reality affects usability [157]. Handheld devices are typically used to provide a window through which augmented information is displayed; this can then be easily and directly manipulated through normal touch interactions familiar to the users. The alternative is the use of headsets, which then require some other form of control, such as gestures, which are more complex to implement and less reliable to use and detect [132]. Many applications suffer from the unfamiliarity of users with particular forms of interaction designed to replace more familiar tools such as keyboards with voice annotation or gestures when using a headset rather than a portable touchscreen [124, 131].

A more complicated usability problem is how systems provide guidance to the users regarding the various options and opportunities available to them. Many users report problems identifying what they can do with systems lacking clear directions on how to move between different activities [121] or modes [95]. This is particularly an issue when systems manage technical and cognitive limitations with information display and density by operating in different modes that require a conscious action by the user to shift between.

More generally the design of augmented reality applications is complicated by the absence of a common design language accepted by creators and users that establishes norms of meaning and behavior that enable the technology to become “invisible” [158] and consequently its use to be natural and efficient. The standard two-dimensional interface of icons, windows, mouse pointers, and keyboard entry, supplemented more recently by the use of direct touch controls, has yet to translate into a similar taxonomy of objects and meanings that can be standardized across different augmented reality applications.

A serious limitation of many current augmented reality systems for education is the lack of compliance with regulatory mandates to support learners with a range of physical and cognitive differences. Almost all augmented reality applications are not accessible and do not embody the requirements of universal design for learning [27, 159, 160]. The overwhelming majority of augmented reality educational applications are directed at visual augmentation of the environment and assume users have good vision. Many headsets are difficult to use with glasses, often being too tight, close to the eyes or humid resulting in fogging of displays. A recent review found that less than 3% of systems presented in the literature had considered the special needs of some groups of learners [27], consistent with earlier reviews noting similar limitations in most augmented reality systems [34, 161]. This is a complex space with many difficult challenges in designing new interfaces that meet the needs of users with different needs, such as touch augmentation for users with vision impairment [18].

13.4.4 Pedagogy

The problem of complexity noted in the Usability section is not just a technical or aesthetic issue but also results from a lack of pedagogical structure in the design and interfaces of augmented reality experiences [26, 74]. Effective interfaces help learners focus on the educational tasks and features and do not distract them or overwhelm them with information – the "cognitive overload" problem [28, 76, 162]. As one reviewer notes, "... with augmented reality, it will be very important for the developers to remember that AR aims at simplifying the user's life by enhancing, augmenting the user's senses, not interfering with them" [23, p. 371].

Scaffolding designed to align the affordances of augmented reality experiences with the learning objectives is an active area of research [163] with some evidence that designs that focus on cognitive structures rather than content may be more effective for learners [60, 164]. This is a complex space with any generalized solutions to the problem of creating virtual representations and activities capable of responding in real time to the learner's task progress and sense of flow yet to be identified [123, 165]. If anything, augmented reality can be so engrossing or interesting in itself that learners get distracted from key learning activities and the task at hand [75].

The pedagogical issues also include barriers and disruption that affect teachers and their work. Many systems make it very difficult for teachers to create content within the experience that align with established lesson plans and curricula [124] or to adapt materials to meet the needs of the local educational context or culture [148]. This results in an alienation of teachers who feel they have lost control over the content of their lessons [34]. As well as being difficult for teachers to create lessons with themselves, augmented reality technologies and activities can present practical issues for teachers needing to integrate their use into normal classrooms, complement them with activities undertaken in other modes and support learners with diverse needs [22, 24, 32, 33, 143]. Many of the examples in the literature continue to align augmented reality experiences with traditional models of assessment [29]. This may not only reflect a lack of commitment to a revised pedagogy but also likely reflects the need to support the entire group of learners when the technology may not be available or suitable for all of them.

A common finding of case studies of the educational use of augmented reality is that teachers express the need for training both in the basics of augmented reality and in the pedagogical models and applications of the technology [54, 88]. Teachers find the logistics of managing the equipment alone challenging. The additional demands and complexity often result in cognitive overload of the teachers attempting to also maintain their role and responsibilities while using the new activities [27, 76]. Consequently, it is not surprising that teacher resistance is reported as an issue in reviews of the literature [27, 34].

13.4.5 Health and Safety

All technologies face concerns over their use as they are adopted and used in new contexts by inexperienced users. Augmented reality is no different in raising a range of health and safety concerns, many of which reflect the intimacy of the technology and its placement as a mediating layer between normal human interactions.

As with other mixed reality tools, augmented reality experienced through headsets can generate significant issues with what is known variously as motion sickness, cybersickness, or simulator sickness [166]. Augmented reality, despite being less immersive than full virtual reality, appears to generate a similarly negative response in some users as any other form of mixed reality [167]. Studies found that 80% of participants experienced some form of simulator sickness. Thirteen percent of participants stopped before the experiment was over because of sickness. Nine percent of these (1.2% of overall) experienced a vomiting response. The strength of the adverse responses was related to duration of exposure [168]. Additionally, discomfort was not immediately alleviated after the experiment ended, with symptoms lasting several hours [168, 169] along with a range of other symptoms, including drowsiness, dizziness, headaches, and disorientation. Golding points out that "Given a sufficiently provocative stimulus nearly all people can be made motion sick" [170, p. 70]. These symptoms can be serious, and safety protocols are strongly urged for all research involving novice users of all mixed reality technologies including augmented reality.

The complex optical mechanisms used for augmentation of the visual field have resulted in disruption to eyesight. These are caused by issues in the visual display affecting the normal operation of the human vision system [55]. The ability of the human brain to adapt to modifications of our vision is well known, but a consequence of these is that adaptations distort vision when the modification or augmentation is removed, and this distortion may continue for a variable length of time [171].

On a practical note, there is evidence of learners losing track of their environment as they're too engrossed in the augmented reality experience [76] and putting themselves in potential harm as a result [172], and it is reported that 10% of users suffer accidents from collisions while using augmented reality. More seriously, use of augmented reality has seen users suffer serious harm or death [173].

Users also report fatigue [174], either through the level of engagement resulting in extended activity, strain from the visual complexity and focus shifting in augmented reality applications [175] or as a result of the equipment putting a strain on the body in various ways, particularly the neck and head for heavier headset devices [135, 136].

Augmented reality has also seen a significant social backlash as members of the community react negatively to others wearing devices on their faces. As well as a range of normal technology concerns relating to the ethics and privacy of their use [23], there is evidence that augmented reality tools can be socially disruptive if their affordances intrude into the perceptions of others [23]. The association of headsets with military applications, and the discomfort of having a camera pointed at them means that many people react nervously to equipment such as that shown in Fig. 13.6 [176, 177].

In extreme cases the visible use of augmented reality equipment can be perceived violently by some, especially in public settings [176–179], with the response to the Google Glass augmented reality technology creating the neologism "Glasshole" to describe users of the technology [178, 179]. In some cases, the violence of the response to the use of the technology has been such that serious health and safety concerns would need to be considered prior to using augmented reality for education in other than controlled environments.

13.5 Future Directions for Educational Use of Augmented Reality

Augmented reality is a powerful tool for learning as demonstrated in the overview provided in this chapter, but it has yet to transition into mainstream use. Despite the limitations noted above, the promise of augmented reality and the extent of the engagement in its development indicated above suggest that it will continue to grow in use despite the immediate concerns of its detractors.

In the short term, it is likely that augmented reality will be dominated by use of consumer mobile devices, such as smartphones and tablets; running software designed around basic activity templates, such as visual instruction manuals for equipment training; and location-based information delivery and gathering, particularly for inquiry-based exploration learning, complemented by content created by commercial publishers as part of the media supports for hybrid textbooks. Software such as Pokémon Go and Ingress also illustrate the potential for more integration of social media with augmented reality, generating collaborative environments capable of supporting social constructivist pedagogies in a wide variety of settings.

More ambitiously, there is the likely development of information appliances that work with the developing capabilities of digital assistants, such as Google's Home Hub and Amazon's Echo, and virtual assistants, such as Apple's Siri and Microsoft's Cortana, to create, in effect, a personal tutor attentive to and responsive to the educational needs of the user in real time [180]. Much as the electronic calculator impacted mathematics education by removing the requirement for laborious calculations, these new tools will change the experience of content acquisition and recall far beyond even the best current efforts of web search engines. The role of the teacher and the process of learning itself will ultimately experience significant disruption once the transmission of content becomes essentially trivial and no longer be the anchor point of formal education as it is today.

Figure 13.7 illustrates how such augmentation might be experienced by a learner, mixing a range of information channels, contextual information and tools into a visual interface that can augment the limitations of human cognition [181]. One problem immediately apparent is the "clutter" or complexity of the information being displayed. As this is user controlled, it is unlikely to generate information overload, but distraction is very much a problem [86], suggesting that such systems would benefit from augmentation of attention [182, 183]. Here that is shown through the use of the system generated alert in the middle of the field of view complemented by a transcript of recent content provided below and the flagging of the teacher in the center as being the person attention should be directed to.

Smart assistants would be monitoring the events in the environment, including the body language and sight lines of other people in order to identify events that require user attention. The assistant's tutoring function is implied by the list of key facts provided on the bottom left. This information would not simply be a keyword search and summary but would be informed by the artificial intelligence's model of the user's knowledge of the subjects being discussed in real time and their interests derived from previous activities. Other useful tools include the ability to identify other people and to integrate information from them appropriately, such as the Twitter trail shown on the right.

Most augmented reality experiences remain very focused on the needs of a single user. A review of the literature found that less than 2% of augmented reality experiments consid-

Fig. 13.7 Augmented reality information appliance concept demonstrating user-focused information streams and attention support

ered collaboration [184]. At a minimum, future systems such as that in Fig. 13.7 could allow for the use of virtual avatars in meetings integrated into the field of view, for example, the other participants might not be in the room physically despite appearing to be so. The first versions of this are already evident in the virtual meeting settings features of products, such as Microsoft's Teams and Zoom [185].

The interface in Fig. 13.7 is a mock-up, but there is very little in it that is not capable of being delivered today, albeit at a very high cost. Challenges still to be resolved include the capability of smart information assistants, which still require significant improvement before they can operate as comprehensively and rapidly as shown here. Also problematic are the field of view shown and the high resolution of the displayed information. Resolving both of these is only a matter of time. Increased network performance and ongoing software development suggest that the assistance function is very close to being usable. Improvements in the understanding of how the visual system operates and ongoing development of high-performance display technology suggest that these aspects are also close to being achieved.

Many educators are already concerned about the distraction caused by the use of various electronic devices by learners, and their sense that these are compromising education or redefining the classroom experience in negative ways [186–193]. In many respects these reflect the growing disconnect between the traditional classroom education processes and the needs and experience of learners familiar with modern information technologies and confident in their use of them [194].

An example of how augmented reality will disrupt this space further can be seen by considering the way that collaborative overlays could be used by groups of learners to modify their shared sense of the classroom with the awareness by the teacher or other learners that this is happening. Virtual overlays or back channels have already been implemented for websites [195] and conferences [196] and not always with positive outcomes [197]. Similar technology now exists to create augmented reality overlays that could be designed to be viewed only by some people creating concerns that they could be used for harassment of various types [198] or simply disrupt activities in spaces [151].

As with all new technologies, a wide range of social and ethical challenges have been identified that will require a process of engagement and renormalization by society as augmented reality becomes a mainstream tool [159, 199]. These include serious concerns about the impact of these tools on individual freedoms and inequality, as well as control over how personal identities are represented and used, and the possibility for deception or misrepresentation of reality [200–202]. This is a particular concern in educational settings, given the legal obligations to protect student information in many jurisdictions.

The literature has identified surveillance and privacy concerns as third parties collect data from AR [203] and potentially misrepresent identity or actions to cause harm [197] or manipulate behaviors, for example, through advertising or fraudulent scams [197]. More generally there are ethical concerns about the way that mixed reality experiences may normalize or promote harmful behaviors or views [204]. There is also the risk that super-realism, or the powerful illusion of presence created by some experiences, could cause issues with the blurring of the line between virtual content and the real world, triggering persistent psychological harm particularly to vulnerable populations, such as children and young adults [199].

Beyond these examples there are also serious concerns around the further inequalities that may be perpetuated as technology, such as this becomes available, initially inevitably at a high-cost premium, with the augmentation of individual intellect, breaking one of the fundamental unnoticed assumptions of equitable education–that all learners experience the same reality.

It is worthwhile for those engaged in augmented reality initiatives to consider five important pedagogical enablers of the technology [205]. The first is the need for augmented reality activities to be integrated into the totality of the learner's experiences. Activities designed to complement and work alongside other experiences will reduce the support resources needed. The role of the teacher needs to be empowered, with affordances designed into the augmented reality environment to allow the teacher to assert control of focus and attention when necessary to scaffold the class. Such affordances should also be to support monitoring and oversight by the teacher so that they maintain their awareness of the learners. Inevitable disruptions to the planned learning activities need to be responded to flexibly and without the necessity of significant technical support or impact on learner progress. Finally, the augmented reality experience needs to be purposeful and focused, delivering the features necessary for learning as simply as possible so as to ensure learners remain focused and not distracted or overwhelmed by unnecessary elaborations.

References

1. Munnerley, D., Bacon, M., Wilson, A., Steele, J., Hedberg, J., Fitzgerald, R.: Confronting an augmented reality. Res. Learn. Technol. **20**, 39–48 (2012)
2. Heilig, M.: Sensorama simulator. US Patent Number 3,050,870 (1962). http://www.mortonheilig.com/SensoramaPatent.pdf
3. Sutherland, E.I.: The ultimate display. Proc. IFIP Congr., 506–508 (1965)
4. Caudell, P.T., Mizell, W.D.: Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, vol. II, pp. 659–669 (1992)
5. Levy, S.: Google Glass 2.0 is a startling second act (2017). https://www.wired.com/story/google-glass-2-is-here/
6. Volkswagen: Virtual technologies (2019). https://www.volkswagenag.com/en/group/research/virtual-technologies.html
7. Fraga-Lamas, P., Fernández-Caramés, T.M., Blanco-Novoa, Ó., Vilar-Montesinos, M.A.: A review on industrial augmented reality systems for the industry 4.0 shipyard. IEEE Access. **6**, 13358–13375 (2018)
8. Mao, C.-C., Chen, F.-Y.: Augmented reality and 3-d visualization effects to enhance battlefield situational awareness. In: Ahram, T., et al. (eds.) IHIET 2019, AISC 1018, pp. 303–309 (2020)
9. Amaguaña, F., Collaguazo, B., Tituaña, J., Aguilar, W.G.: Simulation system based on augmented reality for optimization of training tactics on military operations. In: De Paolis, L.T., Bourdot, P. (eds.) AVR 2018, LNCS 10850, pp. 394–403 (2018)
10. Bridger, G.: BMW group developing augmented reality windshield displays (2011). http://www.motoringfile.com/2011/10/12/bmw-group-developing-augmented-reality-windshield-displays/
11. Popova, M.: Word lens: real-time translation via augmented reality (2010). https://bigthink.com/design-for-good/word-lens-real-time-translation-via-augmented-reality
12. Chiang, C.-L., Lin, Y.-L., Chao, H.-C., Chen, J.-Y., Lai, C.-H.: Effect of augmented reality on astronomical observation instruction. In: Rønningsbakk, L., et al. (eds.) ICITL 2019, LNCS 11937, pp. 184–193. Springer, Cham (2019)
13. Patrício, J.M., Costa, M.C., Manso, A.: A gamified mobile augmented reality system for systems and technologies, 19–22 June 2019, Coimbra, Portugal (2019)
14. Davis, M.: Ingress in geography: portals to academic success? J. Geogr. **116**(2), 89–97 (2017)
15. Mac, R.: The inside story of 'Pokémon Go's' evolution from Google castoff to global phenomenon, Forbes, August 23 (2016). https://www.forbes.com/sites/ryanmac/2016/07/26/monster-game/
16. Albouys-Perrois, J., Laviole, J., Briant, C., Brock, A.M.: Towards a multisensory augmented reality map for blind and low vision people: A participatory design approach. In: Proceedings of CHI 2018, Montréal, QC, Canada (2018)
17. de Oliveira, L.C., Andrade, A.O., de Oliviera, E.C., Soares, A., Cardoso, A., Lamounier, E.: Indoor navigation with mobile augmented reality and beacon technology for wheelchair users. In: Proceedings of IEEE EMBS International Conference on Biomedical & Health Informatics, Orlando, FL, USA (2017)
18. Thevin, L., Brock, A.M.: Augmented reality for people with visual impairments: designing and creating audio-tactile content from existing objects. In: Miesenberger, K., Kouroupetroglou, G. (eds.) Proceedings of ICCHP 2018, LNCS 10897, pp. 193–200 (2018)
19. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality. Found. Trends Hum. Comput. Interact. **8**(2–3), 73–272 (2013)

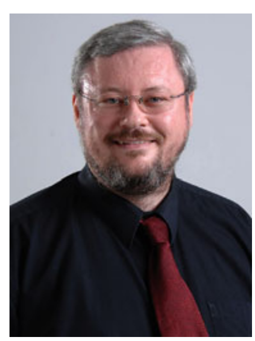
20. Akçayır, M., Akçayır, G.: Advantages and challenges associated with augmented reality for education: a systematic review of the literature. Educ. Res. Rev. **20**(1–11) (2017)
21. Antonioli, M., Blake, C., Sparks, K.: Augmented reality applications in education. J. Technol. Stud. **40**(1), 96–107 (2014)
22. Bacca, J., Baldiris, S., Fabregat, R., Graf, S., Kinshuk: Augmented reality trends in education: a systematic review of research and applications. Educ. Technol. Soc. **17**(4), 133–149 (2014)
23. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. Multimed. Tools Appl. **51**(1), 341–377 (2011)
24. Chen, P., Liu, X., Cheng, W., Huang, R.: A review of using augmented reality in education from 2011 to 2016. In: Popescu, E., et al. (eds.) Innovations in Smart Learning. Lecture Notes in Educational Technology, pp. 13–18. Springer, Singapore (2017)
25. Diegmann, P., Schmidt-kraepelin, M., Van Den Eynden, S., Basten, D.: Benefits of augmented reality in educational environments – a systematic literature review. Wirtschaftsinformatik. **3**(6), 1542–1556 (2015)
26. Garzón, J., Pavón, J., Baldiris, S.: Augmented reality applications for education: five directions for future research. In: De Paolis, L.T., et al. (eds.) AVR 2017, Part I, LNCS 10324, pp. 402–414 (2017)
27. Garzón, J., Pavón, J., Baldiris, S.: Systematic review and meta-analysis of augmented reality in educational settings. Virtual Reality. **23**, 447–459 (2019)
28. Ibáñez, M.B., Delgado-Kloos, C.: Augmented reality for STEM learning: a systematic review. Comput. Educ. **123**, 109–123 (2018)
29. Klimova, A., Bilyatdinova, A., Karsakov, A.: Existing teaching practices in augmented reality. Proc. Comput. Sci. **136**, 5–15 (2018)
30. Lee, K.: Augmented reality in education and training. TechTrends. **56**(2), 13–21 (2012)
31. Mekni, M., Lemieux, A.: Augmented reality: applications, challenges and future trends. Appl. Comput. Sci., 205–214 (2014)
32. Radu, I.: Why should my students use AR? A comparative review of the educational impacts of augmented-reality. In: Proceedings of 11th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2012, Science and Technology Papers, pp. 313–314 (2012)
33. Radu, I.: Augmented reality in education: a meta-review and cross-media analysis. Pers. Ubiquit. Comput. **18**(6), 1533–1543 (2014)
34. Wu, H.K., Lee, S.W.Y., Chang, H.Y., Liang, J.C.: Current status, opportunities and challenges of augmented reality in education. Comput. Educ. **62**, 41–49 (2013)
35. Craig, A.B.: Understanding Augmented Reality: Concepts and Applications. Morgan Kaufmann (2013)
36. Azuma, R.: A survey of augmented reality. Presence Teleop. Virt. **6**, 355–385 (1997)
37. Klopfer, E., Squire, K.: Environmental detectives: the development of an augmented reality platform for environmental simulations. Educ. Technol. Res. Dev. **56**(2), 203–228 (2008)
38. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum, pp. 282–292. SPIE, Bellingham (1995)
39. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Sys. **E77-D**(12), 1321–1329 (1994)
40. Benford, S., Brown, C., Reynard, G., Greenhaigh, C.: Shared spaces: transportation, artificiality, and spatiality. In: CSCW '96 Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, Boston, MA, USA, pp. 77–86 (1996)
41. Brooker, J.: The polytechnic ghost: Pepper's ghost, metempsychosis and the magic lantern at the royal polytechnic institution. Early Pop. Vis. Cult. **5**(2), 189–206 (2007)
42. Microsoft: Microsoft HoloLens hardware (2015). https://www.microsoft.com/microsoft-hololens/en-us/hardware
43. Bauer, V., Nagele, A., Baume, C., Cowlishaw, T., Cooke, H., Pike, C., Healey, P.G.: Designing an interactive and collaborative experience in audio augmented reality. In: Proceedings of Virtual Reality and Augmented Reality: 16th EuroVR International Conference, EuroVR 2019, Tallinn, Estonia, vol. 11883, pp. 305–311 (2019)
44. Tashev, I.J.: Capture, representation, and rendering of 3D audio for virtual and augmented reality. Int. J. Inf. Technol. Secur. **11**(SP2), 49–62 (2019)
45. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. Comput. Graph. Appl. IEEE. **21**(6), 34–47 (2001)
46. Billinghurst, M., Kato, H., Poupyrev, I.: The MagicBook: a transitional AR interface. Comput. Graph. **25**, 745–753 (2001)
47. Kan, T.-W., Teng, C.-H., Chen, M.Y.: QR code based augmented reality applications. In: Furht, B. (ed.) Handbook of Augmented Reality, pp. 339–354. Springer, New York (2011)
48. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A.: KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In: Proceedings of 24th Annual ACM Symposium on User Interface Software and Technology, pp. 559–568 (2011)
49. Azuma, R.: Tracking requirements for augmented reality. Commun. ACM. **36**, 50–51 (1995)
50. Martin, P., Ho, B.-J., Grupen, N., Muñoz, S., Srivastava, M.: Demo abstract: an iBeacon primer for indoor localization. In: Proceedings of BuildSys' 14, Memphis, pp. 190–191 (2014)
51. Joshi, R., Hiwale, A., Birajdar, S., Gound, R.: Indoor navigation with augmented reality. In: Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering ICCCE 2019, pp. 159–165
52. Hung, Y.-H., Chen, C.-H., Huang, S.-W.: Applying augmented reality to enhance learning: a study of different teaching materials. J. Comput. Assist. Learn. **33**, 252–266 (2017)
53. Cabero-Almenara, J., Barroso-Osuna, J., Llorente-Cejudo, C., del Mar Fernández Martínez, M.: Educational uses of augmented reality (AR): experiences in educational science. Sustainability. **11** (2019)
54. Cabero-Almenara, J., Roig-Vila, R.: The motivation of technological scenarios in augmented reality (AR): results of different experiments. Appl. Sci. **9**, 2907 (2019)
55. Liu, C., Chen, X., Liu, S., Zhang, X., Ding, S., Long, Y., Zhou, D.: The exploration on interacting teaching mode of augmented reality based on Hololens. In: Cheung, S.K.S., et al. (eds.) Proceedings of ICTE 2019, CCIS 1048, pp. 91–102 (2019)
56. Bacca, J., Baldirisa, S., Fabregat, R., Kinshuk, K.: Insights into the factors influencing student motivation in augmented reality learning experiences in vocational education and training. Front. Psychol. **9**, 1486 (2018) https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01486/full
57. Bacca, J., Baldiris, S., Fabregat, R., Kinshuk, S.G.: Mobile augmented reality in vocational education and training. Proc. Comput. Sci. **75**, 49–58 (2015). https://doi.org/10.1016/j.procs.2015.12.203
58. Chang, H.-Y., Wu, H.-K., Hsu, Y.-S.: Integrating a mobile augmented reality activity to contextualize student learning of a socioscientific issue. Br. J. Educ. Technol. **44**(3), E95–E99 (2013)

59. Bower, M., Howe, C., McCredie, N., Robinson, A., Grover, D.: Augmented reality in education – cases, places and potentials. Educ. Media Int. **51**(1), 1–15 (2014)
60. Sommerauer, P., Müller, O.: Augmented reality in informal learning environments: a field experiment in a mathematics exhibition. Comput. Educ. **79**, 59–68 (2014)
61. Mayer, R.E.: Multimedia learning: are we asking the right questions? Educ. Psychol. **32**(1), 1–19 (1997)
62. Mayer, R.E.: Multimedia Learning, 2nd edn. Cambridge University Press, Cambridge (2009)
63. Mayer, R.E.: Multimedia learning. Psychol. Learn. Motiv. **41**, 85–139 (2002)
64. Mikropoulos, T.A., Natsis, A.: Educational virtual environments: a ten-year review of empirical research (1999–2009). Comput. Educ. **56**, 769–780 (2011)
65. Brown, J.S., Collins, A., Duguid, P.: Situated cognition and the culture of learning. Educ. Res. **18**(1), 32–42 (1989)
66. Lave, J., Wenger, E.: Situated Learning. Cambridge University Press, Cambridge, MA (1991)
67. Wenger, E.: Communities of Practice. Cambridge University Press, New York (1998)
68. Kolb, D.A.: Experiential Learning. Prentice-Hall, Englewood Cliffs (1984)
69. Moorhouse, N., Tom Dieck, M.C., Jung, T.: An experiential view to children learning in museums with augmented reality. Mus. Manag. Curatorsh. **34**(4), 402–418 (2019)
70. Huang, T.-C., Chen, C.-C., Chou, Y.-W.: Animating eco-education: to see, feel, and discover in an augmented reality-based experiential learning environment. Comput. Educ. **96**, 72–82 (2016)
71. Freina, L., Ott, M.: A literature review on immersive virtual reality in education: state of the art and perspectives. In: Proceedings of eLearning and Software for Education (eLSE), Bucharest, Romania (2015) http://www.itd.cnr.it/download/eLSE%202015%20Freina%20Ott%20Paper.pdf
72. Dunleavy, M.: Persistent design challenges: augmenting reality for learning with wireless mobile devices. In: Proceedings of Invitation Symposia at Society for Information Technology and Teacher Education (SITE), San Diego, CA (2010)
73. Klopfer, E.: Augmented Learning. MIT Press, Cambridge, MA (2008)
74. Squire, K.D., Jan, M., Matthews, J., Wagler, M., Martin, J., Devane, B., Holden, C.: Wherever you go, there you are: the design of local games for learning. In: Sheldon, B., Wiley, D. (eds.) The Design and Use of Simulation Computer Games in Education, pp. 265–296. Sense, Rotterdam (2007)
75. Broll, W., Lindt, I., Herbst, I., Ohlenburg, J., Braun, A.K., Wetzel, R.: Toward next-gen mobile AR games. IEEE Comput. Graph. Appl. **28**(4), 40–48 (2008)
76. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. J. Sci. Educ. Technol. **18**(1), 7–22 (2009)
77. Hew, K.F., Cheung, W.S.: Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: a review of the research. Br. J. Educ. Technol. **41**(1), 33–55 (2010)
78. Neuhofer, B., Buhalis, D., Ladkin, A.: A typology of technology-enhanced tourism experiences. Int. J. Tour. Res. **16**, 340–350 (2014)
79. Squire, K., Jan, M.: Mad city mystery: developing scientific argumentation skills with a place-based augmented reality game on handheld computers. J. Sci. Educ. Technol. **16**(1), 5–29 (2007)
80. Lombard, M., Ditton, T.: At the heart of it all: the concept of presence. J. Comput.-Mediat. Commun. **3**(2) (1997)
81. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos. Trans. R. Soc. Lond. **364**, 3549–3557 (2009)
82. Cummings, J.J., Bailenson, J.N.: How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol. **19**(2), 272–309 (2016)
83. Shin, D.: Empathy and embodied experience in virtual environment: to what extent can virtual reality stimulate empathy and embodied experience? Comput. Hum. Behav. **77**, 64–73 (2018)
84. Slater, M., Sanchez-Vives, M.V.: Enhancing our lives with immersive virtual reality. Front. Robot. AI. **3**, 1–47 (2016)
85. Bursztyn, N., Shelton, B., Walker, A., Pederson, J.: Increasing undergraduate interest to learn geoscience with GPS-based augmented reality field trips on students' own smartphones. GSA Today. **27**(6), 4–10 (2017)
86. Chiang, T.H.C., Yang, S.J.H., Hwang, G.-H.: Students' online interactive patterns in augmented reality-based inquiry activities. Comput. Educ. **78**, 97–108 (2014)
87. Wang, H.-Y., Duh, H.B.-L., Li, N., Lin, T.-J., Sai, C.-C.: An investigation of university students' collaborative inquiry learning behaviors in an augmented reality simulation and a traditional simulation. J. Sci. Educ. Technol. **23**, 682–691 (2014)
88. Kamarainen, A.M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M.S., Dede, C.: EcoMOBILE: integrating augmented reality and probeware with environmental education field trips. Comput. Educ. **68**, 545–556 (2013)
89. Zimmerman, H.T., Land, S.M., Jung, Y.J.: Using augmented reality to support children's situational interest and science learning during context-sensitive informal mobile learning. In: Peña-Ayala, A. (ed.) Mobile, Ubiquitous, and Pervasive Learning. Advances in Intelligent Systems and Computing, vol. 406. Springer, Cham (2016)
90. Sobel, D.: Place-Based Education: Connecting Classroom and Community. The Orion Society, Great Barrington (2004)
91. Tarng, W., Ou, K.-L., Yu, C.-S., Liou, F.-L., Liou, H.-H.: Development of a virtual butterfly ecological system based on augmented reality and mobile learning technologies. Virtual Reality. **19**, 253–266 (2015)
92. Tussyadiah, I.P., Wang, D., Jung, T.H., tom Dieck, M.C.: Virtual reality, presence, and attitude change: empirical evidence from tourism. Tour. Manag. **66**, 140–154 (2018)
93. Chen, Y.C., Chi, H.L., Hung, W.H., Kang, S.C.: Use of tangible and augmented reality models in engineering graphics courses. J. Prof. Issues Eng. Educ. Pract. **137**(4), 267–276 (2011)
94. Kerawalla, L., Luckin, R., Seljeflot, S., Woolard, A.: "Making it real": exploring the potential of augmented reality for teaching primary school science. Virtual Reality. **10**(3), 163–174 (2006)
95. Yip, J., Wong, S.-H., Yick, K.-L., Chan, K., Wong, K.-H.: Improving quality of teaching and learning in classes by using augmented reality video. Comput. Educ. **128**, 88–101 (2019)
96. Arvanitis, T.N., Petrou, A., Knight, J.F., Savas, S., Sotiriou, S., Gargalakos, M., Gialouri, E.: Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities. Pers. Ubiquit. Comput. **13**(3), 243–250 (2009)
97. Cai, S., Wang, X., Chiang, F.-K.: A case study of augmented reality simulation system application in a chemistry course. Comput. Hum. Behav. **37**(31–40) (2014)
98. Sotiriou, S., Bogner, F.X.: Visualizing the invisible: augmented reality as an innovative science education scheme. Adv. Sci. Lett. **1**, 114–122 (2008)
99. Safar, A.H., Al-Jafar, A.A., Al-Yousefi, Z.H.: The effectiveness of using augmented reality apps in teaching the English alphabet to kindergarten children: a case study in the state of Kuwait. EURASIA J. Math. Sci. Technol. Educ. **13**(2), 417–440 (2017)
100. Dünser, A., Walker, L., Horner, H., Bentall, D.: Creating interactive physics education books with augmented reality. In: Proceedings of OzCHI '12: 24th Australian Computer-Human Interaction Conference, pp. 107–114 (2012)

101. Martín-Gutiérrez, J., Saorín, J.L., Contero, M., Alcañiz, M., Pérez-LÓpez, D.C., Ortega, M.: Design and validation of an augmented book for spatial abilities development in engineering students. Comput. Graph. **34**, 77–91 (2010)
102. Koong Lin, H.-C., Chen, M.-C., Chan, C.-K.: Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system. Interact. Learn. Environ. **23**(6), 799–810 (2015)
103. Moorhouse, N., Jung, T.: Augmented reality to enhance the learning experience in cultural heritage tourism: an experiential learning cycle perspective. eReview Tour. Res., 8 (2017) http://agrilife.org/ertr/files/2016/12/RN58.pdf
104. Salmi, H., Thuneberg, H., Vainikainen, M.-P.: Making the invisible observable by Augmented Reality in informal science education context. Int. J. Sci. Educ. Part B. **7**(3), 253–268 (2017)
105. Klopfer, E., Perry, J., Squire, K., Jan, M.-F., Steinkuehler, C.: Mystery at the museum: a collaborative game for museum education, CSCL '05. In: Proceedings of the 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years! Taipei, Taiwan, pp. 316–320
106. Dainenko, Y., Ipalakova, M., Tsoy, D., Shaipiten, A., Bolatov, Z., Chinibayeva, T.: Development of practical tasks in physics with elements of augmented reality for secondary educational institutions. In: De Paolis, L.T., Bourdot, P. (eds.) AVR 2018, LNCS 10850, pp. 404–412 (2018)
107. Freitas, D., Kouroupetroglou, G.: Speech technologies for blind and low vision persons. Technol. Disabil. **20**(2), 135–156 (2008)
108. Frank, J.A., Kapila, V.: Mixed-reality learning environments: integrating mobile interfaces with laboratory test-beds. Comput. Educ. **110**, 88–104 (2017)
109. Rumiński, D., Maik, M., Walczak, K.: Mixed reality stock trading visualization system. In: De Paolis, L.T., Bourdot, P. (eds.) Proceedings of AVR 2018, LNCS 10850, pp. 301–310 (2018)
110. Tarng, W., Ou, K.L., Lu, Y.C., Shih, Y.S., Liou, H.H.: A sun path observation system based on augment reality and mobile learning. Mobile Inf. Syst. **2018**, 1–10 (2018)
111. Tarng, W., Lin, Y.S., Lin, C.P., Ou, K.-L.: Development of a lunar-phase observation system based on augmented reality and mobile learning technologies. Mobile Inf Syst. **2016**, 1–12 (2016)
112. Cai, S., Chiang, F.K., Sun, Y., Lin, C., Lee, J.J.: Applications of augmented reality-based natural interactive learning in magnetic field instruction. Interact. Learn. Environ. **25**(6), 778–791 (2017)
113. Lin, T.-J., Duh, H.B.-L., Li, N., Wang, H.-Y., Tsai, C.-C.: An investigation of learners' collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system. Comput. Educ. **68**, 314–321 (2013)
114. Ibáñez, M.-B., Di-Serio, Á., Villarán-Molina, D., Delgado-Kloos, C.: Augmented reality-based simulators as discovery learning tools: an empirical study. IEEE Trans. Educ. **58**(3), 208–213 (2015)
115. Ibáñez, M.-B., Di-Serio, Á., Villarán, D., Delgado-Kloos, C.: Experimenting with electromagnetism using augmented reality: impact on flow student experience and educational effectiveness. Comput. Educ. **71**, 1–13 (2014)
116. Ma, M., Fallavollita, P., Seelbach, I., Von Der Heide, A.M., Euler, E., Waschke, J., Navab, N.: Personalized augmented reality for anatomy education. Clin. Anat. **29**, 446–453 (2016)
117. Fonseca, D., Martí, N., Redondo, E., Navarro, I., Sánchez, A.: Relationship between student profile, tool use, participation, and academic performance with the use of Augmented Reality technology for visualized architecture models. Comput. Hum. Behav. **31**, 434–445 (2014)
118. Grossman, S.: Professors envision using Google Glass in the classroom, Wired Campus June 20, 2013 (2013). http://chronicle.com/blogs/wiredcampus/professors-envision-using...glass-in-the-classroom/44401
119. Zarraonandia, T., Aedo, I., Diaz, P., Montero, A.: An augmented lecture feedback system to support learner and teacher communication. Br. J. Educ. Technol. **44**(4), 616–628 (2013)
120. Roussou, M.: Learning by doing and learning through play: an exploration of interactivity in virtual environments for children. Comput. Entertain. **2**(1), 1–23 (2004)
121. Lin, H.C.K., Hsieh, M.C., Wang, C.H., Sie, Z.Y., Chang, S.H.: Establishment and usability evaluation of an interactive AR learning system on conservation of fish. Turk. Online J. Educ. Technol. **10**(4), 181–187 (2011)
122. Yu, D., Jin, J.S., Luo, S., Lai, W., Huang, Q.: A useful visualization technique: a literature review for augmented reality and its application, limitation & future direction. In: Huang, L.M., Nguyen, V.Q., Zhang, K. (eds.) Visual Information Communication, pp. 311–337. Springer, Boston (2009)
123. Punako, R.: Computer-supported collaborative learning using augmented and virtual reality in museum education, Doctoral dissertation, Nova Southeastern University (2018). https://nsuworks.nova.edu/gscis_etd/1052
124. Ke, F., Hsu, Y.-C.: Mobile augmented-reality artifact creation as a component of mobile computer-supported collaborative learning. Internet High. Educ. **26**, 33–41 (2015)
125. Gopalan, V., Zulkifli, A.N., Faisal Mohamed, N.F., Alwi, A., Che Mat, R., Abu Bakar, J.A., et al.: Evaluation of e-star: an enhanced science textbook using augmented reality among lower secondary school student. Jurnal Teknologi. (2015)
126. Buker, T.J., Vincenzi, D.A., Deaton, J.E.: The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: reducing apparent latency with predictive compensation. Hum. Factors. **54**(2), 235–249 (2012)
127. Papadakis, G., Mania, K., Koutroulis, E.: A system to measure, control and minimize end-to-end head tracking latency in immersive simulations. In: Paper Presented at the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry, Hong Kong, China (2011)
128. Iftene, A., Trandabăț, D.: Enhancing the attractiveness of learning through augmented reality. Proc. Comput. Sci. **126**, 166–175 (2018)
129. McLellan, H.: Virtual realities. In: Jonassen, D. (ed.) Handbook of Research on Educational Communications and Technology, pp. 461–497. Routledge, New York (2004)
130. Louis, T., Troccaz, J., Rocher-Capellan, A., Bérard, F.: Is it real? Measuring the effect of resolution, latency, frame rate and jitter on the presence of virtual entities. In: Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces, pp. 5–16 (2019)
131. Billinghurst, M.: Grand challenges for augmented reality. Front. Virtual Real. **2**, 578080 (2021)
132. Kim, K., Billinghurst, M., Bruder, G., Duh, H.B.-L., Welch, G.F.: Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008–2017). IEEE Trans. Vis. Comput. Graph. **24**(11), 2947–2932 (2018)
133. Sieluzycki, C., Kaczmarczyk, P., Sobecki, J., Witkowski, K., Maslinski, J., Cieslinski, W.: Microsoft Kinect as a tool to support training in professional sports: augmented reality application to tachi-waza techniques in judo. In: Paper Presented at the 2016 Third European Network Intelligence Conference (ENIC), Wroclaw, Poland (2016)
134. Mahmood, Z., Ali, T., Muhammad, N., Bibi, N., Shahzad, M.I., Azmat, S.: EAR: enhanced augmented reality system for sports entertainment applications. KSII Trans. Internet Inf. Syst. **11**(12), 6069–6091 (2017)
135. Chihara, T., Seo, A.: Evaluation of physical workload affected by mass and center of mass of head-mounted display. Appl. Ergon. **68**, 204–212 (2018)

136. Cometti, C., Païzis, C., Casteleira, A., Pons, G., Babult, N.: Effects of mixed reality head-mounted glasses during 90 minutes of mental and manual tasks on cognitive and physiological functions. PeerJ. **6**, e5847 (2018)
137. Chang, Y.-L., Hou, H.-T., Pan, C.-Y., Sung, Y.-T., Chang, K.-E.: Apply an augmented reality in a mobile guidance to increase sense of place for heritage places. J. Educ. Technol. Soc. **18**(2), 166–178 (2015)
138. Zhan, T., Yin, K., Xiong, J., He, Z., Wu, S.-T.: Augmented reality and virtual reality displays: perspectives and challenges. iScience, 101397 (2020)
139. Trepkowski, C., Eibich, D., Maiero, J., Marquardt, A., Kruijff, E., Feiner, S.: The effect of narrow field of view and information density on visual search performance in augmented reality. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, pp. 575–584 (2019)
140. Kegeleers, M., Miglani, S., Reichert, G.M.W., Salamon, N.Z., Balint, J.T., Lukosch, S.G., Bidarra, R.: STAR: superhuman training in augmented reality. In: Paper Presented at the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities, Delft, Netherlands (2018)
141. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. Robot. Comput. Integr. Manuf. **49**, 215–228 (2018)
142. Furió, D., González-Gancedo, S., Juan, M.-C., Seguí, I., Costa, M.: The effects of the size and weight of a mobile device on an educational game. Comput. Educ. **64**, 24–41 (2013)
143. Dunleavy, M., Dede, C.: Augmented reality teaching and learning. In: Spector, J.M., et al. (eds.) Handbook of Research on Educational Communications and Technology, pp. 735–745. Springer, New York (2014)
144. Sanna, A., Manuri, F., Piumatti, G., Paravati, G., Lamberti, F., Pezzola, P.: A flexible AR-based training system for industrial maintenance. In: De Paolis, L.T., Mongelli, A. (eds.) Proceedings of AVR 2015, LNCS 9254, pp. 314–331. Springer, Cham (2015)
145. Ierache, J., Mangiarua, N.A., Becerra, M.E., Igarza, S.: Framework for the development of augmented reality applications applied to education games. In: De Paolis, L.T., Bourdot, P. (eds.) Proceedings of AVR 2018, LNCS 10850, pp. 340–350. Springer, Cham (2018)
146. Weisner, C.A., Klinker, G.: Overcoming location inaccuracies in augmented reality navigation. In: De Paolis, L.T., et al. (eds.) AVR 2017, Part I, LNCS 10324, pp. 377–388 (2017)
147. Joshi, R., Hiwale, A., Birajdar, S., Gound, R.: Indoor navigation with augmented reality. In: Kumar, A., Mozar, S. (eds.) ICCCE 2019, Lecture Notes in Electrical Engineering, p. 570 (2019)
148. Chiang, T.H., Yang, S.J., Hwang, G.-J.: An augmented reality-based mobile learning system to improve students' learning achievements and motivations in natural science inquiry activities. J. Educ. Technol. Soc. **17**(4), 352–365 (2014)
149. Perry, T.S.: Look out for Apple's AR glasses: with head-up displays, cameras, inertial sensors, and lidar on board, Apple's augmented-reality glasses could redefine wearables. IEEE Spectr. **58**(1), 26–54 (2021)
150. Stein, S.: Lidar on the iPhone 12 Pro is cool now, but it's going to be so much cooler (2021). https://www.cnet.com/how-to/future-of-lidar-cool-now-going-to-be-cooler-apple-iphone-12-and-ipad-pro/
151. Peterson, A.: Holocaust Museum to visitors: please stop catching Pokémon here, The Washington Post (2016). https://www.washingtonpost.com/news/the-switch/wp/2016/07/12/holocaust-museum-to-visitors-please-stop-catching-pokemon-here/
152. Macfarlane, J.: Your navigation app is making traffic unmanageable, IEEE Spectrum (2019). https://spectrum.ieee.org/computing/hardware/your-navigation-app-is-making-traffic-unmanageable
153. Zhang, W., Han, B., Hui, P.: On the networking challenges of mobile augmented reality. In: Proceedings VR/AR Network '17, Los Angeles, CA, USA (2017)
154. Akçayır, M., Akçayır, G., Pektas, H.M., Ocak, M.A.: Augmented reality in science laboratories: the effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories. Comput. Hum. Behav. **57**, 334–342 (2016)
155. Muñoz-Cristóbal, J.A., Jorrin-Abellán, I.M., Asensio-Pérez, J.I., Martinez-Monés, A., Prieto, L.P., Dimitriadis, Y.: Supporting teacher orchestration in ubiquitous learning environments: a study in primary education. Learn. Technol. IEEE Trans. Learn. **8**(1), 83–97 (2015)
156. Ahn, J., Ahn, E., Min, S., Choi, H., Kim, H., Kim, G.J.: Size perception of augmented objects by different AR displays. In: Stephanidis, C. (ed.) HCI International 2019 – Posters. HCII 2019. Communications in Computer and Information Science, vol. 1033. Springer, Cham (2019)
157. Lichty, P.: Making inside the augment: augmented reality and art/design education. In: Geroimenko, V. (ed.) Augmented Reality in Education: A New Technology for Teaching and Learning, pp. 261–278. Springer, Cham (2020)
158. Norman, D.A.: The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution. MIT Press (1998)
159. Carter, M., Egliston, B.: Ethical Implications of Emerging Mixed Reality Technologies. University of Sydney (2020) https://ses.library.usyd.edu.au/handle/2123/22485
160. Rose, D.H., Harbour, W.S., Johnston, C.S., Daley, S.G., Abarbanell, L.: Universal design for learning in postsecondary education: reflections on principles and their application. J. Postsecond. Educ. Disabil. **19**(2), 135–151 (2006)
161. Bacca, J., Fabregat, R., Baldiris, S., Graf, S., Kinshuk: Augmented reality trends in education: a systematic review of research and applications. Educ. Technol. Soc. **17**, 133–149 (2014)
162. Chen, C.-M., Tsai, Y.-N.: Interactive augmented reality system for enhancing library instruction in elementary schools. Comput. Educ. **59**(2), 638–652 (2012)
163. Ibáñez, M.-B., Di-Serio, Á., Villarán-Mollina, D., Delgado-Kloos, C.: Support for augmented reality simulation systems: the effects of scaffolding on learning outcomes and behavior patterns. IEEE Trans. Learn. Technol. **9**(1), 46–56 (2016)
164. Chen, C.-H., Chou, Y.-Y., Huang, C.-Y.: An augmented-reality-based concept map to support mobile learning for science. Asia-Pac. Educ. Res. **25**(4), 567–578 (2016)
165. Csikszentmihalyi, M.: Flow. Harper and Row, New York (1990)
166. Stanney, K., Lawson, B.D., Rokers, B., Dennison, M., Fidopiastis, C., Stoffregen, T., Weech, S., Fulvio, J.M.: Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda. Int. J. Hum. Comput. Interact. **36**(19), 1783–1803 (2020)
167. Hughes, C.L., Fidopiastis, C.M., Stanney, K.M., Bailey, P.S., Ruiz, E., Frye, D.L., Karluk, F.J.: The psychometrics of cybersickness in augmented reality. Front. Virtual Real. **1**, 602954 (2020)
168. Stanney, K.M., Hale, K.S., Nahmens, I., Kennedy, R.S.: What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. Hum. Factors. **45**(3), 504–520 (2003)
169. Ling, Y., Nefs, H.T., Brinkman, W., Qu, C., Heynderickx, I.: The relationship between individual characteristics and experienced presence. Comput. Hum. Behav. **29**(4), 1519–1530 (2013)
170. Golding, J.F.: Motion sickness susceptibility. Auton. Neurosci. **129**(1), 67–76 (2006)
171. Stanney, K.M., Kennedy, R.S.: Aftereffects from virtual environment exposure: how long do they last? Proc. Hum. Factors Ergon. Soc. Annu. Meet. **42**(21), 1476–1480 (1998)

172. Escaravajal-Rodríguez, J.C.: Pokémon GO and its influence on Spanish Facebook users. Apunts: Educacion Fisica y Deportes. **133**, 38–49 (2018)
173. Pokémon GO Death Tracker (2020). http://pokemongodeathtracker.com
174. Moon, J., Koo, D., Kim, K., Shin, I., Kim, H., Kim, J.: Effect of ski simulator training on kinematic and muscle activation of the lower extremities. J. Phys. Ther. Sci. **27**(8), 2629–2632 (2015)
175. Gabbard, J.L., Mehra, D.G., Swan, J.E.: Effects of AR display context switching and focal distance switching on human performance. IEEE Trans. Vis. Comput. Graph. **25**(6), 2228–2241 (2019)
176. Olsson, T., Lagerstam, E., Kärkkäinen, T., Väänänen-Vainio-Mattila, K.: Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. Pers. Ubiquit. Comput. **17**(2), 287–304 (2013)
177. Popper, B.: New evidence emerges in alleged assault on cyborg at Paris McDonald's (2012). http://www.theverge.com/2012/7/19/3169889/steve-mann-cyborg-assault-mcdonalds-eyetap-paris
178. Due, B.L.: The social construction of a glasshole: Google glass and multiactivity in social interaction. PsychNol J. **13**(2–3), 149–178 (2015)
179. Honan, M.: I, Glasshole: my year with Google Glass (2013). https://www.wired.com/2013/12/glasshole/
180. Marshall, S.J.: Futures scenarios as tools for sense-giving and sense-making. In: Shaping the University of the Future. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7620-6_20
181. Stanney, K., Winslow, B., Hale, K., Schmorrow, D.: Chapter xxix: Augmented cognition. In: Boehm-Davis, D.A., Durso, F.T., Lee, J.D. (eds.) APA Handbook of Human Systems Integration, pp. 329–343. APA, Washington, DC (2015)
182. Remington, R.W., Loft, S.: Chapter xxix: Attention and multitasking. In: Boehm-Davis, D.A., Durso, F.T., Lee, J.D. (eds.) APA Handbook of Human Systems Integration, pp. 261–276. American Psychological Association, Washington, DC (2015) 625 pages
183. Vertegaal, R.: Attentive user interfaces. Commun. ACM. **46**(3), 31–33 (2003)
184. Dey, A., Billinghurst, M., Lindeman, R.W., Swan II, J.E.: A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. Front. Robot. AI. **5**, 37 (2018)
185. Carter, R.: Zoom immersive scenes vs Microsoft Teams together mode (2020). https://www.uctoday.com/collaboration/team-collaboration/zoom-immersive-scenes-vs-microsoft-teams-together-mode/
186. McCreary, J.R.: The laptop-free zone. Valparaiso Univ. Law Rev. **43**(3), 989–1044 (2009)
187. Mortkowitz, L.: More colleges, professors shutting down laptops and other digital distractions, Washington Post (2010). http://www.washingtonpost.com/wp-dyn/content/article/2010/04/24/AR2010042402830.html
188. Young, J.R.: College 2.0: teachers without technology strike back, The Chronicle of Higher Education (2010). http://chronicle.com/article/College-20-Teachers-Without/123891/
189. Kraushaar, J.M., Novak, D.C.: Examining the effects of student multitasking with laptops during the lecture. J. Inf. Syst. Educ. **21**(2), 241–251 (2010)
190. Moreno, M.A., Jelenchick, L., Koff, R., Eikoff, J., Diermyer, C., Christakis, D.A.: Internet use and multitasking among older adolescents: an experience sampling approach. Comput. Hum. Behav. **28**, 1097–1102 (2012)
191. Barry, S., Murphy, K., Drew, S.: From deconstructive misalignment to constructive alignment: exploring student uses of mobile technologies in university classrooms. Comput. Educ. **81**, 202–210 (2015)
192. Taneja, A., Fiore, V., Fischer, B.: Cyber-slacking in the classroom: potential for digital distraction in the new age. Comput. Educ. **82**, 141–151 (2015)
193. Sana, F., Weston, T., Cepeda, N.J.: Laptop multitasking hinders classroom learning for both users and nearby peers. Comput. Educ. **62**, 24–31 (2013)
194. Marshall, S.J.: Technology and modern students – the digital natives fallacy. In: Shaping the University of the Future. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7620-6_10
195. Kharbach, M.: 7 great web tools to help teachers annotate web pages (2021). https://www.educatorstechnology.com/2018/02/7-great-web-tools-to-help-teachers.html
196. Kimmons, R., Veletsianos, G.: Education scholars' evolving uses of twitter as a conference backchannel and social commentary platform. Br. J. Educ. Technol. **47**(3), 445–464 (2016)
197. Gesthuizen, R., Rablin, A.: Tweaching: learning that is personal and social. In: Proceedings Australasian Conference on Educational Computing, Adelaide, Australia (2014)
198. Robertson, A.: Is the world ready for virtual graffiti? (2019). https://www.theverge.com/2019/10/12/20908824/mark-ar-google-cloud-anchors-social-art-platform-harassment-moderation
199. Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., Watson, Z., Breen, G., Schwarz, R., Steptoe, W., Szostak, D., Halan, S., Fox, D., Silver, J.: The ethics of realism in virtual and augmented reality. Front. Virtual Real. **1**(1) (2020)
200. Hofmann, B., Haustein, D., Landeweerd, L.: Smart-glasses: exposing and elucidating the ethical issues. Sci. Eng. Ethics. **23**, 701–721 (2017)
201. Hosfelt, D.: Making ethical decisions for the immersive web. arXiv preprint arXiv, 1905.06995 (2019)
202. Pase, S.: Ethical considerations in augmented reality applications. In: Proceedings of the International Conference on E-Learning E-Business, Enterprise Information Systems, and E-Government EEE The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 1. Las Vegas, Nevada, USA, 16–19 July (2012)
203. Peske, M.: The last days of reality. Meanjin Quarterly, Summer 2017 (2017). https://meanjin.com.au/essays/the-last-days-of-reality/
204. Brey, P.: The ethics of representation and action in virtual reality. Ethics Inf. Technol. **1**(1), 5–14 (1999)
205. Cuendet, S., Bonnard, Q., Do-Lenh, S., Dillenbourg, P.: Designing augmented reality for the classroom. Comput. Educ. **68**, 557–569 (2013)



Stephen Marshall (PFHEA) researches in the areas of organizational change in higher education, quality, benchmarking, plagiarism and academic integrity, intellectual property and policy and strategy for effective use of technology by educational organizations. He is co-creator of the internationally recognized and applied e-learning maturity model (eMM, http://e-learning.geek.nz/emm/).

Augmented Reality in Sports and Physical Education

14

Jia Zhang and Yang-Sheng Huang

Contents

14.1 Introduction 355
14.2 Status Quo of PE 357
14.3 Application of IT in PE 358
14.4 Design of an AR-Assisted Learning System for PE 359
14.5 Research Methods and Verification 360
14.6 Learning Outcomes of AR-Assisted Learning 362
14.7 Motor Skill Acquisition by Using AR-PEclass 363
14.8 Learning Motivation Stimulated by AR-PEclass 364
14.9 Conclusion 366
References 366

Abstract

Physical education (PE) courses involve the provision of sports knowledge and the engagement in motor skill practices. Instruction based on augmented reality (AR) has been rarely applied in the hands-on teaching of these courses. Currently, PE teaching is mostly aided by video-assisted instructions. However, such instruction does not provide interactive experience in practices and fails to integrate textbook-based static learning with dynamic learning based on motor skill demonstrations. Because AR can overlay virtual information on a real object, this technology enables learners to, for example, manipulate an interactive three-dimensional human character model overlayed on a textbook while reading the textbook. Thus, AR can overcome the disadvantage of video-assisted instruction. To investigate the effects of AR-assisted instruction on learning outcomes, motor skills of various difficulty levels, and learning motivation in students, two experiments were conducted using different teaching materials; specifically, instructions on basic running drills and Mach drills were offered to students. A quasi-experimental design was adopted. The results indicated that AR-based teaching materials outperformed video-based ones, particularly in the learning outcomes of difficult motor skills.

Keywords

Augmented reality · Educational technology · Physical education · Learning motivation · Motor skill learning

14.1 Introduction

Rapid advances in information technology (IT) and the Internet have resulted in them being applied to a wide extent in daily life and the workplace. Incorporating digital tools into teaching has become a current trend in education. Physical education (PE) involves the provision of instructions on various motor skills and techniques that require accuracy, coordination, and speed of body movements. Instructors must explain and demonstrate relevant techniques as well as guide students in practicing motor skills to facilitate an effective learning process of learners. However, conventional PE models can no longer satisfy learners' needs; thus, PE instructors are faced with various types of challenges. Consequently, unconventional PE strategies must be studied or designed for developing educational, innovative, and inspirational teaching content that can strengthen learning outcomes. Currently, PE instructors mostly emphasize skill demonstrations yet rarely allow students to explore their perceptions of different movements independently. Moreover, PE instructors tend to teach students through lecturing, leaving little leeway for students to think. The use of IT equipment has become prevalent, with improved IT literacy and competency among

J. Zhang (✉) · Y.-S. Huang
Intelligent Interactive Technology Learning Laboratory, Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_14

instructors and increased studies on teaching incorporated with IT. However, the findings of such studies cannot be effectively applied to real PE teaching due to its nature and unique requirements with regard to teaching sites. In addition to teaching academic knowledge and practical skills of sports, PE instructors and professionals are also faced with challenges as in how they can cultivate students' competencies of exploring, integrating, and adopting IT tools, sports knowledge, and sports skills.

Studies have indicated that audiovisual media can diversify the conventional teaching of PE, thereby enhancing students' interest in PE and improving learning outcomes [1, 2]. A study discovered that the provision of intriguing images in a PE class may attract attention and increase student motivation to learn [3]. Specifically, the video learning model has become the most common strategy for the incorporation of audiovisual technology into PE [1, 2]. In video-assisted instruction (VAI), students can repeatedly plan and play specific instructional videos according to their learning progress and needs. Students can practice or modify their movements according to demonstrations in the videos. Thus, the video learning model solves the problem of decreased learning motivation and learning outcomes caused by the failure of instructors, during demonstrations, to attend to each student in a large PE class. Furthermore, in regular training, learners can quickly situate themselves in different sports game scenarios by repeatedly watching and discussing on instructional videos, familiarize themselves with the timing of adopting appropriate strategies, and increase their chance of winning. The use of computer multimedia materials in conjunction with the provision of graphic and textual descriptions has been proved to be effective in promoting learning outcomes [4]. Therefore, an increasing number of studies have focused on adding functions to videos, such as video annotation, according to the characteristics of VAI.

Although it has the aforementioned advantages, video-assisted PE has certain limitations. For example, motor skills are executed in a three-dimensional (3D) space; however, videos generally present movements in a single angle or limited angles. Consequently, learners cannot change the presentation angles in videos for comprehensive observation of the demonstrations of each movement. Describing and demonstrating 3D motor skills to students are essential but difficult in a PE class [5]. To overcome this difficulty, teaching that incorporates 3D instructional materials is proposed. In such a teaching strategy, learners' cognition and memory are strengthened by watching movements demonstrated from various angles [6]. Moreover, learners can explore their body movements, determine problems in the movements, and improve their motor skills. A study demonstrated that with the aid of a 3D dynamic model, learners could interact with instructional materials, which exerted positive effects on their understanding and cognition of movements. In addition, 3D visualized images in the model brought new experiences to the learners and thereby attracted their attention and enhanced their learning motivation.

Another weakness of VAI is the difficulty in integrating videos with printed teaching materials. PE is related to factual knowledge (e.g., the mechanics and principles of different motor skills) and procedural knowledge (e.g., motor skill demonstrations). Text and graphic presentations are suitable for factual knowledge, whereas dynamic presentations are preferable for procedural knowledge. Most VAIs fail to integrate static text and graphics effectively with dynamic videos, which hinders learners from acquiring both types of knowledge. PE is a kinesthetic immersive experience [7, 8]; it involves body movements that change synchronously the perceived visual and auditory information and establish close interconnections between the perceived information and changes to body movements [9–11]. Simply creating a 3D virtual environment for learning can only reproduce the visual effects and cannot enable the integration of the virtual environment with real learning environments for interactive learning. Consequently, learners cannot accomplish immersive learning [12].

To solve the aforementioned problems, studies have proposed the application of augmented reality (AR) to teaching. For example, Klopfer and Squire defined AR as a form of presentation combining the real world with a virtual environment by using positioning technology [13]. Technically, AR is an example of context awareness application. This technology is characterized by connecting virtual information to the real world by using visualization techniques; thus, AR bridges the gap between information and target objects that originally exist in the context awareness framework [14, 15]. The use of AR in education generates novel sensory and interactive experiences for learners, deepens perceived impressiveness, and enhances learning outcomes [13, 16–19]. AR has also been widely applied in various types of teaching, such as museum tours [20–22], the teaching of changes in historic sites [23–26], teaching in off-campus activities [27–29], engineering education [30–32], and astronomy education [33, 34]. The aforementioned instructional uses of AR have demonstrated that this technology helps improve the learning outcomes, motivation, and concentration of learners as well as the attractiveness of physical objects to them [32, 35, 36].

AR combines the real and virtual domains. By integrating this characteristic of AR with PE courses, instructors can combine printed teaching materials (e.g., those concerning the mechanics and principles of motor skills) with virtual objects of 3D teaching materials (e.g., those related to the demonstrations and practice of motor skills), thereby allowing learners to read printed materials while manipulating 3D virtual elements. This enables learners to effectively integrate factual and procedural knowledge [37] and strengthen their immersive learning experiences [8]. The goal of virtuality–reality integration cannot be achieved through conventional teaching, video-based teaching, or 3D scenario simulation. In addition, learners establish a stronger connection with the content they are learning when they use AR technology for learning than when they learn through VAI or 3D simulation [38].

The aforementioned discussion on AR characteristics indicates that AR can overcome the weaknesses of video-assisted learning and improve learning outcomes and motivation. Therefore, this study developed an AR-based motor skill learning system comprising 3D display technology. This study also explored the effects of various teaching models (i.e., VAI and said AR system) on motor skill acquisition. The proposed AR system is expected to outperform the other approach in terms of helping learners to establish correct running concepts, develop running skills, acquire factual knowledge, and enhance their learning motivation.

14.2 Status Quo of PE

Motor skill learning is fundamental in PE. Instructors teach various sport-specific skills by organizing activities, such as movement demonstrations, observations, imitations, and practices. PE instructors must introduce the target motor skill, perform a demonstration of it, and teach students how to practice it. A motor skill is an ability to perform a series of target actions correctly by using the body. Learning a motor skill requires cognitive understanding; thus, learners must be familiar with the attributes and process of performing the target movement, engage in individual or integrative practices, identify mistakes when practicing the target movement, and finally reduce these mistakes for achieving high overall accuracy, coordination, and smoothness for the target movement [39]. Fitts and Posner indicated that the acquisition of motor skills by learners involves three phases: the cognitive, associative, and automatic phases [40]. The greatest challenge facing PE is that instruction must consider the complexity of movements (because a movement usually comprises numerous intricate submovements) as well as their spatiality (direction determination and movement decisions), accuracy (coordination of basic movements and the correct use of each body part), and speed requirement (the learner must complete target movements in an extremely short time) [41, 42]. The teaching of the process, smoothness, coordination, and completeness of a specific movement in PE classes emphasizes the structure and smoothness of the movement. Major concepts of movement structures and smoothness are defined using Laban movement analysis. The German dance artist Laban developed this analysis method to overcome the weaknesses of traditional dance notation systems. Unlike traditional dance notations that record only the time and steps, Laban movement analysis involves quantifying recordings of body movement by using descriptions of the body, time, space, effort, and relationship [43]. Laban categorized human movements on the basis of three elements:

1. Groundedness: Movements must be performed in depth without simply emphasizing the appearance. Instead, the performer should focus on deep muscles or joints as a starting point of the desired movement.
2. Integrity: Performers must connect intermittent or discontinued movements into a whole from the inside out and outside in, thereby achieving temporal and spatial body coordination.
3. Functionality: Each movement represents a distinct function in different fields with various forms of expression. When learning a movement, learners should learn how to use their body parts to execute the movement at the correct time and correctly, to change the environment with the movement, and to perform accurate and coordinated movement within a specified time [44].

Mosston proposed a spectrum of teaching styles. This spectrum is a crucial basis for PE teaching [45]. The spectrum of Mosston comprises 11 PE teaching styles, namely, the command, practice, reciprocal, self-check, inclusive, guided discovery, convergent discovery, divergent production, learner designed, learner initiated, and self-teach styles [45]. The command, practice, reciprocal, self-check, and inclusive styles represent teacher-oriented “reproductive” teaching, whereas the other six styles represent student-oriented “productive” teaching. Existing PE models are predominantly based on reproductive teaching (teacher demonstration); thus, students have few opportunities to explore a movement by themselves. PE instructors are likely to use narrative teaching and rarely allow students to think independently about a movement. When students practice a motor skill, instructors must observe their movements and provide them feedback individually. However, learners might not fully understand the target movement because of inappropriate viewing angles, excessively short time of demonstrations, or the demonstration of numerous movements simultaneously. Moreover, instructors cannot provide one-on-one instruction because of the excessive number of learners in a class. Consequently, beginners might not notice their mistakes when practicing a specific movement. Challenges have been posed to the practice of conventional one-on-one instruction, where PE instructors demonstrate a movement and teach students alternately for helping students acquire motor skill knowledge and perform movements accurately, because a PE class usually comprises 20–30 students. Therefore, an efficient and innovative teaching method must be developed for solving existing problems in PE. When adopting a conventional teaching method for providing motor skill instructions to athletes in a 3D space, coaches need to indicate athletes' mistakes and demonstrate the correct movements within a short time. Such a conventional method is not effective for helping student memorize the movements clearly and accurately or grasp tips on the execution of a specific movement. By using two-dimensional (2D) or 3D multimedia teaching materials that are not limited by space or time, instructors can guide learners and overcome the aforementioned disadvantage of conventional PE teaching [46].

Conventional PE models can no longer meet learners' needs; thus, PE instructors face numerous challenges. To overcome these challenges, a novel PE instructional strategy that incorporated IT must be developed for providing educational, innovative, and inspiring course content to students and thereby improving their learning outcomes.

14.3 Application of IT in PE

IT has substantially enhanced the development of the technology industry and improved the convenience of daily life. In the age of information explosion, the Internet and mobile devices have considerably changed the food, clothing, living, transportation, education, and entertainment aspects of daily human life. IT has been adopted in education, where E-learning has become a new teaching model in fields such as language, natural science, and humanities and social science. However, such technology has rarely been adopted in subjects based on procedural knowledge (e.g., PE, music, and arts). E-learning has changed existing teaching models, replaced textbooks with tablets, and replaced blackboards with electronic whiteboards. The learning styles of students have also changed. Since the introduction of the concept of a flipped classroom, students have had more control over their learning. For example, they can watch instructional videos and engage in hands-on activities on a teaching platform at home and then bring up problems they encounter to teachers in class [47]. E-learning has inspired novel teaching and learning models, such as mobile learning. Mobile learning refers to any mobile device-based learning method that helps learners to learn without temporal and spatial constraints by using the Internet . Facilitating learning outside classroom, mobile learning allows learners to engage in meaningful learning and interactions anywhere and anytime [48]. In IT-incorporated teaching, various visual and auditory multimedia are used to relieve students' cognitive burden they would otherwise face in learning through verbal and textual means. Furthermore, compared with conventional courses, IT-incorporated courses more effectively enhance learners' attention and learning outcomes. Papastergiou and Gerodimos used online multimedia courses as teaching materials to help college students acquire basketball skills (the courses covered an introduction to the sport and its rules as well as skill demonstrations) [49]. An experiment in which the participants were divided into the multimedia group (experimental group) and conventional instruction group (control group) revealed that the tablet-assisted instruction of motor skills was conducive to students' development of basketball skill knowledge.

With the rapid development of sports-related technologies, IT-assisted tools have become widely used in areas such as the analysis of training, competition, and basic movement instruction, sports event commentary, computer-assisted refereeing in professional sports games, the introduction to sports fields and rules, and the analysis of students' movements in practices. Computer multimedia can stimulate students' learning motivation, offer entertaining course content, and improve learning outcomes. Coaches and players expect multimedia IT systems to provide them various sports-related information, which allows players to improve their performance in both training and instruction. According to action research on the incorporation of IT into health and PE learning, the integration of IT, the Internet, and multimedia teaching materials into course content and instructional strategies enables students to accomplish meaningful learning. Furthermore, intriguing and vivid images can attract students' attention and enhance their learning motivation.

Zou adopted an online instructional platform to help elementary school students learn to play table tennis [50]. The results revealed that students who watched demonstration videos on the online platform exhibited superior learning outcomes to those who received conventional instruction. Students who watched the aforementioned videos also had higher learning motivation as well as better skill and knowledge acquisition than those who received conventional instruction. Furthermore, Zou discovered that the use of multimedia tools effectively strengthened student motivation in learning the eight-form Tai Chi. Specifically, participants were divided into three groups, namely, the conventional instruction group, 2D video instruction group, and 3D animation instruction group. After completing the course, the participants completed a questionnaire and interview on perceived learning outcomes and practice responses. The results indicated that the 3D group had higher outcome scores than the 2D group for four Tai Chi movements. The 3D group also had higher outcome scores than the conventional instruction group for all Tai Chi movements. The students who learned using the 3D animation strategy, which facilitated the use of multiple viewing angles, offered free control over the practice speed, and provided interesting and vivid content, experienced fewer difficulties during practice and higher learning interest than their counterparts did in the other two groups. The aforementioned research indicated that compared with conventional instruction, IT application in PE not only improves learning outcomes but also increases students' learning motivation.

Currently, video learning is the most common IT-incorporated PE teaching model. Shooting or making instructional videos of motor skills can help instructors overcome the disadvantages of conventional teaching and thus improve learners' performance in motor skill learning. The incorporation of video learning into PE teaching effectively combines sports theories with motor skill learning with the help of auditory and visual effects as well as digital animation. The uploading of these videos to an E-learning platform allows students to learn anytime and anywhere. In video learning, learners interpret video information by using their visual system, compare their movements with those demonstrated in videos, and identify the correct method of performing the movements. Video learning, in conjunction with feedback provided by instructors based on their professional knowledge to learners, achieves the desired outcome of multimedia-assisted learning.

However, video learning still exhibits several disadvantages. For example, due to spatial constraints, learners must watch instructional videos in the classroom or places with video-playing equipment before going to a suitable location for practice. Thus, video learning can sometimes fail to sustain the learning effect. Learners can sometimes forget what they have learned in a video. Moreover, to facilitate multiple viewing angles by learners for accurate and complete learning, a video must be produced by shooting the same scene several times from various angles. However, such multiple viewing angles can cause excessive cognitive burden on learners because they have to reorganize and reinterpret video information in their mind. Effective integration between PE instructional videos and hands-on activities on-site is not guaranteed because of spatial constraints related to video-playing equipment. Thus, improvements are required in video-learning-based PE instruction. We believe that incorporating 3D model technology in conjunction with mobile learning into PE instruction can solve the problems associated with video learning, help learners learn about motor skills better without being limited by video-learning-related or spatial constraints, and facilitate the successful application of IT in PE instruction.

14.4 Design of an AR-Assisted Learning System for PE

This study developed AR-PEclass, which is a piece of AR-based system for providing motor skill instructions. The AR-PEclass, which uses the virtuality–reality integration characteristic of AR and comprises a 3D motor skill model and

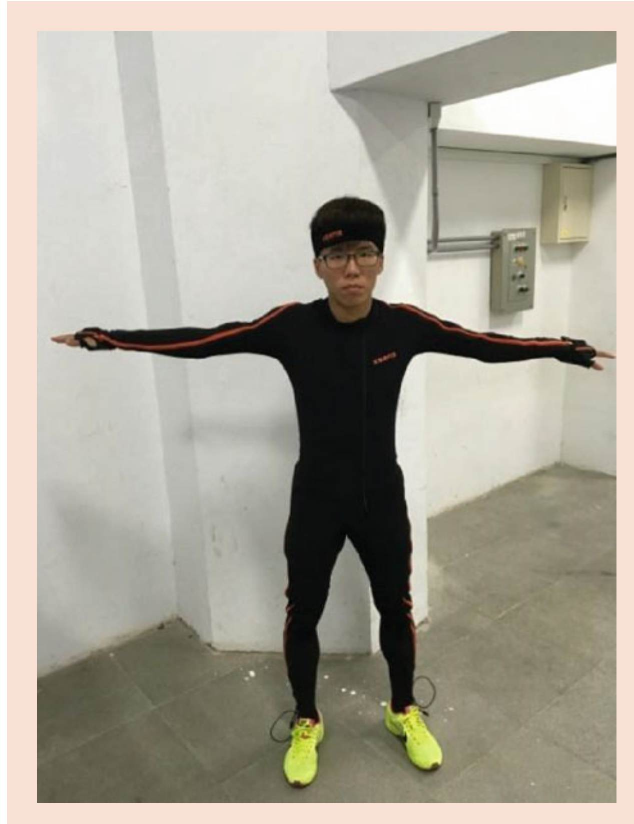


Fig. 14.1 Xsens MVN motion capture system

an AR interface, displays visualized instructions to students. The authors of this study expected AR-PEclass to overcome the problems related to motor skill performance and video viewing angles in conventional PR teaching and existing IT-based PE teaching. When using the AR-PEclass, learners learn about a motor skill by following the guidance provided in the system. By scanning textbook teaching materials with a mobile device, learners will see an integration of the virtual and real environments on the mobile device screen. Learners can then engage in learning by using the displayed information along with the textbook materials.

AR-PEclass was developed using Vuforia, which is a free AR software development kit released by Qualcomm, and Unity 3D, which is a game engine. An Xsens MVN motion capture system (Fig. 14.1) was used to capture 3D movements. The captured movement models were then imported into Unity 3D for system development. Unity 3D allows designers to create game environments and write scripts independently. Finally, the completed scripts were integrated with 3D models to develop AR-PEclass, as illustrated in Fig. 14.2.

As an example of incorporating AR into PE, AR-PEclass can be used together with textbooks by learners. When users

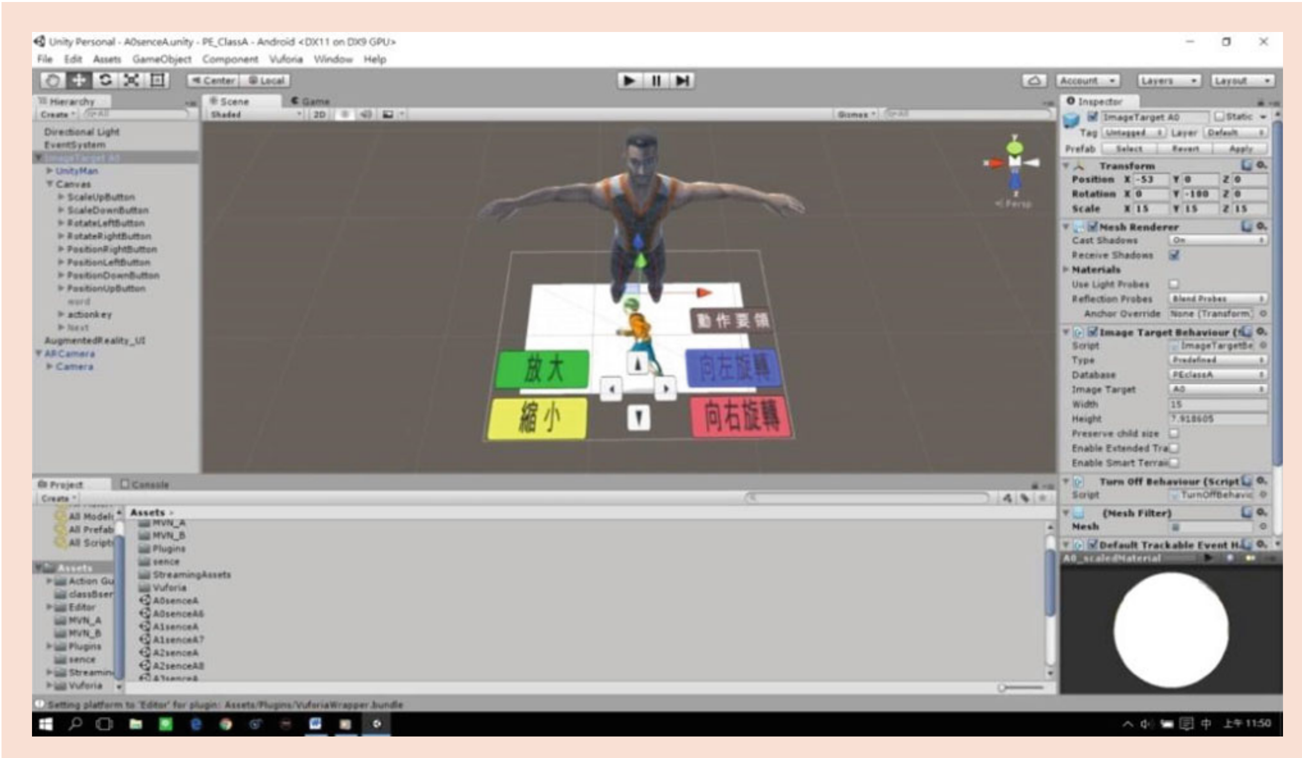


Fig. 14.2 Development interface in Unity 3D

point a mobile device's (e.g., a smartphone) camera at a motor skill picture on a textbook after AR-PEclass is activated, the mobile device screen displays a 3D dynamic model and operational interface, as illustrated in Fig. 14.3. In this interface, four functions are available for users to control a 3D human character model: zooming in and out (changing the model size), rotating left and right (adjusting the viewing angle of the model), arrow keys (adjusting the position of the model on the screen), and movement essentials (providing learners with the relevant operational description of a target motor skill) (Fig. 14.3). When using AR-PEclass for the first time, users must follow instructions provided by the instructor, observe the movement changes exhibited by the 3D model, and complete a worksheet to finish the learning of all motor skills described in the textbook.

Each function of AR-PEclass aims to achieve a specific purpose of learning. For example, the system helps improve learning outcomes by strengthening users' memory and providing them guidance. The aforementioned system also solves the viewing angle problem to improve the effectiveness of motor skill learning. These purposes constitute three learning goals, namely, improvements in learning outcomes, the effectiveness of motor skill learning, and learning motivation. As displayed in Fig. 14.4, learners can engage in visualized learning of a motor skill by controlling the 3D model. The zoom in–zoom out function allows learners to observe locally and in detail how a movement is performed, and the movement essentials strengthen learners' memory of the acquired skills.

14.5 Research Methods and Verification

A quasi-experimental control group pretest–posttest design was adopted to verify the proposed AR-PEclass system. In the conducted experiment, the experimental group learned about motor skills by using textbooks and AR-PEclass, whereas the control group learned about motor skills by watching motor skill instructional videos. The students of both groups were guided by an instructor to read textbooks, use AR-PEclass, or watch instructional videos. Figures 14.5 and 14.6 present photos of the learning activities.

Two seventh-grade classes of students were recruited as the participants of this study. The control and experimental groups comprised 25 and 27 participants, respectively. Student assignment to both classes had been conducted in a manner that the academic performance of students in each class reflected a normal-curve distribution. Moreover, all the students had never previously received professional motor skill training. Instructions on two activities, namely, basic running drills and Mach drills, were provided to both groups.

Fig. 14.3 Operating interface and function description of AR-PEclass [51]

These activities were organized separately to verify the relationship between course difficulty and the learning effectiveness contributed by the use of AR.

The students of the experimental group could scan a picture of a movement from a textbook by using their mobile devices to activate the 3D virtual character of AR-PEclass and engage in interactive learning. The students of the control group were guided by their instructor to read the textbook and watch instructional videos separately. Two experimental activities, namely, the basic running drills and Mach drills, were conducted in both groups of students. Having identical teaching procedures, the two activities were different in the teaching content. Specifically, Mach drills involve dividing running motions into several smaller movements to standardize the running motions for practice. Mach drills are more difficult to execute than basic running movements are because of the higher requirements of movement pace control and smoothness in Mach drills. Prior to the experiment, the students of both groups were introduced to the activity content and learning objectives, during which they also completed a 10-min achievement test on running knowledge (i.e., a pretest). The students of the experimental group received training on mobile device operations to ensure that the experiment would not be interfered with by technical issues. The activities each comprised two 52-min sessions (90 min in total for each activity) and were conducted each over 1 week. In each session, the instructor guided students to learn for 20 min, after which the students engaged in independent practice.

A posttest on running knowledge acquisition and a learning motivation survey were conducted after the experiment, with each lasting 10 min. Subsequently, the motor skills of the two groups were compared. Specifically, three professional instructors graded the motor skills of each participant to determine the learning outcomes. Grading was performed according to the participant's movement accuracy, movement smoothness, and movement integrity.

The pretest and posttest were conducted in a written form, aiming to determine whether the two adopted teaching models (i.e., AR-assisted learning and video learning) improved the students' performance in developing cognitive knowledge of motor skills, which determined the learning outcomes. One pretest and one posttest were conducted for the two activities each. Each test contained 15 items, and 1 point was provided for each correct response. The test content was verified through expert validity, in which the PE instructor who taught the class was asked to complete the test in advance for verifying the test items. A motor skill evaluation scale was developed by referring to the Laban movement analysis method [43] and modified according to the motor skill scale proposed by Arnheim, Sinclair, and Calder [52]. The finalized motor skill evaluation scale used in this study examines the accuracy, smoothness, and integrity of movements related to basic running drills (i.e., walking knee hug, quad walk, walking leg kick, lunge walk, and butt kick) and Mach drills (A walk, B walk, A skip, B skip, three quick steps with one high knee, and three quick steps with one foreleg reach). Each item was rated by the three instructors. Raw scores were standardized into T scores of 1–5, with a maximum of 15 points for the scale. A learning motivation scale, which was also verified through expert validity by the PE teacher, was adapted from the scale of Keller [53], which examines motivation in the dimensions of attention, relevance, confidence, and satisfaction. Each dimension of the modified learning motivation scale comprised six items; thus, this scale comprised a total of 24 items. The Cronbach's alpha values of attention, relevance, confidence, and satisfaction were 0.830, 0.869, 0.846, and 0.862, respectively.

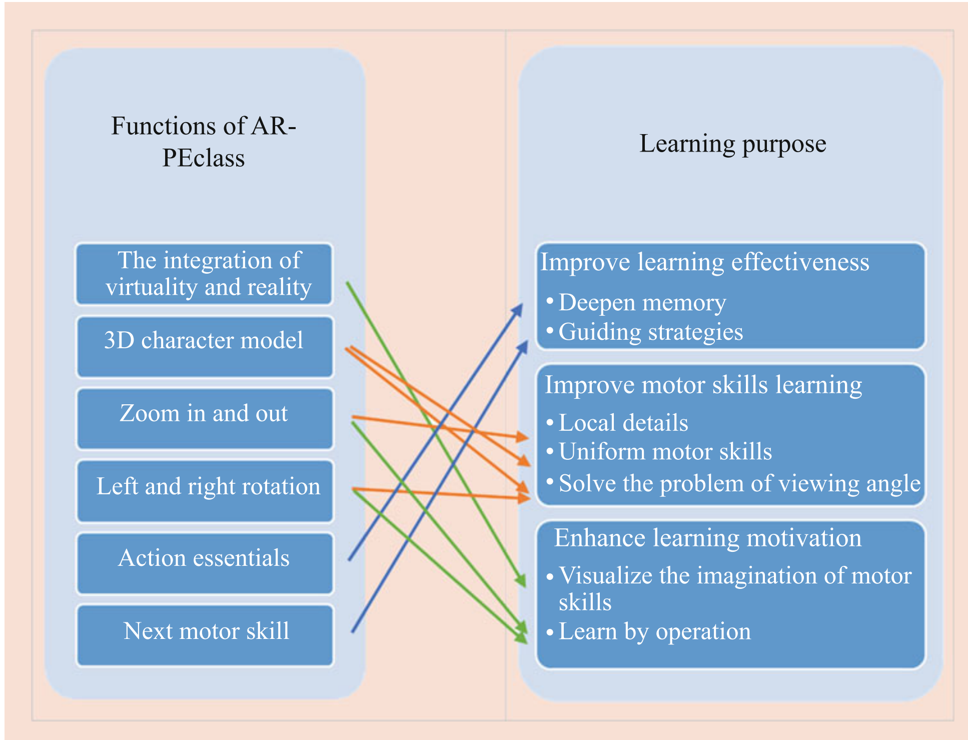


Fig. 14.4 The learning objectives and the functions of AR-PEclass [51]

14.6 Learning Outcomes of AR-Assisted Learning

Table 14.1 presents the pretest and posttest results. A paired-sample *t*-test was conducted on the total pretest and posttest scores of the experimental and control groups. Significant differences were observed between the total pretest and posttest scores of the experimental group ($t = -7.54$, $p = 0.001 < 0.05$) and those of the control group ($t = -3.84$, $p = 0.001 < 0.05$). An analysis of covariance also indicated a significant difference in the scores of the two groups ($F = 15.65$, $p = 0.000 < 0.05$) after the effect of prior knowledge on learning outcomes was excluded. To determine the learning improvements achieved in the two groups, an independent-sample *t*-test was conducted. The results indicated a significant difference in the learning improvements of the two groups ($t = -2.72$, $p = 0.009 < 0.05$). The experimental group exhibited greater learning improvement than the control group did.

The two groups spent different lengths of time in operating the equipment. The students of the experimental group had to familiarize themselves with AR-PEclass (e.g., the button functions and operating procedure), device operation, and teaching model. By contrast, the students of the control group only had to watch specific videos as instructed. Therefore, initially, the experimental procedure of the control group was smoother than that of the experimental group, which undermined the experimental group's performance. However, after the students of the experimental group familiarized themselves with the system operation, the instances of operational mistakes and equipment debugging considerably decreased. They also became better at using the device and buttons.

According to the results presented in Table 14.1, the use of AR and instructional videos significantly affected learning outcomes in motor skill instruction. These outcomes are reflected in the changes between the pretest and posttest scores. The two groups exhibited similar pretest scores; however, the experimental group (i.e., those who engaged in AR-assisted learning) exhibited significantly higher posttest scores than the control group did (i.e., those who engaged in video learning). The experimental group's performance was initially undermined by the less unsmooth teaching procedure and low student familiarity with the system and equipment. Thus, learning smoothness and familiarity with software operation will be crucial factors that affect learning outcomes if AR is to be incorporated into PE teaching.

The user experience for AR-PEclass indicated that the experimental group did not outperform the control group in attaining goals or learning movements with low difficulty [51]. This result was obtained because the advantages of highly complex AR technology are difficult to leverage when learning easy things [54, 55]. However, after the participants mastered system operation, their learning outcomes improved considerably for difficult learning goals and motor skills [51]. This finding indicates that AR-assisted learning is preferable when the learning difficulty is high. The

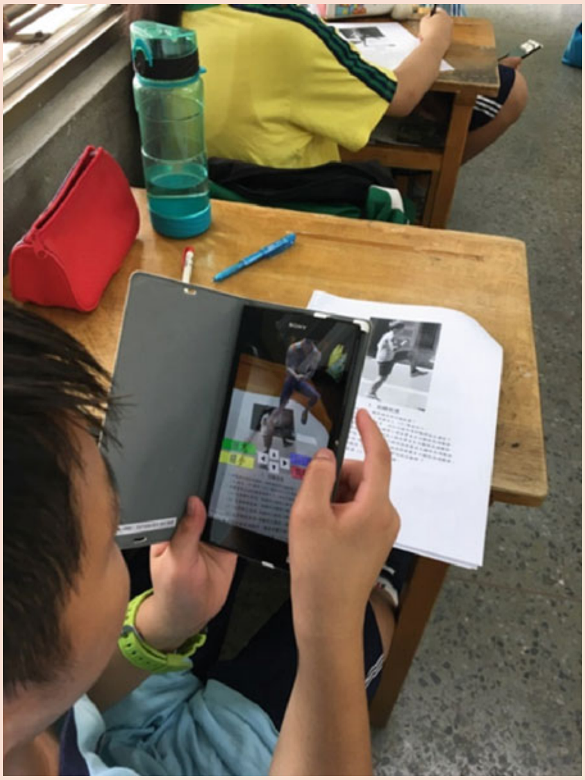


Fig. 14.5 Students of the experimental group using AR-PEclass for learning

Fig. 14.6 Students of the control group watching a video to learn

aforementioned finding is also in agreement with that of Safadel and White [55].

Interviews conducted with the participants indicated that when the target motor skill was difficult to learn, they could fully perceive the benefits provided by the AR system. The system enabled them to enhance their understanding of the skills. A 3D display of teaching materials provides more assistance to learners than conventional lectures or video learning models do with only one viewing angle. This finding is consistent with that of Zhier [6], who concluded that 3D visual teaching materials strengthen learners' memory and cognitive function by providing various viewing angles. This benefit of AR and 3D displays was also indicated by the instructors' assessments of the motor skill performance. Students who received AR-assisted instruction obtained higher instructor scores than those who received conventional instruction [51]. Moreover, the curiosity aroused by the manipulation of 3D models in the teaching materials enhanced the learning motivation, memory, and learning retention of the students [56].

Table 14.1 Means and standard deviations of the pretest and posttest scores

	Group	Number of participants	Mean	Standard deviation
Pretest total score	Control group	25	8.36	2.78
	Experimental group	27	8.37	2.63
Posttest total score	Control group	25	10.36	2.41
	Experimental group	27	12.40	2.83

14.7 Motor Skill Acquisition by Using AR-PEclass

After the learning of motor skills and completion of the achievement test, group practices and then motor skill assessment were conducted to explore the effects of different teaching materials on motor skill learning. Specifically, three instructors were invited to assess the students' motor skills according to the accuracy, smoothness, and integrity of movements. Each motor skill item was rated from 1 (*poor*) to 5 (*excellent*), and a maximum total score of 15 points could be obtained. The means of the total scores provided by the three instructors were determined, and the data collected for the control and experimental groups were then subjected to an independent-sample t-test. Table 14.2 lists the results of this test.

The results indicated significant differences between the performance of the experimental and control groups on 11 motor skills. The experimental group outperformed the control group on all the aforementioned motor skills. The proposed AR-PEclass integrated textbooks with 3D dynamic models, which enabled learners to see and manipulate the 3D models freely. Therefore, the participants who used AR-PEclass exhibited superior learning outcomes to those who

Table 14.2 Independent-sample t-test for the motor skill assessment of the experimental and control groups

Motor skill	Group	Number of participants	Mean	Standard deviation	t	p
Walking knee hug	Control group	25	11.84	1.60	−4.15	0.000***
	Experimental group	27	13.34	0.86		
Quad walk	Control group	25	10.65	1.92	−3.94	0.000***
	Experimental group	27	12.37	1.05		
Walking leg kick	Control group	25	10.49	2.04	−3.08	0.003**
	Experimental group	27	12.14	1.81		
Lunge walk	Control group	25	9.80	1.82	−3.68	0.001**
	Experimental group	27	11.62	1.75		
Butt kick	Control group	25	11.09	1.70	−4.97	0.000***
	Experimental group	27	13.24	1.40		
A walk	Control group	25	9.46	1.47	−8.46	0.000***
	Experimental group	27	12.34	0.88		
B walk	Control group	25	9.36	3.38	−4.78	0.000***
	Experimental group	27	12.74	1.04		
A skip	Control group	25	8.60	2.04	−6.02	0.000***
	Experimental group	27	11.54	1.39		
B skip	Control group	25	8.25	2.05	−5.80	0.000***
	Experimental group	27	10.95	1.12		
Three quick steps with one high knee	Control group	25	6.82	1.75	−8.24	0.000***
	Experimental group	27	10.41	1.38		
Three quick steps with one foreleg reach	Control group	25	6.72	1.95	−7.07	0.000***
	Experimental group	27	9.79	0.65		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

used video-based materials in terms of understanding, performing, and mastering motor skills. This difference in the learning outcomes between the two groups was also reflected in their achievement test scores. In terms of movement proficiency and familiarity, the experimental group obtained higher total scores than the control group did.

During interviews, the students stated that the Mach drills were more difficult to master and understand than the basic running drills were. This opinion was also reflected in the motor skill ratings by instructors. Lower scores were obtained for the final two Mach drill movements, namely, three quick steps with one high knee and three quick steps with one foreleg reach, than for the other movements in both groups. However, the differences between the scores for the aforementioned two movements and those of the other movements were smaller in the experimental group than in the control group. The results indicated that AR-PEclass helped students strengthen their spatial and operational cognition by integrating 2D and 3D virtual objects with real objects for the instruction of difficult and rhythmic movements. By interacting with 3D objects in AR-PEclass, students improved their operational and self-learning competencies. The system achieved virtuality–reality integration by displaying the text and graphics of printed materials in 3D dynamic models; thus, the proposed system successfully integrated printed and digital teaching materials.

The aforementioned results proved that AR-PEclass can facilitate the creation of innovative motor skill teaching models and attractive course content. In practice, the developed system can supplement the deficiencies in current PE teaching, such as the problems related to viewing angles and student ability to learn independently, as indicated by the experimental results for learners' system operation and motor skill practice. AR-assisted learning helps instructors to create innovative teaching models and learners to improve learning outcomes.

14.8 Learning Motivation Stimulated by AR-PEclass

The learning motivation of the students in both groups was analyzed in terms of four factors, namely, the levels of attention, relevance, confidence, and satisfaction (ARCS). A survey on motivation was conducted after the experiment by using the ARCS motivation scale of Keller [53]. Participants who missed any questions or provided contradictory answers to reverse-coded items were excluded to increase the analysis

Table 14.3 Results of the independent-sample *t*-test on the learning motivation of the experimental and control groups

Group	Number of participants	Mean	Standard deviation	t	p
Control group	24	3.88	0.666	−2.65	0.011*
Experimental group	26	4.33	0.520		

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

accuracy. Following data exclusion, 24 and 26 responses were collected from the control and experimental groups, respectively. An independent-sample *t*-test was then conducted, and the results of this test are listed in Table 14.3. The results revealed a significant difference in the mean ARCS scores between the two groups ($t = -2.65$; $p = 0.011 < 0.05$).

Significant differences were observed between the two groups in the dimensions of attention ($t = -3.27$, $p = 0.018 < 0.05$), relevance ($t = -2.11$, $p = 0.04 < 0.05$), and confidence ($t = -3.08$, $p = 0.003 < 0.05$); however, the difference between the two groups in the dimension of satisfaction was not significant ($t = -1.25$, $p = 0.218 > 0.05$). The test results indicated that the experimental group had significantly higher learning motivation than the control group. The mean scores of the participants also verified that AR-based materials outperformed video-based ones for IT-based PE teaching. Specifically, AR-based materials were more effective in stimulating learners' interest and generating positive learning attitudes. Learners also praised the novel AR-assisted teaching model. The ARCS survey confirmed the existence of significant differences between both groups in the attention, relevance, and confidence dimensions of learning motivation. This finding is consistent with that of previous studies; however, the nonsignificant results for the satisfaction dimension contradict the findings of previous studies.

On the basis of the ARCS learning motivation model proposed by Keller [57], we determined that AR-assisted instruction can better meet learners' needs and goals in motor skill learning than VAI can. This finding is obtained because the designed AR-PEclass combines printed textbooks with digital 3D models through its virtuality–reality integration function to facilitate better learning. Furthermore, the developed system's incorporation of digital tools can facilitate the creation of intriguing course content and effective learning. For example, the system functions of multiple viewing angles, 3D character manipulation, and innovative learning can be used to generate positive learning attitudes. Due to the aforementioned reasons, significant differences were observed between the two groups in the attention dimension.

Regarding the dimension of relevance, students engaged in motor skill practices after completing the course and achievement test. These practices aimed to enable students to understand the association between the teaching activities and motor skill learning. In addition, the practices helped the students to stay focused, identify differences between actual movements and those illustrated in teaching materials, and establish self-confidence in motor skill practices. AR-PEclass incorporated printed textbook content into a virtual space and displayed teaching materials by using 3D dynamic models. The AR system allowed students to read textbooks while manipulating the 3D models, adjust the viewing angle, and use all functions displayed in the system interface. Therefore, AR-assisted instruction could stimulate the students' interest and willingness to make efforts to engage in independent learning and achieve learning objectives.

With regard to the dimension of confidence, the students who received AR-assisted instruction grasped motor skills quicker and easier than those receiving VAI because the AR system allowed users to change the viewing angle as well as the position and size of the 3D model. In addition, the AR system, integrating printed and digital materials, facilitated quicker and more detailed understanding compared with the video-based materials. During the interviews, a participant from the experimental group mentioned that "[AR-PEclass] allows us to understand what is a good movement and find out important details." When using the AR system incorporating 3D display to learn about motor skills, the participants exhibited enhanced confidence and completed the learning goals of each course. This result could also be observed in the two groups' different motor skill assessment scores rated by instructors.

The students in both groups had never previously participated in IT-based PE courses. These courses are different from conventional PE courses and allow students to use mobile devices for learning. Therefore, both groups were satisfied with the provided courses. No significant difference was observed in the satisfaction dimension of learning motivation between the two groups. Keller [57] stated that a match between instructional and learning environments can strengthen perceived satisfaction. The AR teaching materials used in this study integrated virtual and real objects by using 2D or 3D techniques; thus, the students' sensory perception of and interactivity with the real world were strengthened [58, 59]. Compared with instructional videos, the combination of a virtual environment with real objects was more appealing to students and more effective in improving learning outcome. Consequently, the experimental group outscored the control group in the satisfaction dimension of learning motivation.

Due to the aforementioned reasons, the students engaging in AR-assisted learning had a considerably higher learning motivation than those who engaged in video learning.

Students who received AR-assisted instruction obtained higher scores in the dimensions of attention, relevance, and confidence than those who received VAI. This result is consistent with those of previous studies [60–62]. In summary, compared with the students who received VAI, those who received AR-assisted instruction had higher motivation in the attention, relevance, and confidence dimensions for learning about motor skills.

14.9 Conclusion

VAI has been widely used in the demonstrations and description of motor skills in PE; however, this teaching model has two disadvantages. First, VAI does not offer interactive experience. Consequently, students cannot view the demonstration and its details from various angles. Moreover, VAI fails to integrate the knowledge of motor skills with actual practice of the skills. AR-based 3D teaching materials help overcome the aforementioned drawbacks and improve learning outcomes. An AR-based model supports learners in acquiring PE motor skills by enabling them to read printed textbooks and manipulate a 3D human character model simultaneously.

The experimental results obtained for two activities (i.e., basic running drills and Mach drills) indicated that the students who used the proposed AR-PEclass outperformed their counterparts who received VAI in terms of learning outcomes, motor skill acquisition, and learning motivation, particularly for the learning of highly difficult movements [51]. The experimental results suggested that in addition to using demonstration materials, learners must be allowed to explore and practice. Allowing students to explore and practice freely is a favorable criterion for supplementary teaching materials. The AR system developed in this study satisfies the aforementioned criterion because it offers interactive experience to learners, enables them to observe a movement from various angles, and provides a kinesthetic immersive experience to them.

During the AR-based PE teaching, students faced difficulties in scanning images from textbooks with their phones. Specifically, the captured images exhibited problems, such as glare, excessively dark shadows, and blurriness, which caused difficulties in image recognition with phones. This issue should be solved by improving the image recognition algorithm of the developed system, avoiding scanning in dim light, and avoiding scanning at locations with excessive interference. AR-assisted learning with 3D display techniques is achieved by overlaying a 3D human character model on a real teaching material. This learning method can only be applied to the individual training of motor skills. Because a single 3D human character model may be insufficient for training for sports competitions, the overlaying information in AR must be able to facilitate scenario-based learning. Accordingly, virtual reality can be integrated with AR to create mixed reality. The integration of these two technologies may fulfil scenario-based learning and may constitute a promising direction for the development and research of training aids for competitive sports.

References

1. Roche, L., Gal-Petitfaux, N.: Using video in physical education teacher education. In Oral presentation at the AIESEP World Congress (2014)
2. Roche, L., Gal-Petitfaux, N.: Using 360° video in physical education teacher education. In: Society for Information Technology & Teacher Education International Conference, pp. 3420–3425. Association for the Advancement of Computing in Education (AACE) (2017)
3. Peng, W., Crouse, J.C., Lin, J.H.: Using active video games for physical activity promotion: a systematic review of the current state of research. Health Educ. Behav. **40**(2), 171–192 (2013)
4. Tabbers, H.K., Martens, R.L., Merriënboer, J.J.: Multimedia instructions and cognitive load theory: effects of modality and cueing. Br. J. Educ. Psychol. **74**(1), 71–81 (2004)
5. Soltani, P., Figueiredo, P., Fernandes, R.J., Vilas-Boas, J.P.: Do player performance, real sport experience, and gender affect movement patterns during equivalent exergame? Comput. Hum. Behav. **63**, 1–8 (2016)
6. Zhier, L.: Research on the physical education teaching system based on 3D human motion capture. Int. J. Signal Process. Image Process. Pattern Recognit. **9**(12), 61–70 (2016)
7. Sears, J.V., Edgington, W.D., Hynes, J.W.: The effect of professional development on physical education instruction in middle schools: thirty middle school physical education teachers in the southwestern United States participate in a research study of professional development opportunities and deficiencies in a field not driven by high-stakes testing. Middle Sch. J. **45**(5), 25–31 (2014)
8. Roth, K., Zittel, L., Pyfer, J., Auxter, D.: Principles and Methods of Adapted Physical Education & Recreation. Jones & Bartlett Publishers (2016)
9. Miller, P.: Learning Styles: The Multimedia of the Mind. Research Report (2001)
10. Willis, M., Hodson, V.K.: Discover Your Child's Learning Style. Prima Pub (1999)
11. Zhang, J., Ogan, A., Liu, T.C., Sung, Y.T., Chang, K.E.: The influence of using augmented reality on textbook support for learners of different learning styles. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 107–114. IEEE (2016)
12. Tüzün, H., Özdinç, F.: The effects of 3D multi-user virtual environments on freshmen university students' conceptual and spatial learning and presence in departmental orientation. Comput. Educ. **94**, 228–240 (2016)
13. Klopfer, E., Squire, K.: Environmental detectives – the development of an augmented reality platform for environmental simulations. Educ. Technol. Res. Dev. **56**(2), 203–228 (2008)
14. Blum, L., Wetzel, R., McCall, R., Oppermann, L., Broll, W.: The final TimeWarp: using form and content to support player experience and presence when designing location-aware mobile augmented reality games. In: Proceedings of the Designing Interactive Systems Conference, pp. 711–720. ACM (2012)
15. Van Krevelen, D.W.F., Poelman, R.: A survey of augmented reality technologies, applications and limitations. Inte. J. Virtual Real. **9**(2), 1 (2010)

16. McCall, R., Wetzel, R., Löschner, J., Braun, A.-K.: Using presence to evaluate an augmented reality location aware game. Pers. Ubiquit. Comput. **15**(1), 25–35 (2010)
17. Wang, X., Chen, R.: An experimental study on collaborative effectiveness of augmented reality potentials in urban design. CoDesign. **5**(4), 229–244 (2009)
18. Wang, K.-H., Chen, L.-C., Chu, P.-Y., Cheng, Y.-M.: A study on the design of augmented reality user interfaces for mobile learning systems in heritage temples. In: International Conference on Virtual and Mixed Reality, pp. 282–290. Springer, Berlin/Heidelberg (2009)
19. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. J. Sci. Educ. Technol. **18**(1), 7–22 (2009)
20. Lee, K.F., Chen, Y.L., Hsieh, H.C., Chin, K.Y.: Application of intuitive mixed reality interactive system to museum guide activity. In: 2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 257–258. IEEE (2017)
21. Westin, J., Foka, A., Chapman, A.: Humanising places: exposing histories of the disenfranchised through augmented reality. Int. J. Herit. Stud. **24**(3), 283–286 (2018)
22. Ghouaiel, N., Garbaya, S., Cieutat, J.M., Jessel, J.P.: Mobile augmented reality in museums: towards enhancing visitor's learning experience. Int. J. Virtual Real. **17**(1) (2017)
23. Chang, Y.L., Hou, H.T., Pan, C.Y., Sung, Y.T., Chang, K.E.: Apply an augmented reality in a mobile guidance to increase sense of place for heritage places. J. Educ. Technol. Soc. **18**(2), 166–178 (2015)
24. Chiu, C.C., Lee, L.C.: System satisfaction survey for the App to integrate search and augmented reality with geographical information technology. Microsyst. Technol., 1–23 (2017)
25. Gottlieb, O.: Time travel, labour history, and the null curriculum: new design knowledge for mobile augmented reality history games. Int. J. Heritage Stud., 1–13 (2017)
26. Tom Dieck, M.C., Jung, T.: A theoretical model of mobile augmented reality acceptance in urban heritage tourism. Curr. Issue Tour. **21**(2), 154–174 (2018)
27. Pombo, L., Marques, M.M., Carlos, V., Guerra, C., Lucas, M., Loureiro, M.J.: Augmented reality and mobile learning in a smart urban park: pupils' perceptions of the EduPARK game. In: Conference on Smart Learning Ecosystems and Regional Development, pp. 90–100. Springer, Cham (2017)
28. Kamarainen, A.M., Metcalf, S., Grotzer, T.A., Brimhall, C., Dede, C.: Atom tracker: designing a mobile augmented reality experience to support instruction about cycles and conservation of matter in outdoor learning environments. Int. J. Des. Learn. **7**(2) (2016)
29. Chu, H.C., Sung, Y.H.: A context-aware progressive inquiry-based augmented reality system to improving students' investigation learning abilities for high school geography courses. In: 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 353–356. IEEE (2016)
30. Dinis, F.M., Guimarães, A.S., Carvalho, B.R., Martins, J.P.P.: Virtual and augmented reality game-based applications to civil engineering education. In: Global Engineering Education Conference (EDUCON), 2017 IEEE, pp. 1683–1688. IEEE (2017)
31. Park, C.S., Le, Q.T., Pedro, A., Lim, C.R.: Interactive building anatomy modeling for experiential building construction education. J. Prof. Issues Eng. Educ. Pract. **142**(3), 04015019 (2015)
32. Martín-Gutiérrez, J., Fabiani, P., Benesova, W., Meneses, M.D., Mora, C.E.: Augmented reality to promote collaborative and autonomous learning in higher education. Comput. Hum. Behav. **51**, 752–761 (2015)
33. Zhang, J., Sung, Y.T., Hou, H.T., Chang, K.E.: The development and evaluation of an augmented reality-based armillary sphere for astronomical observation instruction. Comput. Educ. **73**, 178–188 (2014)
34. Fleck, S., Hachet, M., Bastien, J.M.: Marker-based augmented reality: instructional-design to improve children interactions with astronomical concepts. In: Proceedings of the 14th International Conference on Interaction Design and Children, pp. 21–28. ACM (2015)
35. Zhang, J., Sung, Y.T., Chang, K.E.: Using a mobile digital armillary sphere (MDAS) in astronomical observation for primary school students. In: Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, pp. 2632–2641 (2011)
36. Chang, K.E., Chang, C.T., Hou, H.T., Sung, Y.T., Chao, H.L., Lee, C.M.: Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. Comput. Educ. **71**, 185–197 (2014)
37. Wu, H.K., Lee, S.W.Y., Chang, H.Y., Liang, J.C.: Current status, opportunities and challenges of augmented reality in education. Comput. Educ. **62**, 41–49 (2013)
38. Orman, E.K., Price, H.E., Russell, C.R.: Feasibility of using an augmented immersive virtual reality learning environment to enhance music conducting skills. J. Music Teach. Educ., 1057083717697962 (2017)
39. Pang, Y.: Improving hybrid learning of physical education by video review. In: International Conference on Web-Based Learning, pp. 230–239. Springer, Berlin/Heidelberg (2010)
40. Fitts, P. M., & Posner, M. I. (1967). Human Performance
41. Magill, R.A., Hall, K.G.: A review of the contextual interference effect in motor skill acquisition. Hum. Mov. Sci. **9**(3–5), 241–289 (1990)
42. Stewart, M., Collin, C.: An interactive video tape program designed for preservice teacher education. Technol. Teach. Educ. Annu., 245–251 (1992)
43. Von Laban, R., Lange, R.: Laban's Principles of Dance and Movement Notation. Princeton Book Co Pub (1975)
44. Fogtmann, M.H., Fritsch, J., Kortbek, K.J.: Kinesthetic interaction – revealing the bodily potential in interaction design. In: OZCHI '08: Conference of the Computerhuman Interaction Special Interest Group (CHISIG) of Australia on Computer-Human Interaction. ACM, Cairns (2008)
45. Mosston, M.: Tug-o-war, no more: meeting teaching learning objectives using the spectrum of teaching styles. J. Phys. Educ. Recreat. Dance, 27–31 (1992)
46. Huang, C.H., Chin, S.L., Hsin, L.H., Hung, J.C., Yu, Y.P.: A web-based e-learning platform for physical education. J. Networks. **6**(5), 721 (2011)
47. Chen, Y.C., Hwang, R.H., Wang, C.Y.: Development and evaluation of a Web 2.0 annotation system as a learning tool in an e-learning environment. Comput. Educ. **58**(4), 1094–1105 (2012)
48. Zhao, H., Sullivan, K.P., Mellenius, I.: Participation, interaction and social presence: an exploratory study of collaboration in online peer review groups. Br. J. Educ. Technol. **45**(5), 807–819 (2014)
49. Papastergiou, M., Gerodimos, V.: Can learning of basketball be enhanced through a web-based multimedia course? An experimental study. Educ. Inf. Technol. **18**(3), 459–478 (2013)
50. Zou, J., Liu, Q., Yang, Z.: Development of a Moodle course for schoolchildren's table tennis learning based on Competence Motivation Theory: its effectiveness in comparison to traditional training method. Comput. Educ. **59**(2), 294–303 (2012)
51. Chang, K.E., Zhang, J., Huang, Y.S., Liu, T.C., Sung, Y.T.: Applying augmented reality in physical education on motor skills learning. Interact. Learn. Environ., 1–13 (2019)
52. Arnheim, D.D., Sinclair, W.A., Calder, J.E.: The clumsy child: a program of motor therapy. Aust. J. Spec. Educ. **3**(2), 26–26 (1979)
53. Keller, J.M.: Motivational design of instruction. Instr. Des. Theories Mod. **1**, 383–434 (1983)
54. Ferrara, F., Ponticorvo, M., Di Ferdinando, A., Miglino, O.: Tangible interfaces for cognitive assessment and training in children:

LogicART. In: Smart Education and e-Learning 2016, pp. 329–338. Springer (2016)
55. Safadel, P., White, D.: A comparative analysis of augmented reality and two-dimensional using molecular modeling on student learning. In: Society for Information Technology & Teacher Education International Conference, pp. 1774–1776. Association for the Advancement of Computing in Education (AACE) (2017)
56. Fujimoto, Y., Yamamoto, G., Kato, H., Miyazaki, J.: Relation between location of information displayed by augmented reality and user's memorization. In: Proceedings of the 3rd Augmented Human International Conference, p. 7. ACM (2012)
57. Keller, J.M.: Using the ARCS motivational process in computer-based instruction and distance education. New Dir. Teach. Learn. **1999**(78), 37–47 (1999)
58. Zak, E.: Do you believe in magic? Exploring the conceptualization of augmented reality and its implications for the user in the field of library and information science. Inf. Technol. Libr. **33**(4), 23–50 (2014)
59. Wasko, C.: What teachers need to know about augmented reality enhanced learning environments. TechTrends. **57**(4), 17–21 (2013)
60. Chiang, T.H., Yang, S.J., Hwang, G.J.: An augmented reality-based mobile learning system to improve students' learning achievements and motivations in natural science inquiry activities. J. Educ. Technol. Soc. **17**(4), 352 (2014)
61. Estapa, A., Nadolny, L.: The effect of an augmented reality enhanced mathematics lesson on student achievement and motivation. J. STEM Educ. Innov. Res. **16**(3), 40 (2015)
62. Riera, A.S., Redondo, E., Fonseca, D.: Geo-located teaching using handheld augmented reality: good practices to improve the motivation and qualifications of architecture students. Univ. Access Inf. Soc. **14**(3), 363–374 (2015)

Jia Zhang 's research interests span most aspects of augmented reality and mobile learning. This includes image recognition, real-time AR-based learning, learning behavior prediction in AR environment, as well as augmented learning environment developing. His recent emphasis is using image recognition technology to enhance the augmented reality application effect of education.

Yang-Sheng Huang is a graduate student at National Taiwan Normal University's Graduate Institute of Information and Computer Education. He has won medals in the National Intercollegiate Athletic Games and the National Games of Taiwan.

15 Potentiating Learning Through Augmented Reality and Serious Games

J. M. Cerqueira, B. Cleto, J. M. Moura, C. Sylla, and L. Ferreira

Contents

15.1 Introduction . . . 369

15.2 Augmented Reality in Pedagogical Contexts . . . 371

15.3 Serious Games in Pedagogical Contexts . . . 372
15.3.1 Educational Learning Approaches . . . 373
15.3.2 Design of Serious Games for Education . . . 374
15.3.3 Types of Serious Games Models . . . 374
15.3.4 Serious Games and User Experiences . . . 374

15.4 Serious Games and Augmented Reality . . . 375

15.5 Case Studies . . . 376
15.5.1 Visualizing Platonic Solids in Augmented Reality . . . 376
15.5.2 CodeCubes . . . 379
15.5.3 FootMath . . . 381

15.6 Closing Remarks . . . 386

15.7 Conclusions . . . 386

References . . . 387

Abstract

A significant progress in technological support aroused a general interest in areas such as data visualization, augmented and virtual reality, and artificial intelligence. The innovative capabilities of these technologies increased the potential and relevance of data visualization and interaction services. Traditional teaching and learning methods look insufficient in a context where digitalization invades processes and tools. A strong claim is made for aligning those methods with such technological developments and thus allows students to acquire the skills to successfully integrate the emergent information society. Given the widespread of mobile devices, and the increasing role of computer games in education, their combined use has the potential to play a central role in responding to these demands. This chapter aims at exploring the integration of serious games with virtual environments and technologies as a complement to facilitate and enhance learning. More specifically, we present and discuss the motivation, design, and development process of three tools that use augmented reality in combination with a serious game to teach (i) mathematics, (ii) to explore the platonic solids, (iii) and to teach coding. Further, three pilot user studies are described and discussed and confirm the potential of these tools as powerful new teaching/learning tools for education.

Keywords

Augmented reality · Education · Learning · Mathematics · Playing · Serious games · Virtual reality

15.1 Introduction

Among the various emerging technologies associated with virtual environments, the augmented reality (AR) technology has been gaining the interest of different areas of knowledge offering a wide diversity of use and applicability.

In recent years, research and use of AR technology has increased significantly. Being an innovative and promising technology [1], it offers a new avenue for combining virtuality and reality [2].

Given the technological advancements in the area, today it is possible to extend real environments by adding (non-real) digital information (augmented reality) to it, allowing users to experience a set of new possibilities. However, the limits between the virtual and the augmented reality are very tenuous, often resulting in the joint exploration of

J. M. Cerqueira (✉) · B. Cleto (✉)
Polytechnic Institute of Cávado and Ave, Barcelos, Portugal
e-mail: josecerqueira@aeaf.edu.pt; bgcleto@esmad.ipp.pt

J. M. Moura (✉) · L. Ferreira (✉)
2Ai – School of Technology, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal
e-mail: jmoura@ipca.pt; lufer@ipca.pt

C. Sylla (✉)
CIEC – ITI/LARSyS, University of Minho, Braga, Portugal
e-mail: cristina.sylla@ie.uminho.pt

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_15

both worlds, which results in mixed reality environments. Milgram and Kishino [3] highlighted this division with the *Reality-Virtuality Continuum* spectrum, describing the various "states" of reality and the associated degree of immersion and interactivity (Fig. 15.1).

However, this continuum spectrum does not cover all the possibilities. According to Steve Mann [4], "*(. . .) mixed reality covers a continuum between graphics enhanced by video and video enhanced by graphics . . . however, there are important categories of visual information processors that do not fit within this taxonomy or continuum. (. . .)*." Building on this definition, the concept of *Mediated Reality* [5] should be considered (Fig. 15.2).

According to the *Mediated Reality* approach, which departs from the unmodified reality (*R*), there is a *continuum* toward virtuality (*V*) (horizontal axis) in which the reality is augmented with graphics (*AUG.R*). Simultaneously, the graphics are augmented by reality, creating an augmented virtuality (*AUG.V*). The mediation axis *M* displays the changes, which are visible by moving upward on the mediation axis.

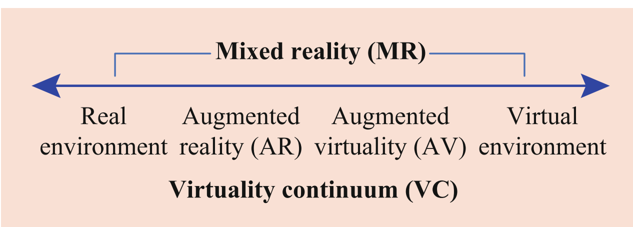


Fig. 15.1 Reality-virtuality continuum model proposed by Milgram and Kishino [3]

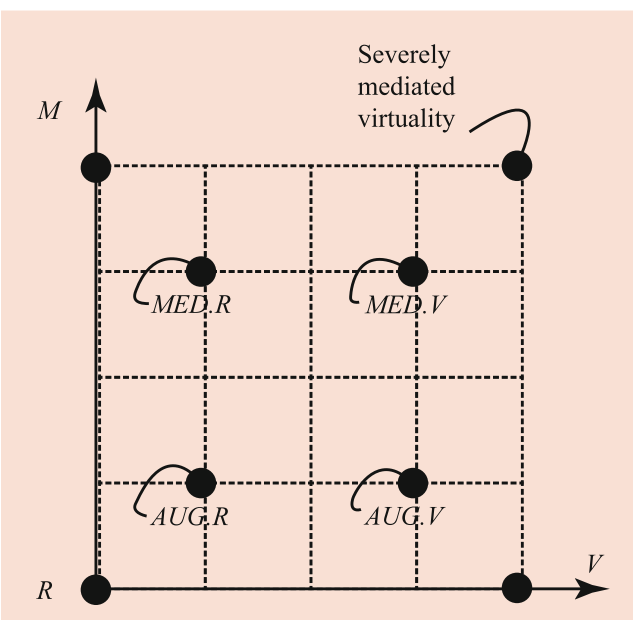


Fig. 15.2 Taxonomy of reality, virtuality, and mediality [5]

Further up on this axis, for example, we can find the Mediated Reality (*MED.R*), the Mediated Virtuality (*MED.V*), or any combination of both. Above and to the right, there are virtual worlds that display a strongly modified version of reality [4].

Independent of all existing perspectives, for better understanding a domain that has several taxonomies, it is relevant to clarify the most relevant concepts: augmented reality, virtual reality, mixed reality, and extended reality.

According to literature, AR is a technology that combines (or mediates) digital (i.e., computer-generated content or content generated by similar electronic equipment) and an existing reality (e.g., physical artifacts), in real time, creating new interaction experiences. As such, AR offers the possibility of using virtual information to reinforce the real environment [6], in real time, tailoring the visualization of the information according to the user's needs [7]. Instead of completely replacing the reality, AR complements it [8], allowing to visualize and interact with the real world without complete immersion.

In parallel with AR, we have also witnessed the continuous growth of virtual reality (VR) technology that can be seen as an artificial simulation (generated by a computer) to recreate (represent) an environment, being it a real or imaginary situation. VR leads to the users' immersion, making them feel that they are experiencing the simulated reality, mainly stimulating their vision and hearing.

When talking about AR, mixed reality (MR) (sometimes also referred as merged reality) and extended reality should be referred as well.

MR is the technology that allows mixing the physical and virtual worlds to create environments (in the spectrum between AR and VR) and forms of visualization where physical and digital objects coexist and can interact. This technology combines computational processing power with human-environment interactivity to ensure that any action in the physical world is reflected in the digital world. In this context, "synthetic" content and real-world content have characteristics that allow them to intertwine and interact with each other (to a certain extent).

Google Glass is an example of an MR device (also called Augmented Reality Smart Glasses), similar to regular glasses. It is a computer that includes an optical screen (a single projection lens) with head support, touch-sensitive, together with a built-in camera and speaker. Although it was a pioneering project, it was conditioned by its low-processing capacity. However, the Google Glass project has been evolving in recent years with artificial intelligence's inclusion in its system.

The extended reality (XR) represents the field that embraces both the real and virtual worlds, that is, AR, VR, and MR technologies. [9] refers to all combined environments that support immersive technologies, whether they are real, virtual, or human-machine interactions generated by com-

Table 15.1 Summary of AR, VR, MR, and XR technologies

Technology	Purpose	References
Augmented reality (AR)	Provides the overlapping of virtual objects with the real world	[8]
Virtual reality (VR)	Provides immersion in an artificial world generated by the computer	[12]
Mixed reality (MR)	Enables interaction, in a real environment, with virtual objects	[13]
Extended reality (XR)	Creates a complete experience targeting all the senses	[9]

puter technology (or wearables). It provides a wide variety and range of virtuality levels, from sensors to immersion. It was considered one of the emerging technologies identified in the 2020 edition of the Horizon Report [10].

According to Accenture [11], the exploration of XR represents an important step in the form of education or training: "*(. . .) the immersive learning is a highly effective way for organizations to deliver formal training while reducing costs (. . .)*."

Table 15.1 summarizes the main characteristics of this technological framework.

Considering the exploration of AR and VR technologies, they cross the most diverse areas of knowledge, seeing, according to the concept of Reality-Virtuality Continuum [3], the two sides of the same coin. There are essentially three ways to explore AR: (i) marker-based AR, the most common, where images (markers) are used in the screening process; (ii) location-based AR, where the location is facilitated due to the existence of various sensors on mobile devices such as the Global Positioning System (GPS), accelerometers, and gyroscopes, among others [14]; and (iii) *markerless AR*, in which the user interacts directly in the real environment.

The growth in the use of AR is partly due to the technological developments of mobile platforms [15], as well as the improvement of the communication infrastructures and services developed on them, namely, the rise of the internet's speed and bandwidth, local networks, and increasingly "ubiquitous" accesses.

AR can be used in multiple types of applications and devices, such as mobile devices with cameras (smartphones, tablets), headphones, glasses, or HDMs (Head-Mounted Displays). Microsoft HoloLens and Magic Leap [1] are examples of HMD-based AR projects. Wearables, such as watches and contact lenses, are another type of device that could be promising targets for this technology in the future.

15.2 Augmented Reality in Pedagogical Contexts

The interest in AR is not restricted to education but extends to many areas of research and application, such as in sciences, engineering, industry, or entertainment [1], [16, 17]. However, the number of available AR-supported mobile applications for many education areas has been growing rapidly and auspiciously.

Considering the education, due either to its potential or to new educational opportunities [15, 18–20], applications that involve AR raise the interest of educators, teachers, and didactics, as educational resources that can be used as added value in the classroom [21], as a complement to more traditional teaching models (in typically aligned rooms) or in more modern models (in flexible learning spaces). Furthermore, AR is also applied in serious games [22], in augmented books [23], and in multiple domain apps, namely, involving georeferencing [24].

The serious game for mathematics presented in [25] includes two mini-games, *Battle Against the Colossus* and *Save the planet!*, and both use AR technology. The HBR AR (*Harvard Business Review Augmented Reality*) application, downloadable from Google Play and the App Store, allows the exploration of several AR experiences, such as described by Michael E. Porter and James E. Heppelmann in the article *Why Every Organization Needs an Augmented Reality Strategy,* on HBR 2017 December Magazine (available on HBR.org).

The same happens with *Pokemon Go*, another example of an AR exploration game, where the entertainment comes from its dynamic georeferenced features.

According to the literature, AR is more recognized in relation to other pedagogical means since it (i) enables a new way of interaction with the real, manipulates virtual objects, and/or allows the observation or experimentation of phenomena that are difficult to observe or experience in the real world [15, 20]. The high level of realism [15, 26], such as the visualization of 3D objects [27] and simulation of dynamic processes that are difficult to observe naturally [28, 29], is an example of improvement; (ii) contributes to free cognitive resources by reducing the users' cognitive load [6, 30]; (iii) contributes to the transformation of the passive to the active learning process, allowing students to interact with the content, be creative, and practice their knowledge in real time [31]; (iv) contributes to increase the motivation and collaboration [28, 29] as well as to increase the interaction, student involvement, and positive attitudes in learning [15, 26, 32–34]; (v) enhances learning by making it more interactive and more attractive [20]; and (vi) is easy to use [15, 35].

Billinghurst and Duenser [36] highlight the importance of AR in the educational area, emphasizing that the "*(. . .) AR educational media could be a valuable and engaging addition to classroom education and overcome some of the limitations of text-based methods, allowing students to absorb the material according to their preferred learning style (. . .)*." The same authors also report that AR in the classroom allows for mediation between the real and the virtual environment

15

Table 15.2 AR and computational thinking projects

Project name	Characteristics/features	References
AR Spot or AR Scratch	The AR Scratch is an extension that adds AR functionality to the Scratch programming environment. It displays virtual objects in a real-world space, seen through a camera using physical markers to control the objects.	[48]
T-Maze	T-Maze is a tangible programming tool for children aged 5 to 9. Children can use T-Maze to create their own maze maps and complete tasks using tangible programming blocks and sensors.	[49]
AR-Maze	The AR-Maze uses tangible programming blocks and mixes virtual and real elements. Children can create their own programs using the programming blocks and execute the code using a mobile device.	[50]
Code Bits	Code Bits is a tangible paper kit that allows creating programs. The code is processed in the Code Bits mobile application.	[51]
Paper Cubes	Using AR, Paper Cubes aims to teach basic computer skills and more advanced programming in the field of artificial intelligence (AI) or machine learning (ML).	[52]
HyperCubes	HyperCubes is an AR platform for children that uses tangible blocks, in a playful way, and allows an introduction to the basic concepts of computational thinking.	[53]

with more fluid interaction, using metaphors through artifacts for manipulating objects and easy switching of visualizations between reality and digital.

AR-supported systems applied in educational contexts enhance a natural interaction with information/objects or 3D events through the hands (using gestures as in the real world) without using the mouse or keyboard [37]. It ensures the continuous interaction and transition between the real and virtual environments, using tangible interfaces for manipulating objects [6, 38], allowing real-time interaction and immersion in a 3D space [8].

The following references are relevant mathematics applications that already explore AR technology for teaching:

- Geometry AR [39]
- GeoGebra Augmented Reality [40]
- Counting With Paula [41]
- AR Flashcards Addition [42]
- Numbeanies Number Forest [43]

Furthermore, other solutions explore AR technology on different domains, contents, or subjects:

- Sparklab (Chemistry) [44]
- Complete Anatomy 2021 (Anatomy) [45]
- Catchy Words AR(Learn Spelling in 3D)[46];
- Plickers (formative assessment) [47]

Table 15.2 enumerates some projects that use AR technologies and tangible programming blocks to introduce and explore computational thinking concepts, like *CodeCubes* does (an application that will be detailed in following sections):

Still in the field of computational thinking, the serious game Cubely [54] allows solving programming enigmas inside a virtual world using an immersive VR programming environment.

Usability and technology requirements are the main challenges on using AR in education. The students may find this technology complicated [15, 35], or the combination of real and virtual objects can make students feel confused about it [15, 20]. For that, students need time to become familiar and feel comfortable with AR [15, 34]. Some usability issues can be directly related to AR technology, which may result from interface design errors, technical problems, or negative attitudes [15, 35]. It is worth noting the importance of educators' involvement to facilitate the development of AR applications favorable for teaching, which increases their potential to be incorporated into education [15, 26].

Table 15.3 presents some applications that, by allowing the creation of AR resources, potentiate processes of co-creation. They allow users (individually or collaboratively) to design, explore, and experience their own creations, either through the use of physical markers or directly on real-world surfaces. These applications were selected considering their potential application on education since they promote active and engaging learning, namely, the development and exploration of topics such as languages, literature, social and exact sciences, technologies, engineering, and coding/programming. Besides, they can be integrated in other platforms or resources (analogic or digital), supporting and promoting *discovery-based learning, design thinking, and inquiry-based* or *problem-based learning*.

15.3 Serious Games in Pedagogical Contexts

Given the new realities arising from the social networks, the continuous growth of interest in digital games is a phenomenon that crosses the most diverse areas of knowledge and the most varied profiles and ages, from the youngest to the oldest.

Table 15.3 AR and co-creation

App name	Characteristics/features	Type of AR experience	Platforms/installations source	Site
Augmented Class!	Is an intuitive, adaptable, and interactive platform that allows users to create and share their own AR educational projects without any technical knowledge. Offers courses and tutorials. Records AR scenes	Marker-based Markerless	Android/Google Play	http://www.augmentedclass.com
CoSpaces	CoSpaces lets students build their own 3D creations (worlds) and animate them with code (using visual programming) and explore them in AR or VR.	Markerless Marker-based (with the Merge Cube add-on)	iOS/App Store Android/Google Play Windows 10/Microsoft	https://cospaces.io/edu
3DBear	Offers various 3D model collections. Allows importing models from Sketchfab and Thingiverse. Offers lessons and challenges for assignments, classroom management, video recording, and 3D printing	Markerless	iOS/App Store Android/Google Play	https://www.3dbear.io
ARLoopa	Offers AR scanner, 3D models library and map, interactive textbooks, 3D product visualizations, and artworks	Marker-based Markerless Location-based	iOS/App Store Android/Google Play	https://app.arloopa.com/
BlippBuilder (the Blippar app is an AR browser)	The Blippar app recognizes images and real-world objects. Allows adding different types of content to AR experiences: 3D models, tracking video, full screen video, animations, background sound, sound effects, Web links, and image gallery. The AR experience can be published to the Blippar app.	Marker-based Markerless	iOS/App Store Android/Google Play	https://www.blippar.com
Merge Cube EDU	Use a Cube (named Merge Cube, printed and assembled) to allow users to hold digital 3D objects	Marker-based	iOS/App Store Android/Google Play Windows 10/Microsoft	https://mergeedu.com
Sketchfab	Is an intuitive and user-friendly app. Offers over a million scenes in 3D	Markerless	iOS/App Store Android/Google Play	https://sketchfab.com

Marc Prensky [55] enumerates several reasons that make digital games so engaging and attractive to millions of people: They are fun, interactive, and adaptable; they have goals, results, feedback, and winning states (rewards); they support conflicts, competition, challenges, opposition, problems, representation, and history.

In the last few years, the video games/digital games industry has grown exponentially, playing an increasingly important role in entertainment. Some successful video games have been adapted and released as films, such as Tomb Raider or Resident Evil. Simultaneously, there is also a tendency to integrate real and virtual environments using optical sensors (present in mobile technologies) and the convergence of various gaming platforms, combined with a new type of interaction based on gestures.

The "highlight" in contemporary culture, digital games, and video games industry originated many types of research [56–58] and studies [59–61] that allowed to "remove" the label of a mere entertainment and leisure product, turning into tools for educational purposes.

15.3.1 Educational Learning Approaches

The new learning approach, *Game-Based Learning* (GBL), emerges [55]. In that approach, games are used to motivate and involve students with the subject of learning, as is the case with serious games (SG) . A more recent trend is that students can create their own games and develop solutions to problems as well as skills in the area of computational thinking. The value of digital games is unquestionable today, playing is instinctive, and the player is rewarded when learning [62].

Educational SG can be defined as games (video or digital) developed for learning improvement, where entertainment is considered an added value [63]. These types of games seem to increase the player's motivation, allowing for the progression and assimilation of new learning content within a continuous and meaningful narrative. They promote decision-making, exposing students to increasing levels of difficulty and allowing learning through trial and error. Furthermore, a safe environment that allows experimentation and exploration, stimulating curiosity (discovery-based learning) and problem-solving , is provided.

15.3.2 Design of Serious Games for Education

Malone and Lepper [64] identified four strengths that can promote a learning environment as an intrinsically motivating game activity: fantasy, curiosity, challenge, and control. These four key elements can promote meaningful learning and must be considered to understand whether a given game meets these criteria and is suitable for use as a learning resource. Attempting to define specific principles for the design of educational games, these authors illustrate what educational games should be:

- Use fantasy to reinforce the learning objectives and stimulate the student's interests.
- Create sensory stimuli (e.g., through audiovisual means) and arouse curiosity and introduce the challenge. Through achieving goals and feedback, the student must feel continually stimulated, and the difficulties must increase in balance with the acquisition of skills in order to prevent the student from being bored, frustrated, or unmotivated.
- Allow the student to feel a sense of control through endogenous feedback provided by the game. That is, they must feel that their actions determine the learning results.

Although games are usually used for fun, they require a strong correlation between playing and learning. They demand various skills from the player, such as memorization, repetition of tasks, analysis, understanding problems, and solving strategies.

15.3.3 Types of Serious Games Models

According to the literature, digital games can be used successfully as complementary learning tools [65], as well as in the exploration of new teaching-learning paradigms. They have certain advantages over traditional learning materials, such as encouraging decision-making and experimenting with different solutions for solving problems [66] and and creating innovative teaching and learning scenarios both in real, virtual or combined environments. In particular, the use of SG with AR in the classroom allows for implementing different methodologies and learning strategies centered on the student/group. In the mathematics classroom, for example, SGs enhance immersive experiences (to varying degrees) in which players can, in an entertaining way, retain information, reason, and solve problems.

The conceptual model or framework for serious games presented by the authors Yusoff, Crowder, Gilbert, and Wills [67] considers relevant the skills/competencies that the player will develop/experience when playing the game. Competencies refer to skills like (i) *cognitive* , that includes memory resources (recall) and higher-order resources such as analysis, synthesis, and evaluation; (ii) psychomotor, that includes the resources of fluid and timely execution; and (iii) *affective*, that includes the ability to identify, adopt, and evaluate appropriate attitudes and points of view.

Considering these aspects, an SG should include moments that allow the student/player to pause and somehow retreat to reflect on their actions. However, during (pause) moments, it is necessary to keep the player involved and immersed in the game world. Simultaneously, the game should allow the student/player to apply the same strategy in identical situations, allowing for interiorization and knowledge transfer.

Another relevant aspect that must be considered when designing a game is the degree of difficulty, that is, the game should not be too easy, as the player can easily be bored and lose interest, or too difficult, as it can generate frustration in the player [68].

Besides, aspects such as interactivity, problem-solving, and feedback are essential for successful and effective learning [57, 69, 70]. The game should promote moments of reflection for those who play it [58] on the topic of learning, creating intrinsic learning [57] that can be flexible and carried out through challenges [71].

SG can facilitate learning in various fields of knowledge [61], for example, in understanding science and mathematics. It can happen when a variety of scenarios can be explored whenever it becomes challenging to manipulate and visualize concepts such as molecules, cells, mathematical graphs, and others [72]. Being in a risk-free environment encourages experimentation and exploration, stimulating curiosity, learning by discovery, and perseverance. However, its use in a learning context must have well-defined learning objectives centered on the learning theme to promote the acquisition of targeted skills and knowledge [56].

15.3.4 Serious Games and User Experiences

In order for the player to experience the sensation of total immersion, losing the sense of time and everything that goes on around him, which Csikszentmihalyi [71] called flow experience, it is important to remove the barriers that exist between him and the game, created by the devices, for instance, using the body itself to communicate with the game and simulating the interaction with the physical world.

Thus, an SG can be a valuable learning tool because when players learn to play, they learn new things. It is learned not only by reading and/or writing but also through numbers, images, symbols, and graphic instruments [59]. The game fits into the conception of knowledge and learning that derives mainly from the theories of Piaget's genetic epistemology [73] and the sociohistorical perspective of Vygotsky [74], both considering that we are not passive toward the environment, which means there is a response to external stimuli and acting on them allows to build and organize knowledge. For Vygotsky, cognitive development does not come only from biology (a genetic process) but is an essentially social process (a concept closely linked to social interaction).

In SG design, the user experience (UX) must be considered, regarded as the result of motivated actions in specific contexts [75]. The motivation, action, and content of the game form the vertices of a triangle representing the experience's needs/conditions [75].

The games' design considers temporality, that is, the user's previous experiences shape present experiences and affect the future ones. Another aspect of game design focuses on information and content creation based on the users' motivational and action needs. Motivational needs are challenging to assess with any methodology, while action needs can be assessed during the game, for example, by recording the players' behavior and emotional responses. This is the case of the *GamePlay Experience* (GX) , that is created during the player-game interaction, when fun and motivating experience is provided for the player [75].

15.4 Serious Games and Augmented Reality

By combining AR interfaces with educational content, another form of interaction between the real world and the virtual world is offered to students. From one side, the possibility of overlapping virtual elements in the real world allows the experience to stop being static. For another, the interaction that allows changes in real time increases the effectiveness and attractiveness of teaching and learning. If SGs that are already used in the classroom are added to the AR component, they provide students with resources to effectively work and cooperate to solve significant problems and build their solutions, narratives, and connections [76].

Games that incorporate AR technology offer a strong interactive component that allows for a "new" way of learning, by discovering, allowing greater cooperation and collaboration, where everyone can simultaneously have the same experience, in real time [77]. This technology facilitates the formalization of ideas through new ways of visualizing, communicating, and interacting between people and information [78], providing a safe, flexible, controlled, and intuitive environment for experiencing physical interactions.

SGs that explore AR (hereafter abbreviated by SG-AR) can use markers that, whenever detected by a mobile device or a webcam, allow for identifying virtual (usually 3D) objects. This type of game (using markers) can be easily used in the classroom on various subjects, from archeology, history, anthropology, geography, and others [79]. It is also possible to resort to AR games that do not use markers but allow anchoring virtual objects to specific locations in the real world, allowing players to interact with digital objects [79].

The type of SG-AR technologies used influences the user experience. The motivation, interaction, appeal, and fun experiences, supporting doing (playing), feeling, and even learning (GX), are essentially related with the following:

- *Type of devices used*: Such as HMD, smartphones, or tablets (mobile devices may require the handling of the device and the bookmarks simultaneously)
- *Type of graphical user interface*: If it is user-friendly and affordable
- *Type of tangible interaction*: Marker-based, markerless, or location-based
- *Game elements*: And inherent dynamics
- *Specificity*: If created for a specific area, for example, education or health, among others
- *Type of content*: Sound, video, 2D/3D models, links, etc., and intrinsic interactions
- Interface usability and gameplay

On the other hand, the success of their usability, reliability, and robustness depends on the following:

- *Tracking capacity*: In the case of using mobile devices, such as smartphones or tablets, the capacity to track movement and position in relation to the surrounding objects
- *Lightning*: To detect markers (in case of marker-based AR) and also to adjust them to virtual objects
- *Reading environment*: The capacity to "read" the environment, namely, the detection of flat surfaces (in case of markerless RA) and in the determination of dimensions and location
- *Placing objects:* The way the virtual objects are placed in a given place/space and their sharing with other SG users
- *Connectivity*: Especially if involving multiplayer or requiring dynamic integration of virtual content

Considering the massification and portability of mobile devices, in particular smartphones, and the increasing tendency to use them in the educational context, the size and interaction services offered are two important limiting factors [80]. The adoption of SG-AR can also be skewed due to the (i) inadequacy or insufficiency of contents (in the case of alignment between education and curricula) and (ii) low-quality designs or reduced performance, inherent to the use of "heavy" models (mainly 3D models).

Other relevant aspects involve health and safety of the user when using these games as educational resources. Since their use may require movement and displacement, the space limitations and possibility of physical damage can be handicaps.

Despite the presented set of criteria for SG-AR selection and its successful use, it was not easy to find relevant literature that supports it. The main source for all these assumptions came from the accumulated experience while developing our applications.

15.5 Case Studies

In the following sections, three educational projects that use AR and gameplay are compiled based on the authors' publications [81–86]:

- *Visualizing Platonic Solids*: Provides an interactive experience that allows students to interact with platonic solids
- *Code Cubes*: A game that uses paper blocks to introduce computational thinking
- *FootMath*: A game that supports students in learning various mathematical functions with a football game in AR
- The order of presentation of these three projects is gradual regarding the complexity of the interaction and the target audience, that is, students between 7 and 17 years of age.

15.5.1 Visualizing Platonic Solids in Augmented Reality

Motivation

Students are often confronted with difficulties in logical thinking and problem solving, leading many times to frustration and interest loss. One of the most significant challenges of education is to develop methodologies and tools that motivate the students and improve results. AR seems to be a promising approach toward real and effective learning.

Aiming at investigating the potential of AR in learning contexts, we have developed an AR mobile application, the platonic solids, that allows 3D representation of the five-geometry convex regular polyhedrons, known as platonic solids. The application provides a virtual interactive environment for the visualization, construction, deconstruction, and manipulation of these polyhedrons, allowing users to explore the properties of each solid, for example, faces, vertices, and edges, as well as the respective spatial constructions (in 3D) with or without animation, from various angles and distances.

Technology and Specifications

The application was developed in Unity [87] and AR Vuforia [88] platforms for Android and iOS platforms. It used 2D markers for detection, that is, drawings representing the solids (Fig. 15.3).

The Vuforia platform indicates the quality and accuracy of the markers classifying them on a given scale. To achieve a high number of visual characteristics and, therefore, a better detection, different colors, textures, and patterns were applied on the various markers, classified with a degree of accuracy over 90%, on the rating scale of the Vuforia platform. This differentiation also allows detecting various markers simultaneously and the visualization of the corresponding solids. To visualize the 3D solids, users must focus the devices' camera on the AR markers.

The application displays the solids in three dimensions and a caption with the respective solid's name (Fig. 15.4a, b). Virtual buttons allow for the visualization of specific elements such as vertices and edges of each solid [81]. The

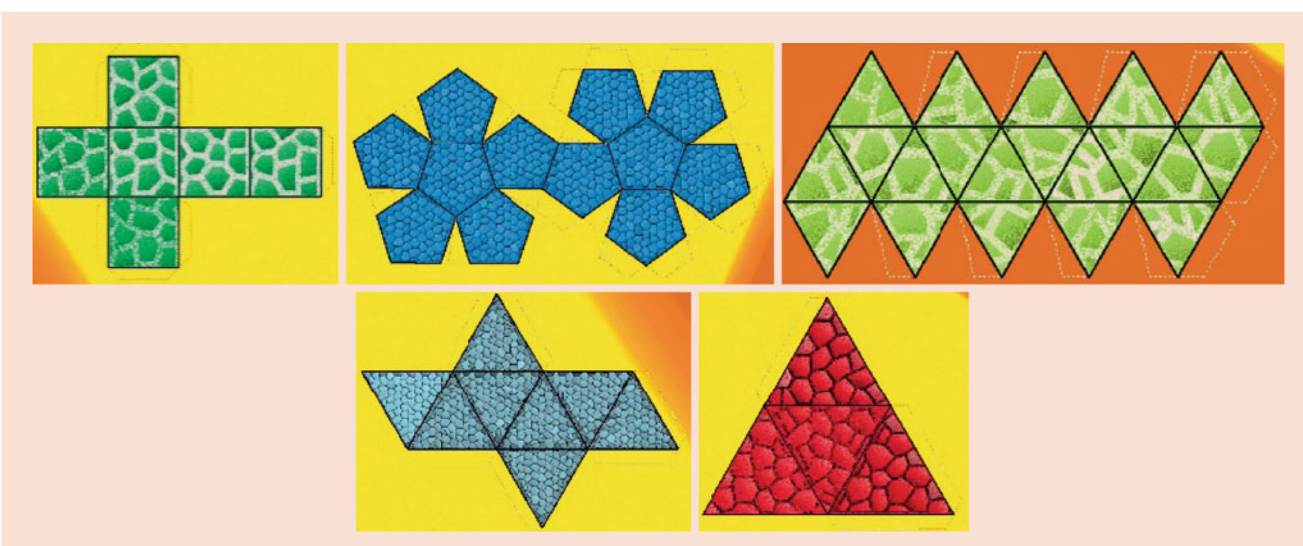


Fig. 15.3 The used markers and the 2D representation of each platonic solid

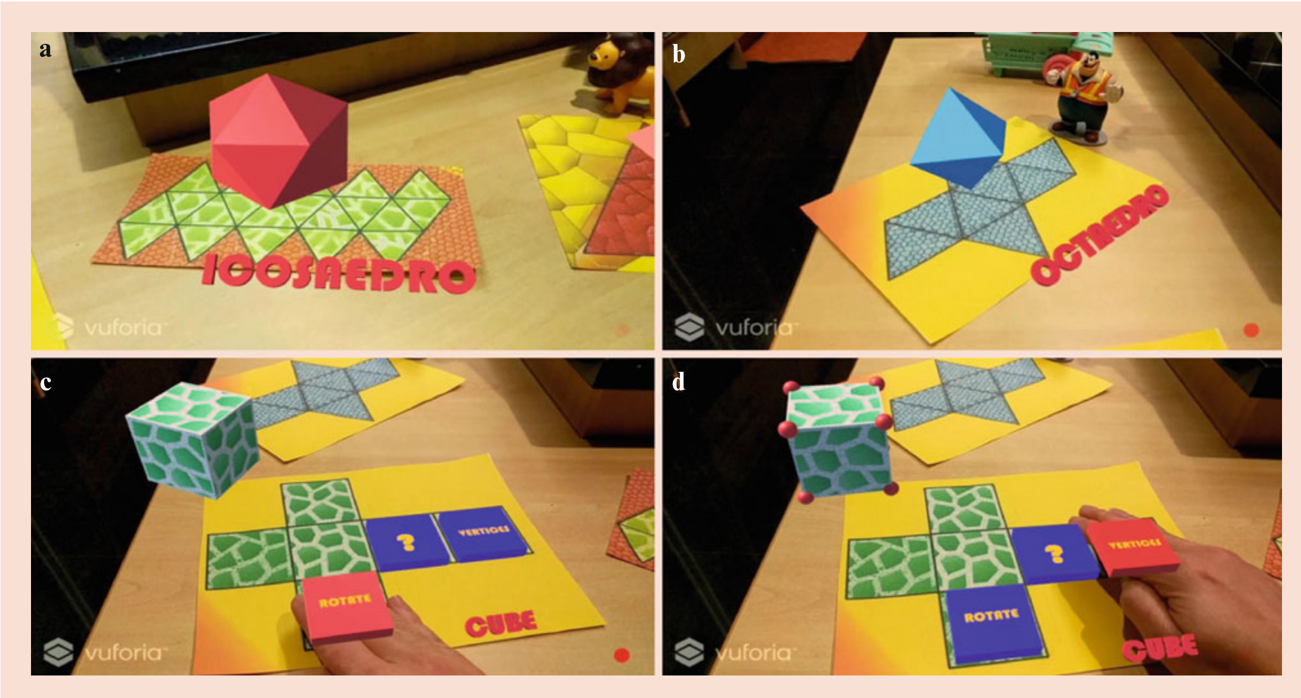


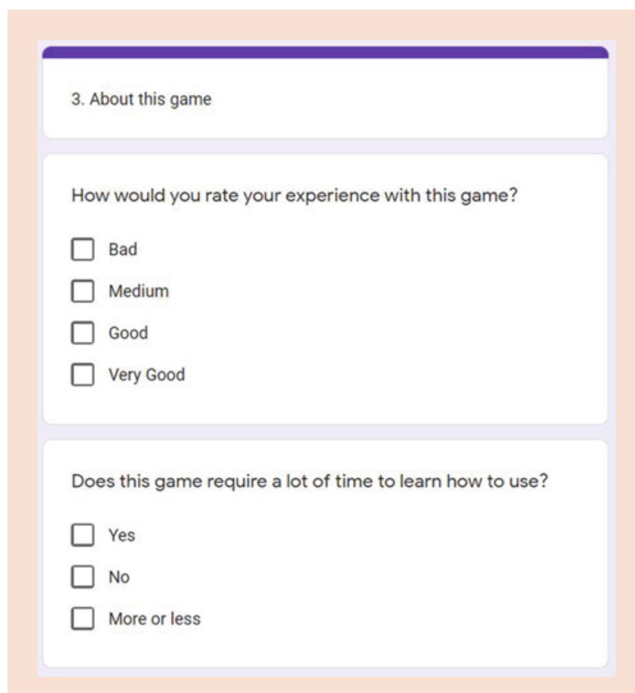
Fig. 15.4 The platonic solids Icosaedro (**a**) and Octaedro (**b**) with respective markers; representations with the virtual buttons for rotating (**c**) and for displaying the vertices (**d**): virtual button (red) active mode

virtual buttons consist of an interaction area that the user activates with the finger or a physical pointer. When the interaction area is activated, the system triggers the corresponding action (Fig. 15.4c, d). However, the use of the virtual buttons might be conditioned by the light conditions, which sometimes hinders the markers' accurate detection. To increase the three-dimensional visualization of the solids and the sense of reality, we used lights that create shadows on the objects' various faces and on the surface where they are displayed.

Pilot Study

The experience data was collected through questionnaires, direct observation, and notes taken at the time.

The application was tested in two distinct moments: Firstly, in the scope of an exhibition [89], the users (four people between the ages of 1 and 20, three people between 31 and 40, and two people between 41 and 50) have experienced the application in Android tablets and personal computers. Later, the experience was in a classroom context with twelve 4th grade students (five males and seven females students) of average age 10. In both moments, after a practical exploration, the users answered to a short survey (using Google platform) to investigate some functional aspects of the application (Fig. 15.5). The survey was applied considering the educational context and the fact that the use of the application occurred before the solids were taught according to the Portuguese curriculum.


3. About this game

How would you rate your experience with this game?

- ☐ Bad
- ☐ Medium
- ☐ Good
- ☐ Very Good

Does this game require a lot of time to learn how to use?

- ☐ Yes
- ☐ No
- ☐ More or less

Fig. 15.5 Part of the questionnaire

The questions were simple and objective such as the following: (a) *Do you find the game interesting to use in an educational context?* (b) *Did you find the game complex?* (c) *How would you rate your experience with this game? (d) Was* *this the first time you saw an augmented reality application?* (e) *Does this game require a lot of time to learn how to use?*

Observations

Globally, the users reported that the application was engaging and with potential to be used in educational contexts. They also expressed the wish to work with this kind of application in other domains.

The students enjoyed visualizing and manipulating the 3D platonic solids. They often collaborated and helped each other with the system using. All the students expressed that they liked using the application and would like to use such tools more regularly in class. All considered that the application facilitates the understanding of abstract concepts and that such materials would potentially facilitate learning and increase their motivation for the learning subject. The students considered that the 3D visualization of the platonic solids helped them understand and visualize their properties. Nevertheless, even very satisfied, the questionnaires revealed that users still find the technology quite hard to use. After an exploratory phase, they classified the experiment as satisfactory or very satisfactory. Fig. 15.6 shows some applied questions and respective answers analysis.

Discussion

The AR application allowed users to create and simulate physical objects corresponding to platonic solids and visualize and manipulate them from various angles and distances according to the supporting devices' movement/rotation. It extended the possibilities of traditional learning, such as the paper book, which only supports a 2D representation of objects and does not allow for any dynamics. Thus, AR can create a mediation environment between the physical and the digital artifact that potentially promotes and facilitates the students' construction of knowledge.

Remarks

An AR application that aims at contributing to the students' better understanding of the platonic solids as a pedagogical resource that allows for a playful exploration of learning contents was described. A first pilot study showed that the students enjoyed using the application and considered it a facilitator for exploring and learning geometry subjects.

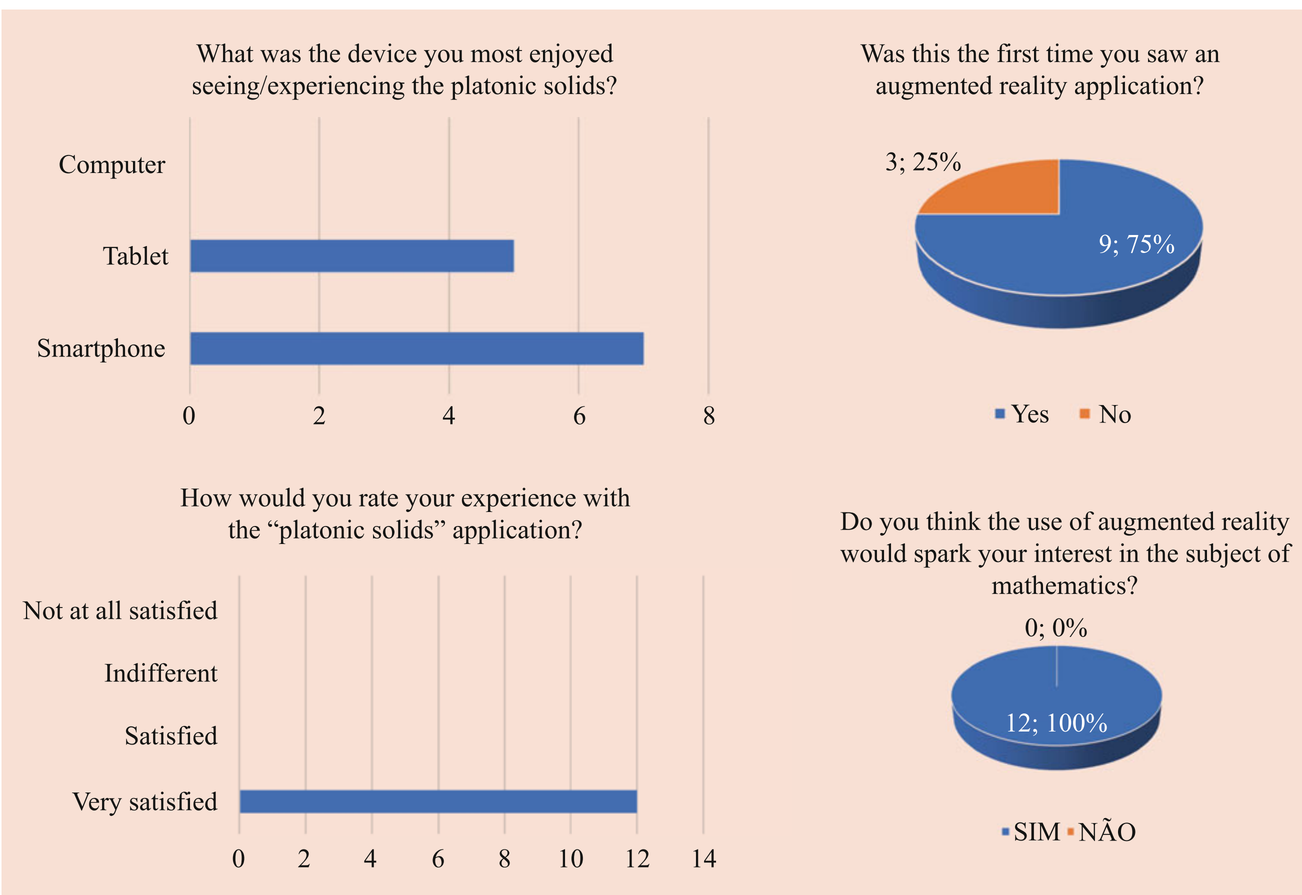


Fig. 15.6 Part of the questionnaire/results

15.5.2 CodeCubes

Motivation

CodeCubes is a hybrid interface that combines (physical) paper squares with AR [82]. This interface aims to combine digital games based on AR technology with educational objectives, to stimulate the development of computational thinking skills, using the physicality that tangible interfaces and AR technology provide [90].

Users manipulate AR markers—physical programming blocks (tangible interface)—that represent programming instructions (Fig. 15.7), allowing them to become familiar with programming instruction sequences using tangible programming blocks [53].

The prototype developed is based on the existing games on the Code.org platform [91], namely, the Angry Birds – Classic Maze [92] mechanics, where the player uses drag and drop blocks of visual programming, ordered in a determined sequence, to program and overcome the various challenges proposed. Compared to these games, the drag and drop blocks feature of the prototype was replaced by square paper (AR marker) that student manipulates like cards and represents the programming instructions capable of virtually implementing it.

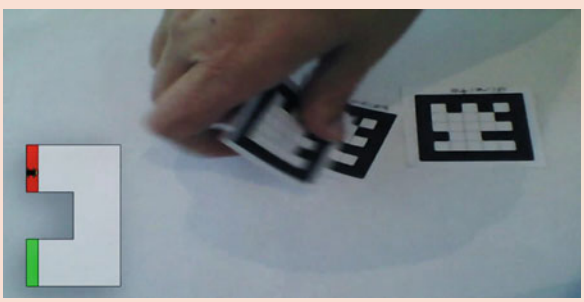


Fig. 15.7 CodeCubes interface

Technical Development

The application was developed using the open-source AR library NyARToolkit [93] for Processing [94] (Java) and explored the Ani library [95] to create the animations and transitions [83].

The game consists of controlling a car's movement under the starting line of a racetrack to successfully reach the finish line. There are three levels of play, in which only the car's track is changed (Fig. 15.8).

The objective is to program a sequence of actions necessary for the car to move and follow the track until the finish line, using only four instructions: right, left, up, and down, represented by each AR marker. The player must place the AR marker representing the desired action in front of the camera in an orderly manner. The car moves forward on the track when executing the defined instructions. The player can execute instruction by instruction or program and execute the complete ordered sequence.

Fig. 15.9 shows three possible programming examples, with the respective sequences, one for each level: 1, 2, and 3. The coordinates and IDs are represented. The coordinates correspond to computer vision detection, from left to right, in the order's programming blocks.

If the player does not execute the instructions or sequence of instructions correctly, but if they can lead the car to reach the finish line, it is possible to change the level. This strategy aims to motivate students to play, allowing them to learn by trial and error and try different solutions. At any time, the player can restart that level. The game is over when the three levels are completed.

Regarding the game screen, in the lower left corner is visible a global perspective of the track to be covered as well as the current position of the car; the pieces are placed in the upper area, and each time a new piece is placed and recognized, an icon with the arrow corresponding to the action to be performed appears in the upper left corner. After placing the parts and recognizing them, each instruction's

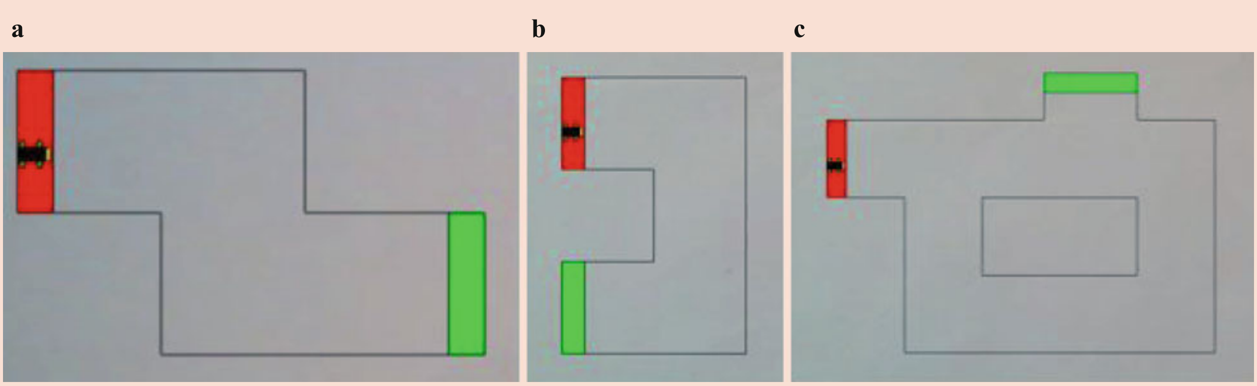


Fig. 15.8 The three levels of the game: level 1 (**a**), level 2 (**b**), and level 3 (**c**)

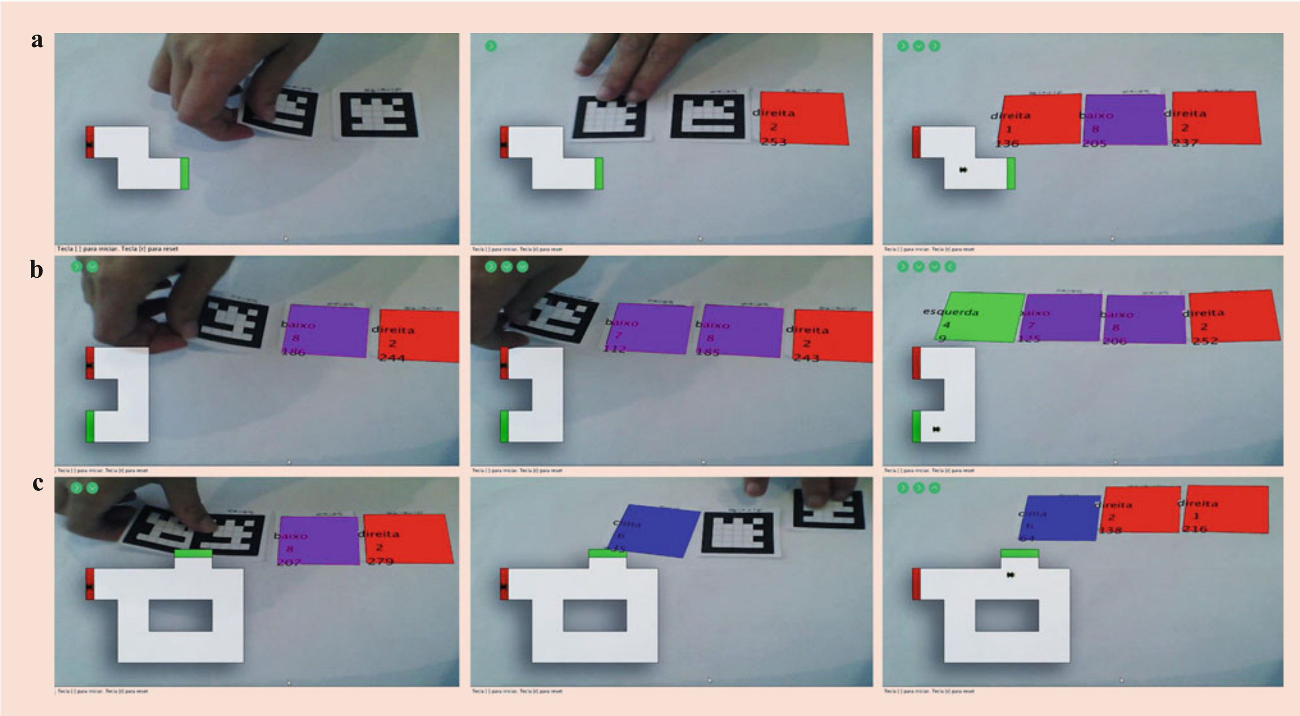


Fig. 15.9 Interaction with the CodeCubes interface at the three levels of the game (level 1 (**a**), level 2 in the (**b**), level 3 (**c**))

processing is performed, leading the car to move on the track. The system carries out the instructions sequentially and in the order they were placed. There is a waiting time between executing the instructions given to understand what is being executed and what they are viewing.

After the sequence execution, and when the car reaches the finish line, a screen congratulates the player, allowing them to play a new level if they activate the restart button. There is always a change in level when the car reaches the finish line, whether the programmed sequence is correct or not.

The player can restart the level and program instruction by instruction or the entire sequence. If the "car leaves the screen" due to a programming error, it always returns to its initial position, being placed on the track's starting line.

Pilot Study

The study was carried out with a group of students from a robotics club in a public school, with an average age 10 years [84]. The parents authorized their students to participate in the study, and the school management authorized the latter.

In these experimentation sessions, the focus was on analyzing any interaction difficulties and the interest, motivation, and the impact that AR technology can have on learning. The tasks proposed to the students were (i) to program a maze, using the Scratch platform, (ii) to program a robot (Lego WeDo) that they use in the club to leave the maze, (iii) to use the Code.org platform to play Classic Maze, and (iv) explore CodeCubes, which was made available to the study's participants.

The following instruments and techniques for data collection were used: (i) direct observation and photographic and audiovisual recording of the tasks the participants had to perform, (ii) interview (carried out in the last session), and (iii) questionnaire (divided into four parts, to be answered at the beginning, after exploring CodeCubes and at the end of all tasks).

The questionnaire was divided into four parts with the following objectives: (1) to characterize the participants, (2) to obtain the opinion of the participants on the application they tested, (3) to assess the impact that AR technology can have on motivation for and on learning, and finally, (4) to evaluate and classify the activity they preferred to carry out, in order to understand the possible difficulties of interaction with CodeCubes and assess whether the use of AR technology increases the interest and motivation of students.

Procedure

There was a total of six sessions, one per week, of 1 h each and at the time and place where the club takes occurs. In the first session, the study participants' activities were presented, and the first part of the questionnaire was answered.

In the remaining sessions, CodeCubes was used to compare with three activities that students are used to doing at the club so that it was possible to verify which activity the students would be most interested in and motivated.

Thus, in the following two sessions (session two and three), students freely explored the provided CodeCubes application, solving the game's three levels, using the AR markers.

In sessions four and five, each student individually performed, in this order, the following tasks: (i) programming CodeCubes level 1, (ii) programming the track using the Scratch platform (the maze was already designed, with students only having to choose a character and do the programming), and (iii) programming a Lego WeDo® robot to follow the same track as that of the CodeCubes at level 1 of the game.

Thus, the same tasks were performed in three different ways, and it was possible to compare and evaluate the activities. Besides, and as a final task, the students used the Code.org platform to play Classic Maze – Angry. Finally, in the sixth session, an interview was held with all participants. The observation focused on their performance, the interaction, and how they performed, resolved, and found solutions to overcome the proposed challenges.

Discussion and Results

Nine participants answered the questionnaires: four males and five females, aged between 9 and 13 years old.

In this first part of the questionnaire, we also tried to understand whether they were familiar with AR technology. It was found that only one student answered "yes" to the question "*Have you ever had any experience with augmented reality programs or games?*" All respondents enjoyed the game; eight rated the experience as "very satisfactory." Nine found the application easy to use, and eight mentioned that they did not need help to use the application despite the totality indicating that it did not take time to learn using it.

All respondents expressed that they would like this resource to be used in classes and books/manuals and considered that it would arouse interest to learn the contents of the other subjects, thus improving their learning. This result was confirmed in the interview when one of the interviewees referred, "*As I was more interested, I had better results*."

For evaluating the activities carried out, a Likert scale with five points was used (*I did not like it, I liked it a little, indifferent, I liked it*, and *I liked it a lot*). It was inferred by the answers given that the activity "programming with AR" was the one they liked the most.

Regarding the question "*What would you change in the game?*," the interview confirmed the answers given in the questionnaires: They referred to the inclusion of more characters and sounds; a game was suggested to allow for the programming of "a character who picked up garbage, the garbage increased from level to level and won whoever managed to fill the bin"; another suggestion said that the game should have a scoring system.

Study Limitations

It was found that participants at the beginning (level 1) preferred to execute the instruction-by-instruction code. They found that placing several pieces in the following levels, they were thinking about the car's trajectory while handling the pieces.

The experience allowed us to perceive a very positive impact. Nevertheless, it must be considered that the sample was not sufficiently representative with the results presented here confined to a very restricted universe. The application to other classes can confirm the results obtained in terms of the AR tool's effectiveness, either as a motivation/interest factor or as a tool that can improve or increase learning. The studies carried out were inconclusive on the relationship between the use of technology and the promotion of learning; this question required a study with more time and resources, comparing CodeCubes with the activities used to teach this topic.

Remarks

The results reveal that the participants who had no previous knowledge and experience related to AR technology did not show any difficulties interacting with the game.

From what was observed, the participants had no difficulty in using the AR markers, the programming pieces. In addition, they showed a lot of interest and curiosity, with no resistance to participate in the study nor to the application's use, which proved its intuitiveness and easiness.

Given the characteristics of the interface that CodeCubes offers, in most cases, it was not necessary to explain its use, which consequently allowed for the autonomous exploration of the game.

The study made it possible to conclude that CodeCubes was very well accepted, with all the activities carried out being the favorite and the one they liked the most. Participants responded enthusiastically and positively to the proposed tasks. During the sessions, they were very motivated and interested in carrying out the proposed activities.

As future work, a multiplayer interface will expand the interaction between groups toward creating a collaborative environment in the game. In this way, an attempt will be made to expand the initial study on AR in serious games to a level of exploration and investigation of possible influences on students' social skills, in particular, to see whether a possible "multiplayer with AR" extension can promote collaborative problem-solving.

Finally, other levels and different challenges could be added to the game. This upgrade should allow for a broader exploration of programming knowledge, for example, by including new blocks of instructions to use repetition structures, selection, and functions.

15.5.3 FootMath

Motivation

FootMath is a serious game intended for high school and secondary mathematical contexts with a *Digital Game-Based*

Learning (DGBL) approach [55]. The game uses AR technology to visualize, manipulate, and explore mathematical concepts, particularly the affine, quadratic, exponential, and trigonometric functions (sine and cosine).

Being a serious digital game with AR technology, it is a digital artifact for education, providing tangible mediation on mathematical functions exploration (Fig. 15.10).

The design of this serious AR game starts from the idea of providing teachers and students with a teaching and learning resource that potentially facilitates the exploration of a mathematical subject that is often considered difficult to learn: functions. Thus, the difficulties that students face in learning mathematics were considered.

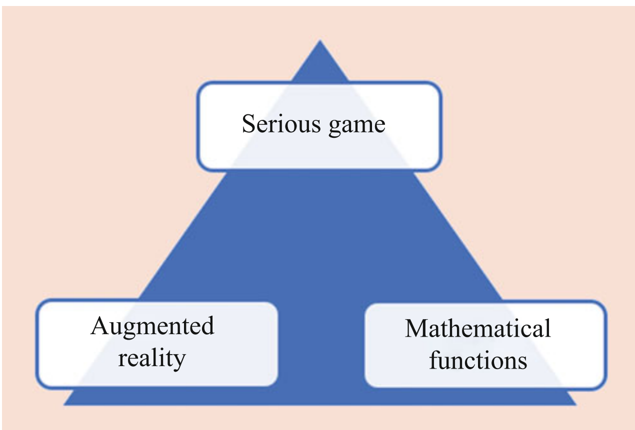


Fig. 15.10 Digital game combining AR and mathematical functions

Disruptively, considering that traditional math teaching and learning resources use scientific manuals and calculators to solve problems, FootMath simulates a 3D football game [85]. Users can manipulate and explore different functions and respective parameters to get the ball into the goal, that is, score goals (Fig. 15.11).

Technical Development

An AR Software Development Kit (SDK) allows the incorporation (by reusing) of implemented AR components, such as recognition and tracking [96]. Although any existing SDK has advantages and disadvantages, the Vuforia SDK was selected considering the target mobile devices' reliability and compatibility. It should be noted that Vuforia provides a Web-based environment where users can create and manage bookmarks and obtain the necessary licenses to test and publish applications. Thus, FootMath was developed in Vuforia to run on Android mobile platforms since most students have Android smartphones and/or tablets.

FootMath was designed to incorporate AR and to work with 2D physical markers. In the implementation, the markers are part of the tangible interface and serve to trigger the exploration of the various mathematical functions. The set of physical markers was created on the Vuforia Web platform and then printed on paper.

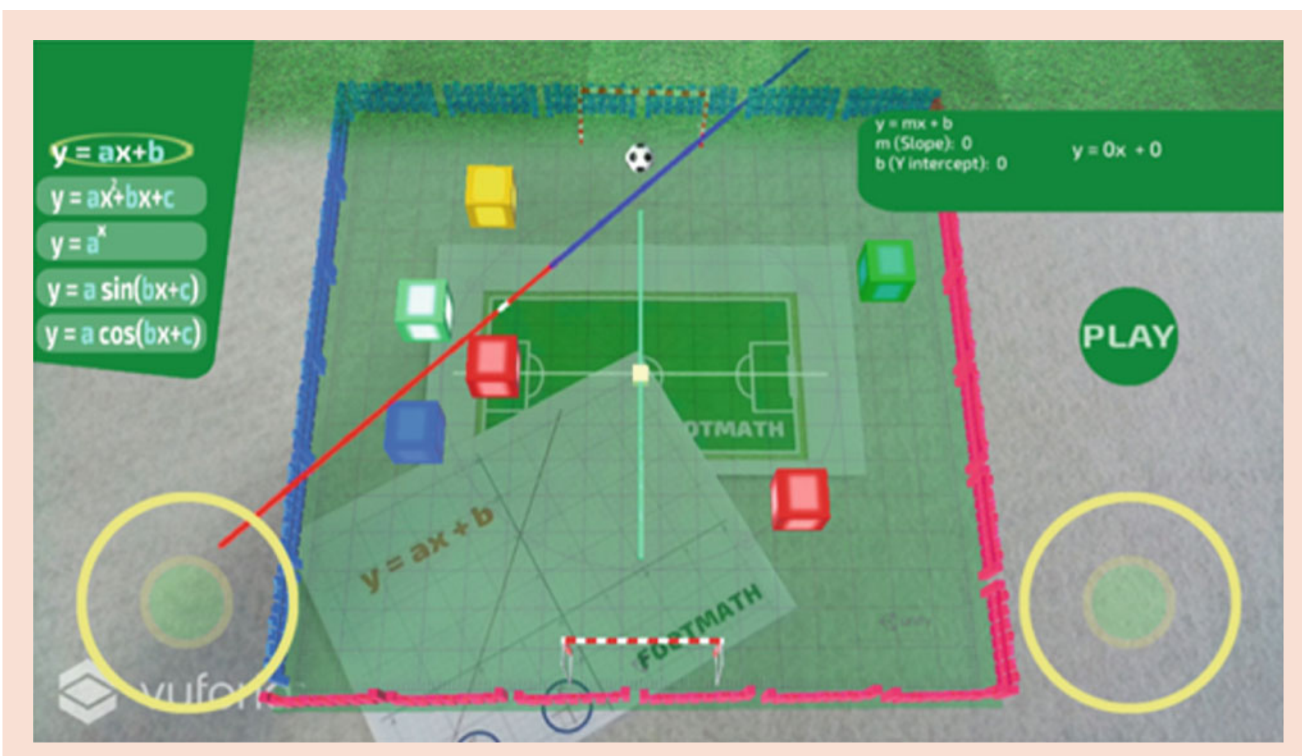


Fig. 15.11 FootMath game (function $y = ax + b$ represented in the game area)

For a good detection, each marker must have a high rating in terms of visual characteristics (features). Thus, different patterns and colors were applied to the various markers, and classified with a degree of accuracy above 90%, on the classification scale of the Vuforia platform [81]. This accuracy degree allows two markers to be detected simultaneously and the corresponding functions to be displayed. In this way, it allows the comparison of various functions and their simultaneous exploration on the football field [86].

FootMath presents "Football" as its central theme. To provide a fun game experience, based on the "Football" metaphor, it was tried to create a simulation with a simplified and attractive perspective [62]. The process of interacting with the ball was one of the critical points of the game's design and development, mainly due to the exogenous effects intended. For this, it was necessary to apply and experiment with different types of physical variables, such as speed and angles, to the "materials," taking into account aspects such as friction, rebound, bounce, and the elasticity (dynamic or static friction and bounciness configurable in the development platform). Considering this, different physics types were applied to the game ball, to the football field's ground, and to its surrounding objects (on the stage).

In order to not transform the simulation of the game into simple gameplay based on the selection of functions with the physical markers and the ball shot, some obstacles (in the form of cubes) were created with random positioning to stimulate the so-called ball rebound and effects resulting from the collision with the ball. In this way, it was possible to implement in the simulator with unpredictability in the ball's collision with these obstacles (opponents of the player).

Other aspects that proved to be important were the ball's speed and acceleration. In the initial phase, there were difficulties in programming/developing ways to control the ball, that partially conditioned the prototype's development in terms of fluidity and spontaneity in the simulator's behavior. To overcome this problem, it was necessary to refine the simulator with different experiences in terms of the speed and acceleration of the ball (considering aspects such as the mass of the rigid body that constitutes the ball, the physics of the material, the applied force, and the initial speed).

Exploring FootMath

The game FootMath uses 2D physical markers (Fig. 15.12) to trigger the various functions. When the camera detects a marker, the FootMath displays a 3D football field with a menu on the left with five buttons (selection). The buttons show the following functions: $y = ax + b$ (linear); $y = ax^2 + bx + c$ (quadratic); $y = a^x$ (exponential); $y = a\sin(bx + c)$ (sine); and $y = a\cos(bx + c)$ (cosine). Each physical marker causes it to be displayed on the football field the proper graph with the respective function (Fig. 15.13).

An interface menu supported by two joysticks, allows to manipulate the existing function parameters (*a, b, c, d*). The joystick in the lower right corner allows the player to manipulate parameters *a* and *b*; the joystick in the lower left corner allows the player to manipulate parameters *c* and *d*.

Each equation button shows the parameters that the user can manipulate. When the user changes the values with the joystick, the parameter display is shown in the menu's upper right corner. Furthermore, the user can change the function parameters, triggering the "plot" in the virtual environment.

After defining the correct position to score a goal, manipulating the function according to the ball's position, the user "kicks" the ball by touching the *Play* button. A small cube, located on the left side at the beginning of the function, begins its trajectory along with the function. When the ball is on the same trajectory as the function, the cube collides with it, allowing the goal to be scored.

To increase the difficulty of scoring a goal, random cubes circulate on the football field, influencing the ball's trajectory. Therefore, depending on the ball's position and the cubes, the user needs to decide which mathematical function is the best to score a goal.

In one of the final stages of the development of FootMath (a more solid and robust version), a first informal assessment was made, regarding the design and usability, with a group of 24 of 7th- and 8th-grade students. The students explored the different functionalities of FootMath during an Information and Computer Technology (ICT) class (due to the availability of resources) and expressed their satisfaction in using AR to explore mathematical functions. In this stage, FootMath presented very good results (feedback given by the students), highlighting the curiosity for experimenting with physical markers and the way of playing.

Pilot Study

Considering (i) the interest in the population's particular characteristics, that is, teachers [97]; (ii) the interest to study teachers' experience and perceptions; and (iii) the Creswell and Poth [98] recommendation that the sample size should be between 5 and 25 participants, the exploratory study was carried out with two intentional groups of 11 teachers each, coming from several schools of the 3rd cycle of basic education and secondary, from different areas: one group in mathematics and another in ICT.

From the exploration of FootMath, the study focused on discovering the benefits and challenges related to the potential use of serious games with AR, exploring basic mathematical functions as a playful and educational tool capable of motivating and involving users, which can be used to facilitate problem-based learning and logical reasoning. It was also tried to determine if the game environment would

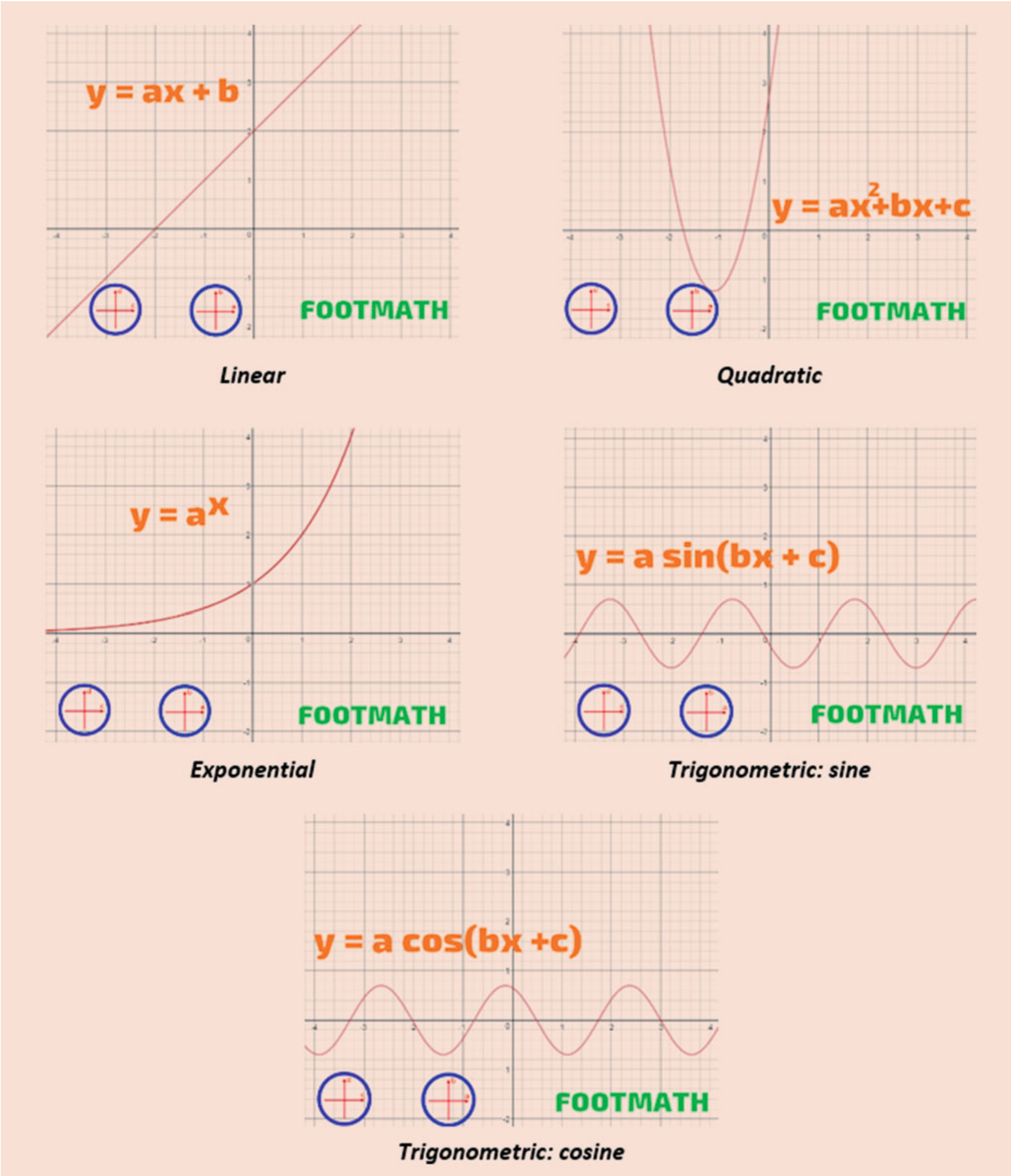


Fig. 15.12 AR 2D markers for the functions: linear, $y = ax + b$; quadratic, $y = ax^2 + bx + c$; exponential, $y = ax$; and trigonometric, sine, $y = a \sin(bx + c)$ and cosine, $y = a \cos(bx + c)$

provide exploration, reflection, the discovery of solutions, and logical reasoning in an attractive, motivating, and engaging way.

FootMath was considered a promising and innovative game to be applied in mathematics teaching from the analysis of the obtained results.

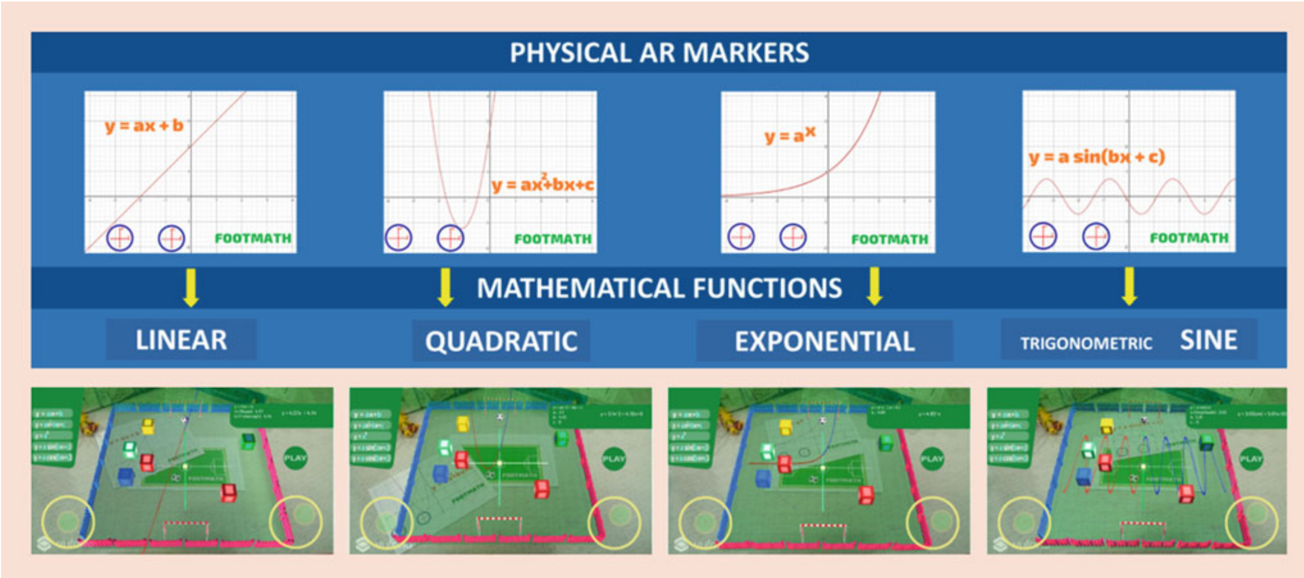


Fig. 15.13 FootMath game. (functions linear, quadratic, exponential, and trigonometric sine represented in the game area)

Procedure

The two groups were formed accordingly to the two workshops that took place previously and in different periods. A set of activities for each workshop was planned and described in a task guide. At the beginning of the work sessions with the teachers, AR's theoretical concepts were briefly introduced and discussed to know if they were familiar with the AR topic. In the following, videos of AR applications were presented, such as *platonic solids*, *geometry-augmented reality*, and *Pupil Augmented Learning Project* at Carnegie Mellon University, as well as this one, *FootMath*.

The first workshop was for the mathematics teachers. They had the opportunity to try the three AR apps in the following order: platonic solids, geometry-augmented reality, and then FootMath. For this purpose, three tablets and app markers were prepared. In the case of geometry-augmented reality, some of the participants installed and experimented with their own smartphones. During the experimentation and exploration of the FootMath game, using direct observation, dialoguing with the participants and collecting their feedback, the relevant data/information resulting from the experience, was recorded.

In the second workshop, ICT teachers, following the script, had also the opportunity to view and try the same AR applications.

At the end of each of the workshops, a group interview using open questioning was conducted, seeking to explore and discover the impact resulting from using the apps.

Discussion and Results

Regarding the direct observation of the experiments carried out, it was possible to register an active participation and a high involvement in all teachers' proposed activities. However, they showed greater interest in the activities performed on tablets compared to those performed on smartphones.

In the first stage, they explored the FootMath game's operation, both in the positioning of the tablet/smartphone in relation to the markers and in the process of using the joysticks. Teachers quickly learned how to use joystick markers and were able to use them properly. The improvement in the experience with AR was evident as they were using the game.

Overall, the game provided a lot of curiosity, interest, interaction, and attention. In the group interviews, teachers from both groups were unanimous in considering that FootMath can be used in the classroom to complement the teaching-learning processes of subjects related to mathematical functions. Additionally, they considered that it could contribute to creating rich learning scenarios and improve the consolidation of subjects that involve functions. Everyone agreed with the game's pertinence and identified a positive impact that can enhance students' learning and motivation.

Intrinsically related to motivation, they highlighted benefits such as the attention and satisfaction provided by this educational resource's experience. Regarding learning based on the "problem solving" model, teachers were asked to comment on whether it would be possible to "pour" pedagogical instructions into the game experience, following the paradigm's strategies in question. Several suggestions were made, coming essentially from mathematics teachers: in the linear function, $y = ax + b$, scoring goals under certain conditions, for example, with a positive or negative slope (value of a); in the quadratic function, $y = ax^2 + bx + c$, scoring goals under certain conditions, for example, the concavity of the parabola facing up or facing down; and other situations, such as scoring goals when the function graph is positioned at its zeros. In this way, several complementary scenarios were identified to provide problem-solving through the game experience with interaction and mediation based on AR.

Remarks

With this serious AR game development, it was intended to contribute to a better understanding of using AR technology in the learning environment. We also intended to explore a new space offered by emerging technology, bringing new interaction paradigms to the teaching environment. We tried to assess whether AR provides and enhances learning.

The results reinforce a positive impact concerning the possible use of AR mathematics game apps in the classroom. Teachers, faced with this new learning opportunity, agreed that it has several benefits and can enhance the learning of mathematical functions.

It can be concluded that this type of application can provide a tangible experience with several benefits and complement formal approaches to motivate and engage. It offers a fun environment that allows for the visualization and exploration of basic mathematical functions. FootMath provides the player with an interactive and interpretive experience that involves doing, feeling, and learning.

15.6 Closing Remarks

This section resumes the main results from the experiences and evaluations processes of the described applications. Globally, and considering the *platonic solids*, the results allowed the identification of distinct behaviors, namely, (i) whether there was learning, (ii) the pedagogical potential of the applications, and (iii) founded difficulties, both in handling the apps and in the contents covered. Participants' interaction with the application was also observed. The students were unanimous in saying that they would like this type of applications to be used on learning subjects, both in general and in relation to the mathematic solids, and found the idea of books and worksheets with AR interesting. Everyone mentioned that this type of application can increase/improve their learning, and its use would arouse their interest in mathematics subjects. Furthermore, the students indicated that the application helped them to visualize/understand the solids (in 3D).

Considering the *CodeCubes,* the found results indicated that this technology, being new, aroused much curiosity and interest. In the interview, the students said that it was fun to watch and control the car moving, manipulating the paper markers, while it was possible to visualize their hands and table where they were programming. Seeing the virtual elements overlapping the real environment and be able to control the visual elements through the physical manipulation of objects fascinates them.

The questionnaire results had shown that the technology has the potential to be used in the classroom and integrated into educational activities and also sustained that AR contributes to increase students' motivation and improve their performance, as well as to conclude that the use of this technology makes learning more interesting and attractive, enhancing knowledge.

Considering the *FootMath*, the interesting results (after students' feedback) emphasized their experience with physical markers and the gameplay, as well as confirmed a positive impact regarding the possible use of this game in the classroom. Teachers, on this new learning opportunity based on a digital AR game, agree that it has several benefits and can enhance the learning of the basic functions of mathematics.

Finally, we are conscious that the carried-out studies cannot be sufficiently conclusive. We believe the relation between the use of technology and the promotion of learning requires deeper studies with more users, time, and resources, toward the comparison of explored games (*platonic solids, CodeCubes, and FootMath*) with the activities that are used for the teaching and learning this domain.

15.7 Conclusions

In this chapter, three educational/didactic resources using AR technology have been described. The *Platonic Solid AR* experience, the *AR CodeCubes* app, and the AR serious game with mathematical functions, *FootMath*, demonstrate AR serious games' potential to teach and learn mathematics and programming.

These solutions use a type of interaction mediated by AR technology to provide a more tangible experience with several benefits. The combination of serious games with AR technology enhances these benefits, offering fun environments that can be applied in different contexts of teaching approaches. It was possible to contribute to a better understanding of AR technology's usefulness and applicability in serious games at the educational level. The results sustain that AR offers new interaction paradigms to teaching/learning environments.

Following the work developed and the results obtained, it was unanimous in noting that the AR technology (i) shows potential to be used in the classroom and integrated into educational activities, (ii) contributes to increase students' motivation and improve their performance, and (iii) makes learning more exciting and attractive, enhancing knowledge acquisition from the very beginning.

Concerning the research on the impact of using these resources, it must be highlighted that everyone's enthusiasm was constant, and the participants were receptive to integrate this technology in the classroom, with the potential to be applied in learning the contents of other disciplines. The results also demonstrate a high level of interest and motivation in promoting active learning through playful AR technology activities.

Furthermore, the AR combination with digital serious games confirmed a positive impact on motivation, interest, attraction, and performance, enhancing the acquisition and practical application of knowledge. It seems to be a promising approach toward practical learning, allowing the exploration of abstract mathematics' concepts and programming, offering an innovative and more practical and objective way to learning it.

Overall, the work developed contributed to show that integrating virtual models/solutions that include AR technology and serious games in the virtual environments constitutes an auspicious approach that awakens multiple possibilities to explore and apply in education.

As a final though, and in the sequence of these experimentations, we believe the integration of these new kind of tools in the teaching curriculum is relevant. However, its acceptance depends on further research and experimentations, involving specialists on pedagogical and didactic domains as well.

References

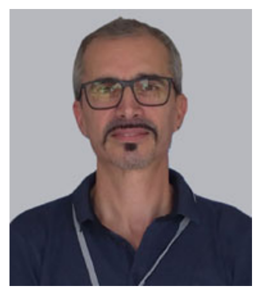
1. Gandolfi, E.: Virtual reality and augmented reality. In: Kennedy, R.E., Ferdig, K. (eds.) Handbook of Research on K-12 Online and Blended Learning, 2nd edn, pp. 545–561. ETC (2018)
2. Chen, Y., Wang, Q., Chen, H., Song, X., Tang, H., Tian, M.: An overview of augmented reality technology. J. Phys. Conf. Ser. **1237**, 022082 (2019). https://doi.org/10.1088/1742-6596/1237/2/022082
3. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **E77-D**, 1321–1329 (1994)
4. Mann, S.: "Mediated reality with implementations for everyday life," Presence Connect, vol. 1, (2002).
5. Mann, S.: Mediated Reality. In: M.I.T. Media Lab Perceptual Computing Section, Cambridge, MA (1994)
6. Kesim, M., Ozarslan, Y.: Augmented reality in education: current technologies and the potential for education. Procedia-Social Behav. Sci. **47**, 297–302 (2012). https://doi.org/10.1016/j.sbspro.2012.06.654
7. Liarokapis, F., Macan, L., Malone, G., Rebolledo-Mendez, G., Freitas, S.D.: A pervasive augmented reality serious game. In: 2009 Conference in Games and Virtual Worlds for Serious Applications, pp. 148–155. IEEE (2009). https://doi.org/10.1109/VS-GAMES.2009.40
8. Azuma, R.T.: A survey of augmented reality. Presence Teleoperators Virtual Environ.**6**(4), 355–385 (1997). https://doi.org/10.1162/pres.1997.6.4.355
9. Çöltekin, A., et al.: Extended reality in spatial sciences: a review of research challenges and future directions. *ISPRS Int. J. Geo-Inf.* **9**(7), 439 (2020). https://doi.org/10.3390/ijgi9070439
10. Brown, M. et al.: "2020 Educause horizon report teaching and learning edition," Louisville, CO, 2020. [Online]. Available: https://www.learntechlib.org/p/215670.
11. Raghavan, R., Rao, P.: "Immersive learning for the future workforce accenture extended reality (Xr)," 2018. Accessed 20 Jan 2021. [Online]. Available: https://www.accenture.com/_acnmedia/pdf-86/accenture-extended-reality-immersive-training.pdf.
12. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. Multimed. Tools Appl. **51**(1), 341–377 (2011). https://doi.org/10.1007/s11042-010-0660-6
13. Holz, T., Campbell, A.G., O'Hare, G.M.P., Stafford, J.W., Martin, A., Dragone, M.: MiRA—Mixed reality agents. Int. J. Hum. Comput. Stud. **69**(4), 251–268 (2011). https://doi.org/10.1016/j.ijhcs.2010.10.001
14. Chen, C., Tsai, Y.: Interactive augmented reality system for enhancing library instruction in elementary schools. Comput. Educ. **59**(2), 638–652 (2012). https://doi.org/10.1016/j.compedu.2012.03.001
15. Khan, T., Johnston, K., Ophoff, J.: The impact of an augmented reality application on learning motivation of students. Adv. Human-Comput Interact. **2019** (2019). https://doi.org/10.1155/2019/7208494
16. Lin, T.-J., Duh, H.B.-L., Li, N., Wang, H.-Y., Tsai, C.-C.: An investigation of learners' collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system. Comput. Educ. **68**, 314–321 (2013). https://doi.org/10.1016/j.compedu.2013.05.011
17. Nincarean, D., Alia, M.B., Halim, N.D.A., Rahman, M.H.A.: Mobile augmented reality: the potential for education. Procedia – Soc. Behav. Sci. **103**, 657–664 (2013). https://doi.org/10.1016/j.sbspro.2013.10.385
18. Estapa, A., Nadolny, L.: The effect of an augmented reality enhanced mathematics lesson on student achievement and motivation. J. STEM Educ. **16**(3), 40–49 (2015)
19. Iftene, A., Trandabăt, D.: Enhancing the attractiveness of learning through augmented reality. Procedia Comput. Sci. **126**, 166–175 (2018). https://doi.org/10.1016/j.procS.2018.07.220
20. Wu, H.-K., Lee, S.W.-Y., Chang, H.-Y., Liang, J.-C.: Current status, opportunities and challenges of augmented reality in education. Comput. Educ. **62**, 41–49 (2013). https://doi.org/10.1016/j.compedu.2012.10.024
21. Velev, D., Zlateva, P.: Virtual reality challenges in education and training. Int. J. Learn. Teach. **3**(1), 33–37 (2017). https://doi.org/10.18178/ijlt.3.1.33-37
22. Pellens, M., Hounsell, M., Silva, A.: Augmented reality and serious games: a systematic literature mapping. In: 2017 19th Symposium on Virtual and Augmented Reality (SVR), pp. 227–235 (2017). https://doi.org/10.1109/SVR.2017.37
23. Bakri, F., Marsal, O., Muliyati, D.: Textbooks equipped with augmented reality technology for physics topic in high-school. J. Penelit. Pengemb. Pendidik. Fis. **5**(2), 113–122 (2019). https://doi.org/10.21009/1.05206
24. Behzadan, A.H., Menassa, C.C., Kamat, V.R.: Georeferenced augmented reality for discovery-based learning in civil engineering. In: Transforming Engineering Education, pp. 199–228. American Society of Civil Engineers (2018)
25. Rebollo, C., Remolar, I., Rossano, V., Lanzilotti, R.: Multimedia augmented reality game for learning math. Multimed. Tools Appl., 1–18 (2021). https://doi.org/10.1007/s11042-021-10821-3
26. Wei, X., Weng, D., Liu, Y., Wang, Y.: Teaching based on augmented reality for a technical creative design course. Comput. Educ. **81**, 221–234 (2015). https://doi.org/10.1016/j.compedu.2014.10.017
27. Goh, E.S., Sunar, M.S., Ismail, A.: 3D object manipulation techniques in handheld mobile augmented reality interface: a review. *IEEE Access*. ***24***(7), 1 (2019). https://doi.org/10.1109/ACCESS.2019.2906394
28. Dünser, A., Walker, L., Horner, H., Bentall, D.: Creating interactive physics education books with augmented reality."In: 24th

15

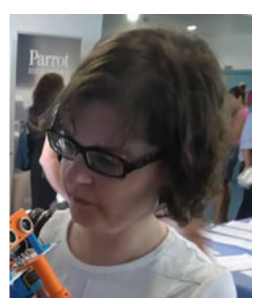
Australian Computer-Human Interaction Conference, pp. 107–114, 2012 doi: https://doi.org/10.1145/2414536.2414554.
29. Kaufmann, H.: The potential of augmented reality in dynamic geometry education. In: 12th International Conference On Geometry and Graphics (ISGG), Ago, pp. 6–10 (2006)
30. Bower, M., Howe, C., McCredie, N., Robinson, A., Grover, D.: Augmented reality in education — cases, places, and potentials. EMI. Educ. Media Int. **51** (2014). https://doi.org/10.1080/09523987.2014.889400
31. Diegmann, P., Schmidt-Kraepelin, M., Van den Eynden, S., Basten, D.: Benefits of augmented reality in educational environments – a systematic literature review. In: 12th International Conference on Wirtschaftsinformatik, March 4–6 2015, Osnabrück, 2015, pp. 1542–1556, Accessed 16 Feb 2021. [Online]. Available: https://aisel.aisnet.org/wi2015/103.
32. Bacca, J., Baldiris, S., Fabregat, R., Graf, S., Kinshuk: Augmented reality trends in education: a systematic review of research and applications. Educ. Technol. Soc. **17**(4), 133–149 (2014)
33. Di Serio, Á., Ibáñez, M.B., Kloos, C.D.: Impact of an augmented reality system on students' motivation for a visual art course. Comput. Educ. **68**, 586–596 (2013). https://doi.org/10.1016/j.compedu.2012.03.002
34. Gopalan, V., Zulkifli, A. N., Bakar, J. A. A.: A study of students' motivation using the augmented reality science textbook. In: AIP Conference Proceedings, 2016, vol. 1761, no. 1, p. 20040, doi: https://doi.org/10.1063/1.4960880.
35. Akçayır, M., Akçayır, G.: Advantages and challenges associated with augmented reality for education: a systematic review of the literature. Educ. Res. Rev. **20**, 1–11 (2017). https://doi.org/10.1016/j.edurev.2016.11.002
36. Billinghurst, M., Dünser, A.: Augmented reality in the classroom. Comput. (Long. Beach. Calif). **45**, 56–63 (2012). https://doi.org/10.1109/MC.2012.111
37. Svedström, T.: Gesture interfaces. Aalto Univ. Sch. Sci. Technol. Inf. Nat. Sci. (2010)
38. Billinghurst, M.: Augmented reality in education. New horizons Learn.**12**(5), 1–5 (2002)
39. Ventura. Geometry AR. https://www.venturaes.com/iosapps/geometryar.html (2020). Accessed 04 Mar 2021
40. GeoGebra. App downloads – GeoGebra. https://www.geogebra.org/download (2020). Accessed 04 Mar 2021
41. Studios, O.: Android apps by Omens Studios – APK Support. https://apk.support/developer/Omens+Studios (2018). Accessed 04 Mar 2021
42. Heath, M., Callahan, D.: Home – AR Flashcards. https://arflashcards.com/ (2019). Accessed 04 Mar 2021
43. Inness, S.: Numbeanies number forest. http://www.simoneinness.com/projects/numbeanies.html (2020). Accessed 04 Mar 2021
44. "Sparklab – Chemistry app in AR/VR – Apps no Google Play." https://play.google.com/store/apps/details?id=ge.arx.sparklab.android&hl=pt&gl=US Accessed 11 Apr 2021
45. "Complete anatomy | advanced 3D anatomy platform." https://3d4medical.com/ Accessed 11 Apr 2021
46. "Catchy Words AR on the App Store." https://apps.apple.com/us/app/catchy-words-ar/id1266039244 Accessed 11 Apr 2021
47. "Plickers." https://get.plickers.com/ Accessed 11 Apr 2021
48. Radu, I., MacIntyre, B.: Augmented-reality scratch: a children's authoring environment for augmented-reality experiences. In: Proceedings of the 8th International Conference on Interaction Design and Children, pp. 210–213 (2009). https://doi.org/10.1145/1551788.1551831
49. Wang, D., Zhang, C., Wang, H.: T-Maze: a tangible programming tool for children. In: Proceedings of the 10th International Conference on Interaction Design and Children, pp. 127–135 (2011). https://doi.org/10.1145/1999030.1999045
50. Jin, Q., Wang, D., Deng, X., Zheng, N., Chiu, S.: AR-Maze: a tangible programming tool for children based on AR technology. In: Proceedings of the 17th ACM Conference on Interaction Design and Children, pp. 611–616 (2018). https://doi.org/10.1145/3202185.3210784
51. Goyal, S., Vijay, R.S., Monga, C., Kalita, P.: Code bits: an inexpensive tangible computational thinking toolkit for K-12 curriculum. In: Proceedings of the TEI' 16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 441–447 (2016). https://doi.org/10.1145/2839462.2856541
52. Fusté, A., Amores, J., Ha, D., Jongejan, J., Pitaru, A.: Paper cubes: evolving 3D characters in augmented reality using recurrent neural networks. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 31–33 (2017)
53. Fuste, A., Schmandt, C.: HyperCubes: a playful introduction to computational thinking in augmented reality. In: Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, pp. 379–387 (2019). https://doi.org/10.1145/3341215.3356264
54. Vincur, J., Konopka, M., Tvarozek, J., Hoang, M., Navrat, P.: Cubely: Virtual Reality Block-Based Programming Environment. (2017). doi: https://doi.org/10.1145/3139131.3141785.
55. Prensky, M.: Digital Game-Based Learning, vol. 1. McGraw-Hill, New York (2001)
56. Gros, B.: The impact of digital games in education. *First Monday*. **8**(7), 6–26 (2003)
57. Malone, T.: Toward a theory of intrinsically motivating instruction. *Cogn. Sci.***5**(4), 333–369 (1981). https://doi.org/10.1207/s15516709cog0504_2
58. Pivec, M., Kearney, P.: Games for learning and learning from games. *Organizacija*. **40**(6) (2007)
59. Gee, J.P.: What video games have to teach us about learning and literacy. *Comput. Entertain.***1**(1), 20 (2003). https://doi.org/10.1145/950566.950595
60. Gee, J.P.: Learning by design: games as learning machines. *Interact. Educ. Multimed. IEM*. **8**, 15–23 (2004)
61. Gee, J. P.: *Good Video Games and Good Learning: Collected Essays on Video Games, Learning and Literacy*, 2nd Ed., no. April 2008. New York: P. Lang, (2007).
62. Martinho, C., Santos, P., Prada, R.: Design e Desenvolvimento de Jogos. FCA (2014)
63. Li, J., Spek, E., Feijs, L., Wang, F., Hu, J.: Augmented reality games for learning: a literature review. In: Streitz, N., Markopoulos, P. (eds.) Lecture Notes in Computer Science. Distributed, Ambient and Pervasive Interactions, vol. 10291, pp. 612–626. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-58697-7_46
64. Malone, T., Lepper, M.: Intrinsic motivation and instructional effectiveness in computer-based education. In: Snow, R.E., Farr, M.J. (eds.) Conative and Affective Process Analyses, pp. 255–286. Lawrence Erlbaum Associates, Mahwah (1987)
65. Drigas, A.S., Pappas, M.A.: On line and other game-based learning for mathematics. *Int. J. Online Eng.***11**(4), 62–67 (2015). https://doi.org/10.3991/ijoe.v11i4.4742
66. Griffiths, M.: The educational benefits of videogames. *Educ. Heal.***20**(3), 47–51 (2002)
67. Yusoff, A., Crowder, R., Gilbert, L., Wills, G.: A conceptual framework for serious games. In: Ninth IEEE 2009 International Conference on Advanced Learning Technologies, pp. 21–23. IEEE (2009). https://doi.org/10.1109/ICALT.2009.19
68. Verhaegh, J., Fontijn, W., Jacobs, A.: On the benefits of tangible interfaces for educational games. In: 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, pp. 141–145 (2008). https://doi.org/10.1109/DIGITEL.2008.37

69. Prensky, M.: Don't Bother Me, Mom—I'm Learning: How Computer and Video Games are Preparing Your Kids for 21st Century Success—and How You Can Help! (2006)
70. Zyda, M.: From visual simulation to virtual reality to games. Computer (Long. Beach. Calif). **38**(9), 25–32 (2005). https://doi.org/10.1109/MC.2005.297
71. Csikszentmihalyi, M.: Creativity: Flow and the Psychology of Discovery And Invention. HarperPerennial (1996)
72. Mitchell, A., Savill-Smith, C.: The use of Computer and Video Games for Learning: A Review of the Literature. Learning and Skills Development Agency, London (2004)
73. Piaget, J.: The theory of stages in cognitive development. In: Green, D., Ford, M.P., Flamer, G.B. (eds.) Measurement and Piaget, pp. 1–11. McGraw-Hill, New York (1971)
74. Vygotsky, L.: Play and its role in the mental development of the child. J Russ East Eur Psychol. **42**(4), 35–54 (1933)
75. Nacke, L., Drachen, A., Goebel, S.: Methods for evaluating gameplay experience in a serious gaming context. *Int. J. Comput. Sci. Sport*. **9**(2) (2010)
76. Schrier, K.: Using augmented reality games to teach 21st century skills. In: ACM SIGGRAPH 2006 Educators Program, pp. 15–es. ACM, New York (2006)
77. Leitão, R.: Aprendizagem baseada em jogos: realidade aumentada no ensino de sólidos geométricos. Universidade Aberta, R Leitão (2013)
78. Tori, R., Kirner, C., Siscoutto, R.A.: Fundamentos e tecnologia de realidade virtual e aumentada. Editora SBC, Porto Alegre (2006)
79. Yuen, S.C.-Y., Yaoyuneyong, G., Johnson, E.: Augmented reality: an overview and five directions for AR in education. *J. Educ. Technol. Dev. Exch.***4**(1), 11 (2011). https://doi.org/10.18785/jetde.0401.10
80. Ferrer, V., Perdomo, A., Rashed-Ali, H., Fies, C., Quarles, J.: How does usability impact motivation in augmented reality serious games for education? In: 2013 5th International Conference on Games and virtual worlds for serious applications (VS-GAMES), pp. 1–8 (2013)
81. Cerqueira, J., Cleto, B., Moura, J.M., Sylla, C.: Visualizing platonic solids with augmented reality. In: Proceedings of the 17th ACM Conference on Interaction Design and Children – IDC '18, pp. 489–492 (2018). https://doi.org/10.1145/3202185.3210761
82. Cleto, B., Moura, J.M., Ferreira, L., Sylla, C.: Codecubes – playing with cubes and learning to code. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, vol. 265, pp. 538–543 (2019). https://doi.org/10.1007/978-3-030-06134-0_58
83. Cleto, B., Sylla, C., Ferreira, L., Moura, J.M.: CodeCubes: coding with augmented reality. In: OpenAccess Series in Informatics, vol. 81, (2020). https://doi.org/10.4230/OASIcs.ICPEC.2020.7
84. Cleto, B., Sylla, C., Ferreira, L., Moura, J.M.: 'Play and learn': exploring CodeCubes. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, vol. 307 LNICST, pp. 34–42 (2020). https://doi.org/10.1007/978-3-030-40180-1_4
85. Cerqueira, J., Moura, J.M., Sylla, C., Ferreira, L.: An augmented reality mathematics serious game. In: OpenAccess Series in Informatics, vol. 81, (2020). https://doi.org/10.4230/OASIcs.ICPEC.2020.6
86. Cerqueira, J., Sylla, C., Moura, J.M., Ferreira, L.: Learning basic mathematical functions with augmented reality. Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST. **265**, 508–513 (2019). https://doi.org/10.1007/978-3-030-06134-0_53
87. "2D 3D Game Creator & Editor | Augmented / Virtual Reality Software | Game Engine | Unity." https://unity.com/products/unity-platform Accessed 13 Apr 2021
88. "Vuforia Developer Portal |." https://developer.vuforia.com/ Accessed 13 Apr 2021
89. "Exposição Digital Games @ IPCA no GNRation – EST." https://est.ipca.pt/noticia/exposicao-digital-games-ipca-no-gnration/ Accessed 11 Apr 2021
90. Sabuncuoğlu, A., Erkaya, M., Buruk, O.T., Göksun, T.: Code notes: designing a low-cost tangible coding tool for/with children. In: Proceedings of the 17th ACM Conference on Interaction Design and Children, pp. 644–649 (2018). https://doi.org/10.1145/3202185.3210791
91. "Learn today, build a brighter tomorrow. | Code.org." https://code.org/ Accessed 11 Apr 2021
92. "Code.org – Classic Maze." https://studio.code.org/hoc/1 Accessed 11 Apr 2021
93. "NyARToolkit for processing | NyARToolkit project." https://nyatla.jp/nyartoolkit/wp/?page_id=166 Accessed 11 Apr 2021
94. "Processing.org." https://processing.org/ Accessed 11 Apr 2021
95. "Ani – An animation library for Processing." http://www.looksgood.de/libraries/Ani/ Accessed 11 Apr 2021
96. Herpich, F., Guarese, R., Tarouco, L.: A comparative analysis of augmented reality frameworks aimed at the development of educational applications. *Creat. Educ.***08**, 1433–1451 (2017). https://doi.org/10.4236/ce.2017.89101
97. Johnson, R., Christensen, L.: Educational Research: Quantitative, Qualitative, and Mixed Approaches. SAGE Publications, Thousand Oaks (2019)
98. Creswell, J., Poth, C.: Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Sage publications, London (2017)

15



José Cerqueira: holds a master's degree in engineering on Development of Digital Games from Polytechnic Institute of Cávado and Ave and BA in Math and Computer Science from University of Minho. He is currently an ITC teacher and is involved in research and development projects related with augmented reality, virtual reality, and math serious games. He had published several articles about augmented reality and serious games for education.



Bárbara Cleto: Invited lecturer at Escola Superior De Media Artes E Design (ESMAD/IPP) Collaborator member of uniMAD, Teacher of

Computer Science, Student of PhD Program Multimedia Education – University of Aveiro (UA) Master's in engineering in Digital Games Development – Polytechnic Institute of Cávado and Ave (IPCA) Post-graduation in Multimedia Communication (UA), Degree in Electronics and Computer Science Teaching (UA)

João Martinho Moura: is a multi-awarded artist-researcher working in media art and virtual reality. He has collaborated in projects at international institutions such as ESA, INL, NATO, and UNESCO and was laureate at the European Commission's STARTS and MINDSPACES science/art initiatives. In 2013, he received the National Multimedia Art and Culture Award in Lisbon for his contributions to the field of media arts.

Cristina Sylla: received her PhD from the University of Minho in 2014. She develops her research at the Research Centre on Child Studies at UMinho and at the ITI (Interactive Technologies Institute)/LARSyS, University of Madeira. Her work focuses on the design of innovative tools at the intersection of traditional games and playful activities and new interactive technological solutions. CIEC – ITI/LARSyS, University of Minho, Braga

Luís Ferreira: is associate professor and integrated researcher at 2Ai – Applied Artificial Intelligence Laboratory in School of Technology of the Polytechnic Institute of Cávado and Ave, Portugal. He received his PhD in Industrial and Systems Engineering from University of Minho in 2013. His research interests include Computer Science, Architectures and Integration of Information Systems and Services, Data Science, and Artificial Intelligence.

Check for updates

16 Augmented Reality for Cultural Heritage

Bruno Fanini, Alfonsina Pagano, Eva Pietroni, Daniele Ferdani, Emanuel Demetrescu, and Augusto Palombini

Contents

16.1 Virtual Technologies for Cultural Heritage Legibility 391

16.2 State of the Art on AR in CH Contexts 393
16.2.1 Historical Background 393
16.2.2 Usage of AR Between Real and Virtual 394

16.3 Relevant Case Studies in CH: Different Approaches to AR 395
16.3.1 A Selection of Case Studies 395
16.3.2 Analytical-Comparative Summary of Use Cases 398

16.4 Technical Limitations and Challenges of AR for CH 400

16.5 Multisensorial AR Experiences 403

16.6 Efficacy and Effectiveness of AR 404
16.6.1 The Importance of User Experience Design for AR Systems 404
16.6.2 Efficacy and Effectiveness of AR Systems in Relation to the Context of Use 405

16.7 Directions and Future Perspectives 407

References 408

Abstract

Augmented reality applications for Cultural Heritage have been implemented in the last years. The use of AR is currently diffused for many purposes, from technical and managing activities to dissemination. On the communication side, the main potential of such approach is the extension of human sight as to cover simultaneously the current situation of a point of interest (monuments, archaeological sites, artifacts, etc.) and the reconstruction of its ancient condition in different historical periods. Moreover, it allows to compare different possible hypotheses and evaluate the reliability of each of them in the general context of the known data. Despite the current limits of AR (due to the approximation in device positioning), which still influence the use of applications for mobile use, such technology is very promising in the fields of tourism, education, and entertainment; it allows to enrich 3D scenarios with different types of content. In the next future, it is reasonable to imagine a huge amount of information, in different formats, potentially reachable by people simply having a look at the remains of the ancient past and choosing their favorite topics. At the same time, the connection of AR with other emerging technological infrastructures (such as IoT) will allow to collect, view, and manage simultaneously lots of real-time diagnostic data in the same framework. This will foster the research activity toward the definition of new communication metaphors and cognitive solutions.

Keywords

Interaction · User experience · Storytelling · Immersiveness · Cultural Heritage reading and interpretation

16.1 Virtual Technologies for Cultural Heritage Legibility

Interaction with places, objects, and people *generates stories*, those stories that cultural objects are witnesses of into museums and archaeological areas. Nevertheless, most of the time, these stories do not emerge and are not made evident because nothing or no one is able to collect them and promote their existence, both in real and virtual forms. Yet these stories intrigue visitors; they involve them from an emotional point of view, and allow them to keep memory of what just heard once they come back home. Technology makes this process of content's *understanding and memorization*

B. Fanini · A. Pagano · E. Pietroni · D. Ferdani · E. Demetrescu · A. Palombini (✉)
CNR ISPC, Rome, Italy
e-mail: bruno.fanini@cnr.it; alfonsina.pagano@ispc.cnr.it; daniele.ferdani@cnr.it; eva.pietroni@cnr.it; emanuel.demetrescu@cnr.it; augusto.palombini@cnr.it

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_16

easier, as it makes "visible" concepts and meanings related to cultural objects and sites. It therefore facilitates the mental process of abstraction in every visitor, turning into legible and recognizable what usually is not.

However, technology is a "mean-through" – an operative modality or a tool used to strengthen certain communication paradigms or to convey information from a transmitter to a receiver. The selection of the target users to whom address the communication, the quantity and quality of information to be transferred, the context where such communication takes place, and the experience designed for them cannot be decided starting from the technological support. Rather, these choices must be influenced by the overall purpose of our communicative process and by the evaluation of the best user experience design in specific conditions of use. Technology in museums needs to be invisible and unobtrusive, which means, for example, that if *head-mounted displays* (HMD) are used for augmented reality (AR), these cannot represent a limit to the understanding of what the visitor is observing or these cannot tire him while wearing it; for the same reason, the placement of an interactive installation into a museum context cannot fail to respect the visit pathway, and it must take into account the flow of visitors and the possibility of enjoying the technological experience collectively. Definitely, technology needs to be well calibrated when used for cultural purposes, and it needs to follow precise criteria and rules in order to take the best out of its possibilities [55].

Exploiting the potential offered by virtual reality technology not only improves the communication of Cultural Heritage but also makes the cultural offer more attractive to different user groups. When we talk about cultural innovation we refer to the most innovative technologies to improve visitors' enjoyment. Actually, what we are doing is improving the *User eXperience* (UX) of the cultural object or site they are living, so to transform that moment into an "augmented," and even customizable story – a new piece of memory. Indicators such as the usability of the multimedia applications, the visibility of the elements that set up the graphic interfaces, the satisfaction generated by the use and the vision/perception of real and virtual environments, and the comprehensibility and memorization of the cultural contents transmitted are just some of the aspects that are important to be evaluated. Likely, the study of the users' visit path, the time spent using digital applications, and the users' behavior toward the museum items, actively or passively, in a group or alone, are crucial elements for the design of new user experiences, and for the success of any cultural communication project. Innovating, when dealing with Culture, means renewing the perception and the knowledge of a site or an artifact and enhancing, to a certain extent, its meaning and its social value. Innovation therefore brings Culture to everyone's attention, making it understandable but without trivializing it, emphasizing its points of convergence with modern times without simplifying its creative and social processes. This is even more true if AR is used, given the intrinsic and unique potential of the technological medium.

Working on Cultural Heritage through virtual technologies allows to establish a more intimate and profound "dialogue" with visitors who, always more, look for interactivity and sensory immersiveness during the cultural experience. Multi-projection systems, holographic applications, immersive viewers, and AR systems let users immerse themselves in reconstructed scenarios of the other ages, to interact with precious artifacts and ancient characters, while personalizing and integrating their itineraries. Technologies like AR that enters museums and cultural places, if well exploited, can increase both the concrete meanings of heritage and the personal meanings emerged by the direct experience of users.

The AR technology applied to Cultural Heritage helps integrating the visualization needs with scientific information to increase legibility and contextalization of cultural sites [27]. Indeed, *"the visualization of places of historical interest closed to the public or no longer existing and the re-contextualization of historical or archaeological objects are examples of the first trend; the appeal to novelty, to focus on beautifully executed virtual reconstructions of the second. From these first experimentation, the trend has been towards a growing mix of these two aspects, with the aim of creating virtual scientific reconstructions, but at the same time beautiful and engaging, also increasing the amount of accessory information made available [to the public]. In short, the [two faces of AR] are increasingly linked and functional to each other: for teaching, for tourism, for scientific research. The AR, after all, allows an emotional approach that has the great power to address not only the rational part, but also the irrational and intimate part of the users. In particular, the reference to the daily life aspects of history and to the approach of micro history, which is in any case able to guide towards macro events, [allowing the progressive involvement] of the public"* - [19].

AR technology is therefore different than virtual reality: if the latter simulates a complete reality, totally reconstructed by means of digital tools, the former is a *"mixture between the perception of the surrounding reality and the images generated by a system in order to provide the user with additional information while moving and interacting with the actual environment that surrounds him"*- [11].

Section 16.2 of this contribution will illustrate the general foreground of AR definitions and, more generally, of *mixed reality* (MR), and their application fields. Section 16.3 will present case studies and provide a general grid of references for analysis on potentialities and limitations of AR in Cultural Heritage domain. Section 16.4 will allow to go further into the topic through reflections, practical examples, and comparisons with respect to the effectiveness of AR and its technological solutions, in terms of reading and understanding the Cultural Heritage. Finally, Sect. 16.6 will give an overview of the strengths and weaknesses of the virtual technologies and of the methods that can be used, depending on the context, the conditions of use, the target, and the methods of interaction, so as to outline best practices and guidelines for future augmented reality applications. Section 16.7 will close the document, envisioning possible directions of future narrative and technological developments in AR/MR.

16.2 State of the Art on AR in CH Contexts

While the scientific community agrees on including "virtual reality" among the representations of the "cybernetic world" – considering it as a fully synthetic world, completely disconnected from any real context – the debate on "augmented reality" (AR) is still open. Over the years, many attempts have been made to provide a common and broadly accepted definition of the term "augmented reality" [45]. Work by [64] formulated base criteria that have been widely accepted, defining AR as an integration of virtual object into a real environment. However, it was argued that, since AR is still a developing technology, it has not yet reached its full potential [30]. As a result, the definition of AR can change according to the context, method of implementation [67], or technologies [5]. Still today there is a lot of confusion so that AR is frequently overlapped or mixed up with the term "mixed reality" (MR). According to the scientific literature, there several existing notions of AR and MR [6, 25, 37, 62]. The most common are:

– **AR as a specific case of MR: The "virtuality continuum"**. In definition, theorized by Milgram and Kishino, the term "mixed reality" refers to a *subset of virtual reality (VR) technologies that involve the merging of real and virtual worlds somewhere along the "virtuality continuum" which connects completely real environments to completely virtual ones* (Fig. 16.1) [45]. Thus, MR include both augmented reality, where digital information and virtual objects augment images of the real world, and augmented virtuality, where real objects augment artificial computer-generated images (http://www.v-must.net/virtual-museums/glossary).
– **AR as AR as a variant of virtual environments**. Azuma defined AR as "a variation of virtual environments" which "allows the user to see the real world, with virtual objects superimposed upon or composited with the real world." According to his AR, system is not technology dependent but depends on three characteristics: (1) combines real and virtual, (2) is interactive in real time, and (3) is registered in three dimensions [5].
– **MR as evolution of AR**. Interaction-based notion. It considers MR as a specific case of AR. While AR is defined as a condition in which virtual objects are visually overlapped or juxtaposed to the real world, MR is instead interpreted as a step beyond AR in which virtual objects placed in the real world can be interacted with [74].
– **AR as synonymous of MR**. In this notion, both terms, AR and MR, can be used interchangeably or considered similar entities [26, 38].

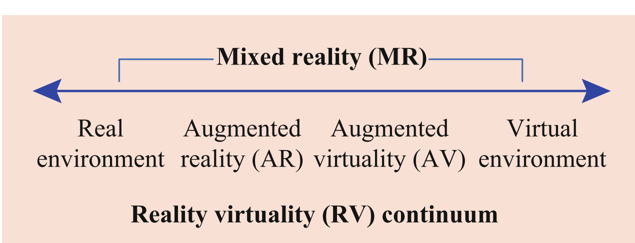


Fig. 16.1 Schematic representation of "virtuality continuum"

Given this premise, in the following sections, we will refer to AR following the Milgram and Kishino definition because in our opinion it describes the *virtuality continuum* in a clear and concise way that abstracts from technology.

16.2.1 Historical Background

The birth of AR dates back to around 1960 and can be attributed to Ivan Sutherland who, in one of his well-known work, postulated an "Ultimate Display" [65]. Technological advances between 1980 and 1990 allowed the AR to emerge as an independent research field. The term "augmented reality" first appeared in 1992 where see-through virtual reality goggles were provided to assist workers in the airplane factory by displaying wire bundle assembly schematics [14]. In the Cultural Heritage field, applied AR technologies started at the beginning of the twenty-first century. One of the first and successful applications of AR is Archeoguide which the first experiment was made on the archaeological site of Olympia, Greece [70]. This first application demonstrated the potential of connecting such technologies with cultural and heritage tourism. AR indeed allows to enhance the visiting experience at cultural sites by providing "augmented" information and three-dimensional interpretations [4, 66] improving learning process though edutainment [48]. Furthermore, using virtual reconstructions of heritage sites and manufacts into AR applications became widespread for dissemination purposes into virtual archaeology and virtual museums fields. It is the case of projects like:

– *Lifeplus*, a 3D reconstruction of ancient Pompeian frescoes featuring groups of virtual animated characters in

an immersive AR environment. The project was based on video captured in real time and augmented with realistic 3D simulations [50];
- *Arich*, an AR system for visualization of archaeological artifacts, and *Arco*, an augmented representation of cultural objects. The goal of the former was to integrate the video and 3D digitized object in an augmented reality environment. The latter aimed at creating a robust digital capture, reconstruction, and presentation through AR interface of cultural objects [46];
- *Prisma*, a project for the implementation of a new 3D visualization prototype based on AR technologies for cultural tourism applications. The project aimed at combining tourist binoculars with AR technologies in order to enhance real scene with multimedia personalized interactive information and user-friendly interfaces [24];
- *iTacitus*, a European research project funded in Seventh Framework Programme in which an augmented reality presentation system for remote cultural heritage sites was developed. The project allowed to superimpose cultural visual media on real cultural heritage sites via video see-through. It used Mobile AR Guide Framework based on markerless tracking on mobile computers [75];
- *CityViewAR system*, an AR application for supporting learning activities where students can use smartphone to see the buildings of Christchurch as they were before the 2011 earthquake [10];
- *March*, a multimedia interactive AR solution to discover prehistoric engravings by highlighting in real time, on the smartphone, the engravings framed with the camera [15];
- *Trends* show how activity in relation to project, patents, and publications related to augmented reality have significantly increased since 2010 [31].

In the last decade, thanks to the evolution of AR dedicated software and devices as well as the development of computer graphics, the application of augmented technologies in Cultural Heritage domain is constantly growing. In-depth analysis of some virtuous and cutting-edge case studies can be found in Sect. 16.3.

16.2.2 Usage of AR Between Real and Virtual

AR does not have a single approach to fruition but can be enjoyed in different ways depending on the objectives, the context of fruition, and the users to be reached. A possible classification, suggested by Milgram [45], distinguishes between three modes of fruition according to the technology used. The first two modes involve the use of displays and depending on whether or not they allow to view reality are defined "see-through AR display" or "monitor-based AR display." The third mode whereas allows the enjoyment of virtual content superimposed on reality without the use of viewers or media that stand between the user and the physical world.

1. See-Through AR Display System equipped with semi-transparent displays that allow to directly view the surrounding world. The virtual contents are displayed, thanks to an illusory technique better known as Pepper's ghost and born in the sixteenth century. This technique uses a game of reflections to project virtual elements (images or videos) from a display on a transparent glass, plexiglass, or other plastic film to give the illusion that these elements appear inside environments and interact with objects or physical persons. A very widespread example of this technology is the so-called holographic displays which are very common in the commercial sector for advertising small products (Fig. 16.2). Portable see-through AR displays (like HoloLens, Magic Leap One) are based on this approach: they are wearable HMD that offer stereoscopic fruition. Such solutions are still under development.

2. Monitor-Based AR Display System of analog or digital superimposition or juxtaposition of virtual and real images shown on display. This means that the reality is not actually "achieved," as in the previous case, through a transparent plate, but it is displayed on a monitor. In case of superimposition, following a real-time video capture, on the real-time shot of reality (which is transmitted to the display from the camera of the device used), virtual contents are then superimposed, thanks to dedicated apps. In case of juxtaposition, the display shows only a virtual content (e.g., a 3D reconstruction) of the object that the camera of the device is filming. The comparison between real and virtual is made by the user who moves the eye from the monitor to the physical context and vice versa to identify similarities and differences. The visual correspondence between the two contexts (real and virtual) is generally guaranteed by tracking tools (markers, geometric recognition algorithms) and/or spatial positioning tools (GPS, beacons, network) that allow the AR system to match the geometries of the real object with those of the virtual object displayed on the monitor (Fig. 16.3).

3. Projector-Based AR/Spatial AR The system allows the superimposition of virtual reality directly on the physical world (usually architecture or objects), without the use of viewers or devices, but through projections. Virtual contents such as lights, images, or videos are projected and made to coincide with real surfaces using projection mapping techniques. Thanks to the advantage of being device-free, it has been applied in many fields such as medicine, industrial design, art installations in parks and theatres, etc. One of the best known examples of success on a national level is the

Fig. 16.2 Example of see-through AR display in a museum context where virtual contents are displayed on the real object (a gold fibula) using Pepper's ghost technique. The augmented content related to the object is projected by an "invisible" display placed above the object and reflected by a transparent panel inclined at 45°. Holographic showcases developed within the European project CEMEC, by Allard Pierson Museum (APM), Amsterdam, and CNR ISPC (former ITABC), 2018

Fig. 16.3 Examples of monitor-based AR display. The tablet camera captures the environment in real time and the AR app overlays interactive virtual content such as Mars and Venus statue reconstruction and information hot-spots

show *Journey to Ancient Rome* by Piero Angela and Paco Lanciano (http://www.viaggioneifori.it/), realized inside the Forum of Augustus and the Forum of Caesar where the story is accompanied by a visual narration with virtual reconstructions projected directly upon the historical architectures (Fig. 16.4).

16.3 Relevant Case Studies in CH: Different Approaches to AR

There are many examples of AR in the field of Cultural Heritage. For the purpose of this section, we tried to select seven case studies that are representative of the modalities and technologies related to such technological solution (see Sect. 16.2.2). In some cases, they are pioneering and research applications, as in the case of *Olympia*, but already quite complete at conceptual level, while, in other cases, they are products developed mainly for tourism and commercial purposes. See Sect. 16.3.1 and Fig. 16.6 for a schematic summary of use cases and Sect. 16.3.2 for a comparative analysis.

16.3.1 A Selection of Case Studies

The diffusion of AR solutions in cultural heritage is a trending topic. In the last European calls of Horizon 2020, several projects like iMARECULTURE (https://imareculture.eu/) or ARETE (https://cordis.europa.eu/project/id/856533), which use AR as tools for education and for improving access to Cultural Heritage, have been financed. In this section selected examples of relevant case studies developed in the last decade are presented.

Olympia, 2000

A milestone in the history of AR applied to Cultural Heritage is certainly the *Archeoguide* Experience [70], in 2000: the earliest applications of an immersive AR system for archaeological sites. *Archeoguide* (Augmented Reality based Cultural Heritage On-site GUIDE) is an augmented reality-based system offering the chance to observe monuments in archaeological sites together with their overlapping reconstruction. The system has been conceived to be used as a customized guide and to provide multimedia information to visitors adapting to their characteristics and at the same time adapting the vision to the visitor's position with respect to the monuments. AG provides a mobile unit for each user consisting of a wearable computer (laptop) and a see-through

Fig. 16.4 Examples of projector-based AR: the application "Revealing Flashlight" [59] at Mercati di Traiano, Rome

head-mounted display (a viewer to be worn as a pair of glasses containing a video camera and a speaker). The project was carried out, with funding from the EU IST Framework Program by a consortium of European organizations, such as Intracom (Greece), Fraunhofer Institute (Germany), Centro de Computação Gráfica (Portugal), etc. The first archaeological site where *Archeoguide* (AG) has been implemented is the ancient Olympia in Greece with the aim of having a system available to visitors for the 2004 Athens Olympic Games.

The project (http://www.Archeoguide.it/old/) has been tested and evaluated by end users, site operators, and developers.

Drawings on Glass at Carnuntum (Austria), 2011

In the archaeological site of Carnuntum (Austria), glass walls have been located in front of ancient damaged monuments [35], with essential drawings representing the original shape of the building (https://archaeology-travel.com/top-picks/ingenious-archaeological-reconstructions/). This simple, cheap in technology and understandable solution represents the basic example of AR application for Cultural Heritage reconstruction. Such an approach implies the presence of an external medium between the user and the cultural object.

Jumieges 3D, Normandy (France), 2013

Jumieges 3D is is one of the earliest examples of AR used to visit an historical site. It was conceived as a new way of visiting the Jumieges Abbey (www.youtube.com/watch?v=k7SgLVuUgYI) (Normandy, France) with AR technology, the "nicest ruin in France," as Victor Hugo said. It appears great and very romantic, but it is quite hard to imagine the impressive size of the site before its destruction. Conseil General de Seine Maritime, owner of the site, decided to develop a new cutting-edge device for presenting the site to the public, using augmented reality and high value content. The mobile application has been developed by Art Graphique et Patrimoine (AGP) Group in 2013. The tool is available on-site and through an app to be downloaded at home. Jumieges 3D experience offers a new and spectacular way of visiting the site: superimposing, through a mobile device, the shape of the abbey as it was in five different periods, from the ninth to the eighteenth century, thanks to augmented reality. The application works both by geopositioning system and thanks to some hot spots on the ground, shaped as of decorations. The user finds the right position and then is able to activate the 360° AR contents and look at the original architecture elements. At the same time, he is constantly aware of his position on the abbey map, thanks to the GPS function.

The application works by geopositioning system and on specific hot spot placed on the ground. The user finds the right position and activates the panoramic AR contents while looking at the original architecture elements. At the same time, he is constantly aware of his position on the abbey map, thanks to the GPS function.

ViaggiArte: Fontanelle Cemetery, Naples (Italy), 2019

The Fontanelle Cemetery in Naples is a cave containing the bones of about 40,000 dead people, located in the city center [18]. ViaggiArte, developed by IRISS-CNR, CUEIM, and bbPlanet, is a AR application based on QR code positioning (GPS positioning is not available since the place is underground). The system allows to see characters, in form of ghosts, who dramatically tell users the history and insights of such a place. Since many decades, the local people celebrate the dead persons of this cemetery, giving them names and personality, creating characters, and virtually keeping in touch with them. Such a phenomenon has been studied by anthropologists. Thus, it is particularly effective to arrange a virtual representation in which actors play on the basis of a specific screenplay, through the well-known chroma key technique, which allows to transpose the contents in AR on mobile devices, as well as scenographic effects, such as the ghost transparency and the fading of virtual images. In these kinds of products, the QR code is a useful expedient to allow the positioning in an underground environment. Moreover, a high precision is not mandatory for playing the characters, as they can also move in the virtual scene without losing the perspective all around.

CEMEC: The Box of Stories, EU Project, 2015–2019

The holographic showcase has been conceived and realized by CNR-ITABC, in the frame of the CEMEC project (Connecting European Early Medieval Collections - Creative Europe Program), between 2015 and 2019 (http://www.itabc.cnr.it/progetti/cemec) [54]. It is a particular showcase aimed to re-create around the original artifact (exhibited inside), its original sensory dimension, through a new kind of dramatic approach. For this reason, it is called "the Box of Stories." Since 2017, such an application traveled through many European museums in the frame of an exhibition about the Middle Ages. The showcase has been conceived as a sort of theatre stage, providing settings for stage direction, scene devices synchronization, audio, scenography, and projections. The projection is performed through the well-known Pepper's ghost technique, spread in the eighteenth century, for theatre plays, now enhanced thanks to digital technologies. The collection objects are integrated in a mixed reality approach. Visitors can find connections with the exposed object and its virtual representation, in order to understand how and why it has been linked to its symbolic and meanings. This holographic showcase, including the artifact, allows visitors to keep the attention constantly on it. Not its virtual replica, but the original itself, is the center of our attention along the time of the whole experience: all the virtual animations and the fragments of stories originate from its real figures and details, thus creating an experience of mixed reality (see Fig. 16.2).

The Santa Maria Antiqua Videomapping, Rome (Italy), 2015

The Santa Maria Antiqua Church, on the Palatine hill slope, built in the sixth century (www.katatexilux.com/storm) AD is famous for its paintings: about 250 square meters, since Rome's foundation lasting up to the seventh century. It's a unique example for the documentation of the medieval and byzantine art. The earthquake in 847 buried such frescoes and paintings, thus saving them from the destruction. A series of multimedia installation has recently been developed by Katatexilux [12], among them a videomapping AR experience, working since 2015. The project aims at leading the visitor in reading the building history, along an emotional and scientific path. The videomapping is conceived to reconstruct and display – in an immersive approach – the lacking parts of frescoes and chapels at the sides of the presbytery (Fig. 16.5).

The Revealing Flashlight, EU Project, 2013

A further example of AR applications (no handled displays, no head-mounted displays) is a system which implies user's natural interaction. The Revealing Flashlight is a spatial AR application (https://vimeo.com/109284170) based on natural interaction, created by CNRS with the support of various experts [59]. It is conceived to light end stress non-visible details on 2D surfaces, or to display new virtual texture layers related to disappeared decorations, directly projecting them on the artifacts. Moving one finger, the user points specific areas of the object's surface. A spot of "light," revealing the hidden content, is then projected on it. Although the system may be based on a simple "2D on 2D" matching, a detailed 3D model and the preciseness of the projecting calibration are important for an accurate (re)projection of the contents on the surface. The user takes advantage of his/her finger as a *torch*, driving the spot onto the original content

Fig. 16.5 Santa Maria Antiqua videomapping, Rome 2015. Courtesy of Katatexilux

16

and decorations, exploiting a Leap motion device (Fig. 16.4, right) which allows to track the virtual torch. This application has been presented at the Allard Pierson Museum (APM), in Amsterdam, for the *Eternal Egypt exhibition* (July 12, 2013 to January 5, 2014) and then for the *Keys2Rome exhibition*, in the Rome and Alexandria venues (2014).

iMARECULTURE, EU Project, 2016–2020

iMARECULTURE – iMmersive serious games and Augmented REality as tools to raise awareness and access to European underwater CULTURal heritagE – is a European-funded project focusing on raising European identity awareness using maritime and underwater cultural interaction and exchange in the Mediterranean Sea (http://www.imareculture.eu/). One of the main goals of this project is using digital technologies to allow the wide public to virtually access unreachable maritime heritage like shipwrecks and submerged sites. Commercial ship routes joining Europe with other cultures are vivid examples of cultural interaction, while shipwrecks and submerged sites, unreachable to the wide public, are excellent samples that can benefit from immersive technologies, augmented and virtual reality. In order to achieve this goal, the project developed different technologies to be used according to users. One the one hand, it supports dry visits by providing immersive experience by means of VR Cave and 3D info kiosks; on the other hand, it enhances diver experience using underwater augmented reality accessible by tablet or underwater housing. An interesting experiment within this project was performed in the Underwater Archaeological Park of Baiae (Naples) using two Underwater Augmented Reality (UWAR) technologies to allow divers to better understand the underwater heritage and archaeological sites. The technologies used adopt hybrid tracking techniques to perform an augmented visualization of the actual conditions and of a hypothetical 3D reconstruction of the archaeological remains. The first of these UWAR technologies uses a marker-based tracking with inertial sensors, and the second adopts a markerless approach which integrates acoustic localization and visual-inertial odometry [13].

ARETE, EU Project, 2019–2023

To conclude this selection of AR projects, we must cite ARETE, (https://cordis.europa.eu/project/id/856533) Augmented Reality Interactive Educational System. The project is still ongoing and aims at promoting AR in education and at exploiting the opportunities offered by multiuser interactions with AR technologies evaluated within three different pilot studies: STEM, English literacy skills, and positive behavior intervention [43]. First investigations have been already performed, and researches describing how AR technology can be used to provide holistic education and engage students in learning STEAM subjects have been published [29]. The project is still ongoing Developments will be carried out between 2021 and 2022, while the evaluation phase will be completed in 2022.

16.3.2 Analytical-Comparative Summary of Use Cases

Figure 16.6 summarizes the main aspects of the selected case studies. The basic characteristics of each of them have been taken into consideration, both generally (year, author, type of cultural heritage, website, active/non-active project) and technically (type of multimedia containers used, software framework, devices involved, AR user positioning, and orientation system). It is relevant to say that different technologies may highlight diverse strengths and weaknesses or particular applications may not be fully representative of the technology used. The table stands as a tentative sample useful to outline the overall potential and limitations of the AR in Cultural Heritage field (see *infra* Sect. 16.4).

The case studies considered have been developed in Europe; they are based on different technologies ranging from HMD to monitors (smartphone or touch screen), from projection mapping to see-through visualizations based on holograms. All the case studies cover the last decade except the pioneering European project *Olympia*, which dates back to 2000. It is interesting that the majority of the examples have been mainly developed by specialized companies (supported by academic and scientific partners as in the case of *Jumieges* and *Santa Maria Antiqua*); the same happened for the European project *Archeoguide* and *Olympia* – where companies developed the prototype. In recent years, however, the most significant examples have been developed within European projects (iMARECULTURE, ARETE) where the focus has shifted to multidisciplinarity (creators of cultural content "validated" on scientific criteria, developers of original hardware/software solutions, content dissemination experts) and interaction between research institutes and universities and private companies. This trend marks an acceleration in the development of new solutions but also new metaphors for communication and use of AR technologies, as in the case of use in the underwater environment (iMARECULTURE). In several cases, AR has been applied in multi scale, from small contexts such as museum objects and showcases (*Box of Stories* and *Revealing Flashlight*) to architectural (*Jumieges*, *Santa Maria Antiqua*) and territorial contexts (*Olympia*, *Carnuntum*) and even expanding the traditional boundaries with use in the underwater environment (iMARECULTURE). The quality of 3D models involved and the frameworks used are closely linked to the technology available at the time of the creation of the AR application. It should be noted that the formats of the 3D models and the software used in the case of *Olympia* are now obsolete and practically unusable (VRML

Case study	**Olimpia**	**Jumièges 3D**	**Carnuntum**	**ViaggiARTE**	**Revealing flashlight**	**Santa maria antiqua videomapping**	**CEMEC – The box of stories**	**iMARECULTURE**	**Arete**
Hardware	Head-mounted display	Display	Display	Display	Projection mapping	Projection mapping	Holographic showcase	Display	Display
Year	2000	2012	2011	2017	2014	2016	2015–2019	2016–2020	2019–2023
Authors	Intracom S.A., Greece; Fraunhofer institute IGD, Germany; Center for computer graphics (ZGDV), Germany; Center for computer graphics (CCG), Portugal	Art Graphique et Patrimoine (AGP). France	7Reasons, LBI. Austria	ViaggiArte, italy	Inria, CNRS, IMN, Labri Univ. Bordeaux. France	Parco archeologico del Colosseo - Santa Maria Antiqua (Foro Romano) Progetto KatatexiLux, Università della Tuscia. Italy	CNR-ITABC and EVOCA in collaboration with CEMEC consortium. Italy, Belgium, Netherlands, Greece and Hungary	Technologiko Panepistimio Kyprou and the iMARECULTURE consortium. Cyprus, Czechia, Canada, Bosnia and Herzegovina, France, Italy, Portugal, Hungary	University college dublin, National university of ireland (IRL) in collaboration with ARETE consortium: Ireland, Netherland, United Kingdom, Italy, Spain, Belgium, German
Field of appliance	Archeological site	Archeological site	Archeological site	Cemetery	Museum objects	Historical architecture and wall paintings	Museum objects	Maritime archaeology	Education (English literacy; STEM; Positive behaviour interventions and support)
Interaction	Interactive	Interactive	Interactive	Interactive	Interactive	Not interactive	Not interactive	Interactive	Interactive and collaborative
Visual content	VRML, pre-calculated lighting	Panoramic renders (360°)	3D models	Video	Renders	Renders and computer graphic animations	Computer graphic animation and video	Computer graphic animation and video, 3D models,	3D content; Classroom and lesson information data; Activity data; Interaction data and user messages; Learning records
Audio content	–	Actors' voices, music	–	Actors' voices, music and sound effects	Computer graphic animation and video	Audio narrations and music	Actors' voices, music and sound effects	–	Audio
Device	Wearable computer and see-trough display	Mobile	Mobile	Mobile	Projectors and leap motion	Projectors	See-trough glass and monitor	Underwater tablets	HMDs, smartphones or tablets
Framework	AVALON toolkit VR, VRML and Java interface. Orientation via C++ code		Wikitude SDK	Wikitude SDK	Proprietary software	Unity 3D	VVVV	Unity 3D, ARKit	Unity 3D, Vuforia, ARLEM
Positioning / orientation system	GPS and optical tracking through *Fourier* transformation	GPS and ground bushings to recognize POIs	Image-based	QR code	Hand tracking and projector calibration with projection mapping techniques.	Manual alignment of projectors	Optical tracking calibrated with projection mapping techniques	Hybrid traking: 1) markers + inertial measurement unit (IMU) 2) markerless: acoustic localization system + visual inertial odometry	Markerless-based AR
Website	https://www.archeoguide.it/old/	https://itunes.apple.com/fr/app/jumieges-3d/id556799877?mt=8	https://www.7reasons.net	https://www.neatec.it/en/viaggiarte/	https://hal.inria.fr/hal-00986905v2	https://www.katatexilux.com/storm	https://cemec-eu.net/ https://www.itabc.cnr.it/progetti/cemec	https://imareculture.eu/ https://cordis.europa.eu/project/id/727153	https://cordis.europa.eu/project/id/856533

Fig. 16.6 Case studies selected to draft out criteria and paradigms of effectiveness of AR uses in CH contexts (see *infra* Sect. 16.3.1)

16

format) due to technical reasons and for the overall visual quality of the models.

Although there is no certainty that this will happen again in 10 years onward, technical limitations and requirements (calculation capacity of mobile devices, latency, etc. – see next section) are strictly related to interactive AR. Today, these limitations indeed affect the overall visual quality and resolution of 3D content and related production pipelines. We foresee important hardware upgrades in the next few years that will on one hand high interactive visual quality and on the other better reusability of the assets. Those applications that follow the visualization metaphor of the "augmented" content juxtaposed to reality like *Jumieges* no longer rely on 3D real-time models but on panoramic renderings (static images or video) which allow to achieve a very high visual quality. Such content, even after 7–8 years, maintains its potential reusability in new applications, whereas in other case studies, different techniques have been used to estimate the positioning and orientation of the user during the AR experience: for the juxtaposition of virtual and real, the GPS (and the magnetometer) has been considered suitable and sufficient at the scope (*Carnuntum*), while for interactive monitor-based AR applications, or see-through AR, position and orientation are calculated using different image recognition algorithms. In general, this grants a sufficient tracking robustness during the experience.

Visualization, as already mentioned, is one of the major topics emerging from analyzed case studies. Usability has been generally evaluated as a strength point in case studies of Fig. 16.6: they were considered simple and immediate to use. This aspect support the effectiveness and intuitiveness of the system which is inherent in the AR metaphor (see *infra* Sect. 16.6). Nevertheless, *Olympia* prototype was rightly considered too uncomfortable because of the backpack and helmet-mounted camera, while the HMD viewers used outdoors with AR juxtaposition metaphor lead to usability issues due to excessive overheating and battery consumption.

Generally about AR experiences, in the case studies where relevant content and storytelling technique have been applied, this turned to be a positive aspect in terms of perceived realism of virtuality and effectiveness for the users (*Olympia*, *Box of Stories*, *Santa Maria Antiqua*). The strengths or weaknesses of applications' content pertained mainly to the scientific validation obtained by experts in the field – who guaranteed their authoritativeness (*Jumieges*, *Box of Stories*, *Santa Maria Antiqua*); other important issues have been the inclusion of audio and soundscape in the storytelling as well as the simplicity of texts (*Box of Stories*); less successful in terms of originality seemed to be texts with a more descriptive vocation, like audio-guide style. Last interesting aspect is the chance to see these applications still running or not (all of them except for *Olympia*) which testifies how this approach allows a certain longevity to the use cases related to AR.

16.4 Technical Limitations and Challenges of AR for CH

This section presents an overview of technical challenges, issues, and limitations typically faced in previously described case studies and in general literature that should be taken into account during the design and development of AR applications for CH.

For the sake of clarity, the scheme in Fig. 16.7 sorts AR devices depending on display medium proximity to the user. Such classification is also helpful to easily highlight *physical invasiveness* of specific AR technologies (e.g., Is the device wearable? Is it handheld?):

– **Class A**: optical see-through (OST) and video see-through (VST) devices or AR contact lenses with the display medium very close to users' eyes. This category includes devices, tools, and systems with a high degree of "intimacy" with the user.
– **Class B**: handheld AR devices or manipulable systems within the range of approximately 1 mt.
– **Class C**: spatial see-through frames/panels applicable to specific contexts that users find in spot locations.

Field of view (FoV): in order to obtain a consistent fruition during the AR experience, the device should allow the virtual content to be presented inside a good portion of the user's field of view (FoV), considering 220° as an approximate horizontal FoV limit of human vision. Within see-through AR devices for instance (class A, see Fig. 16.7), the first version of Microsoft HoloLens (https://www.microsoft.com/it-it/hololens) has a relatively limited FoV (around $30 \times 17°$), while the new HoloLens 2 offers a wider range (43×29). The FoV enlargement at hardware level indeed should take into consideration also device weight and form factor, to maintain a proper comfort during the session. Consumer-level see-through AR devices (also including Magic Leap One) actually provide a limited FoV compared to consumer-level HMDs for immersive VR (such as Oculus Rift, HTC Vive, SONY PlayStation VR, etc.). The limited field of view of see-through AR devices, therefore, has to be taken into consideration when designing the experience (Fig. 16.8).

Vergence-accommodation conflict (VAC): within stereoscopic see-through AR (class A), the so-called *vergence-accommodation conflict* issue [69] occurs when the user brain receives mismatching cues between the distance of a virtual 3D object (vergence) and the focusing distance (accommodation) required for the eyes to focus on that object. In order to solve this problem, new solutions are being investigated hardware-wise [3, 21, 33] mostly dealing with special lenses or materials that offer multifocal planes. Such issue affects all consumer-level HMDs, including see-through AR displays and those employed for immersive VR.

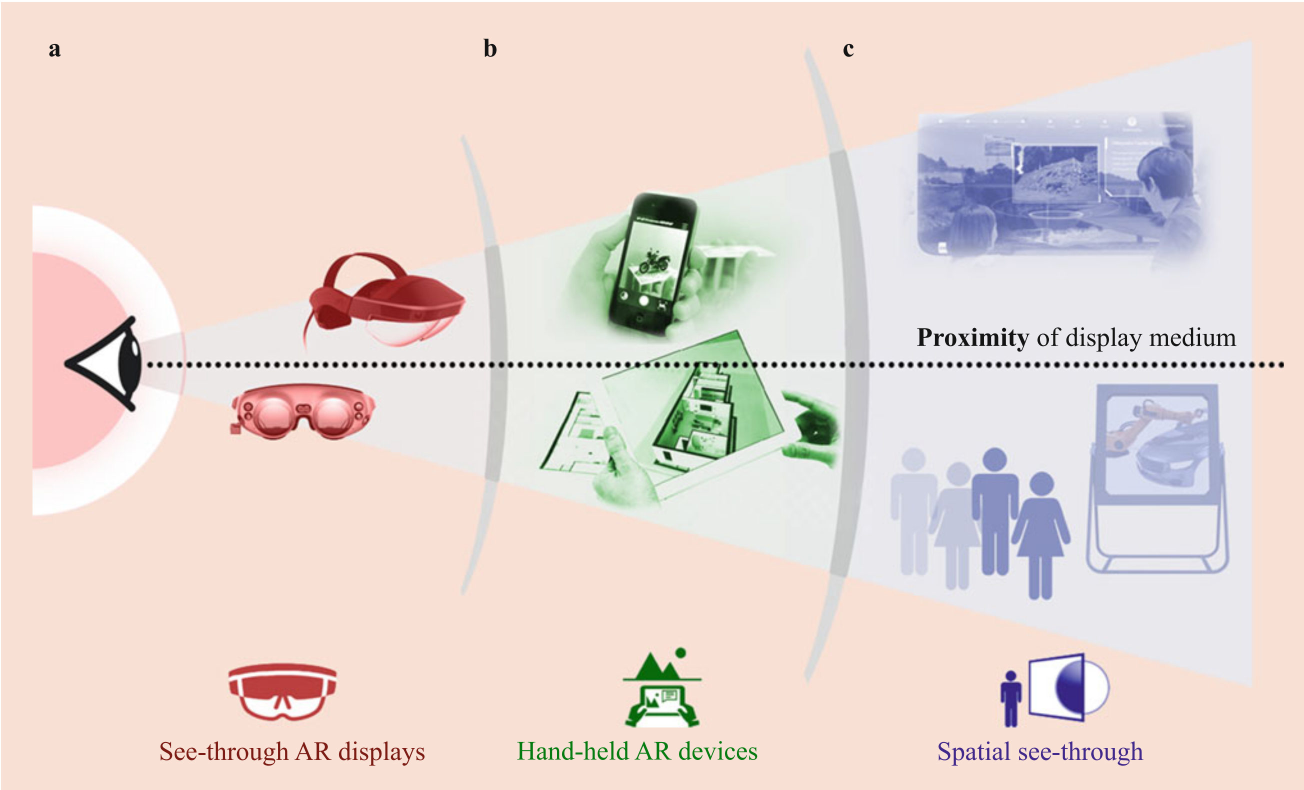


Fig. 16.7 AR devices classified by display medium proximity

Latency: one of the most prominent challenges within the discussed context is represented by the computational load required by the device to orient/position the virtual layer on top of the real one (classes A and B). Such issue results in perceived latency and discrepancies during the AR experience, thus affecting its consistency and interactivity. Specifically, real-time processing of (potentially complex) input data is required to guarantee a proper tracking and response to real context and its variations [58]. These problems are even more prominent in consumer mobile contexts (class B – handheld AR through smartphones/tablets) where the low latency is a strong requirement to produce a convincing and effective AR experience for the user. Intensive computation is generally needed in fact, for both *marker-based* AR (recognition of specific markers arranged in the real world – and *markerless* AR (exploitation of features, patterns, or internal device sensors that allow to overlap the virtual content to real world). Within marker-based AR sector, recent advancements of web technologies are allowing the development of frameworks such as AR.js (https://github.com/jeromeetienne/AR.js/blob/master/README.md) to deploy online AR apps using common mobile web browsers, maintaining low latency. Regarding web deployment and markerless AR in Cultural Heritage, the mobile device compass can be exploited in specific museum locations to augment rooms and objects with matching orientation [7], removing all heavy computations involved in feature recognition and tracking.

Occlusion: AR content is generally overlapped to the real-world imagery, for example, via video or optical see-through displays (class A) or using handheld devices (class B). However, such overlays are not effective when displaying data in 3D, since occlusion between the real and computer-generated objects is not addressed. In literature, in order to solve this issue, *depth* has to be taken into consideration [23, 28, 71] to guarantee a consistent rendering that allows real objects to occlude virtual ones.

Ambient lighting: the ambient lighting of the real context generally plays an important role. Both marker-based and markerless AR that depend on feature tracking may suffer from real ambient conditions: scarce or sudden lighting changes may strongly affect tracking or pattern recognition, thus impacting user experience. Furthermore, excessive or extreme lighting conditions (e.g., outdoor contexts) may also affect vision clarity on the display (class A).

Virtual lighting: there are several research studies that focus on virtual lighting (see, for instance, [1, 2]) to create a consistent rendering of virtual layer on top of the real one. Little importance in fact is generally given on replicating the visual qualities of the real-world scene in a plausible manner. Even though the virtual object is correctly tracked and placed

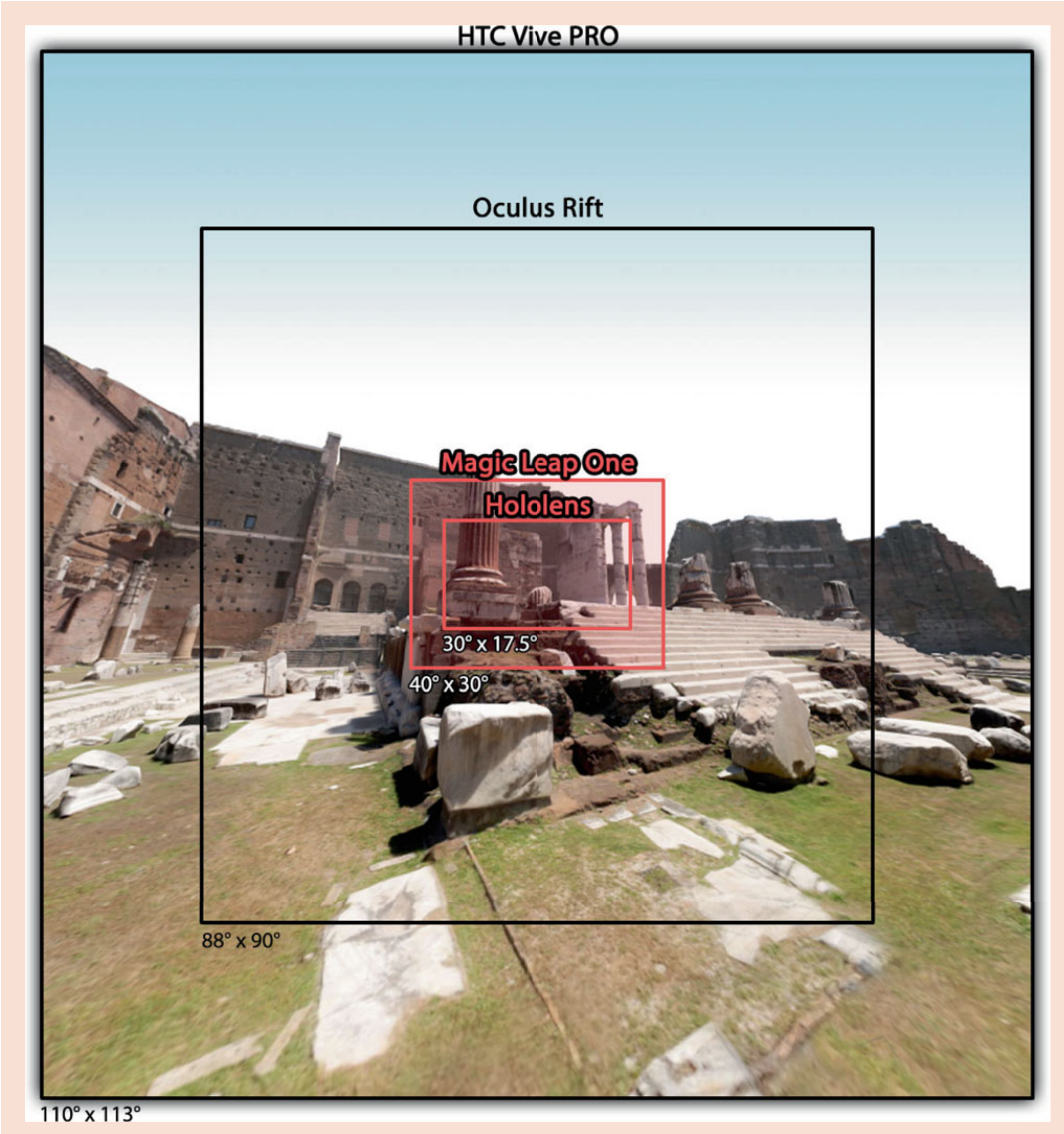


Fig. 16.8 A comparison between different HMD FoV to present virtual content

in the real context, its visual appearance immediately gives away its artificiality. This issue creates a visual mismatch, since materials and lighting are not properly simulated with respect to surrounding real environment. The process of acquiring lighting information can be complex: this is the reason why approximate methods are generally adopted to *infer* a lighting model for virtual content.

Connectivity: depending on the AR device used and the specific experience, virtual content can be served by remote locations (via local hot spot arranged in the real space, or via Internet connection). This may represent a bottleneck when dealing with large amounts of visitors in a museum or archaeological site and/or the network connection is slow or absent. Such challenges are faced in both these cases:

– When downloading a stand-alone application once at the beginning of the experience (e.g., handheld AR devices, class B)
– When using a streaming approach along predefined areas, usually via Wi-Fi networks or hot spot

Content optimization: producing virtual content for AR (generally 3D objects, artifacts, buildings, etc.) must be carried out with proper care from a technical point of view. The limited rendering capabilities of certain see-through AR displays or mobile devices (e.g., classes A and B) require proper workflow to optimize 3D content, thus maintaining interactive frame rates. Content creators have to follow specific guidelines (partially overlapping with those adopted in serious games 3D production pipelines) including optimization of geometries, controlled number of polygons, texture optimization, and pre-computation of specific effects.

Calibration: just like HMD for immersive VR, see-through AR devices typically require preliminary routines for calibration and setup in order to provide a correct experience. These routines are generally performed for each user, like the cases of Microsoft HoloLens and Magic Leap One. Regarding handheld AR with markerless tracking, preparatory steps can be needed to build virtual 3D reference frames where artificial content will take place. For spatial see-through displays (class C), preliminary setups are indeed required to align virtual content to the real context (museum, archaeological site, etc.).

16.5 Multisensorial AR Experiences

A natural evolution for AR applied to Cultural Heritage can surely be the **empowerment of multisensory experiences** in virtual, augmented, and mixed environments. Indeed, since the early stages and applications of VR in Cultural Heritage, the interest toward multisensorial immersive environments has been high, but in the last decade, the advances in the fields of multisensorial AR frameworks seem to open new perspectives.

In order to let the cultural visit experience be effective and overwhelming, the exchange between the user and the place of Culture needs to take place from the first eye contact. The meaning of this exchange is linked to the perception of *stimuli* that reach our senses and which, for evolutionary reasons, are linked to the need to understand if the surrounding object/person/space constitutes a threat or not; such perception is processed extremely fast (few seconds).

Multisensoriality in AR, or what is called by Covaci et al. "Mulsemedia" [16], takes indeed into account:

1. **Perceptual-motor functions**: set of positional sensations which, together with motor programming and muscle contraction/relaxation, generate movements in the VR world.
2. **Attention**: cognitive process that allows to select some environmental *stimuli* among the many available in the VR world, at a given moment, and to ignore others.
3. **Executive functions**: functional modules of the mind, which regulate the planning, control, and coordination processes of the cognitive system when we experience a VR world.
4. **Learning and memory**: process through which new information about the VR world around us is acquired and stored, which involves a behavioral modification in response to internal and external *stimuli* [42].
5. **Language**: set of symbolic codes (of a verbal or nonverbal nature) that allow digital information to be transmitted, stored, and processed.
6. **Social cognition**: subjective interpretation that the person constructs with respect to his own social environment, considering the configuration of the factors inherent to the person himself and the situation in which he finds himself acting.

That is why, for a new empowered multisensorial AR experience, we need to pursue [9, 60, 61]:

1. **An interactive approach**: possibility of using devices for widespread interactivity which take into account touch (e.g., touch screen), smell (e.g., environmental speakers), sight (e.g., eye tracking), and movement through sensors (e.g., Kinect, Leap Motion, webcam) in a combined and multi-layered manner;
2. **The fusion of different communication formats**: integration of different expressive and artistic modalities for the reading and understanding of objects/places of Culture, through the use of soundscapes, video graphics, 2D and 3D graphics, comics, animations, contents accessible through immersive reality expedients, stimulation of the olfactory system, large-scale projections, etc.

Examples of AR devices enabling touch sensations can be seen in [49]. Author actually developed a mobile platform to augment both the sense of touch, sight, and hearing and all of them in combination to constitute what is called a multisensory experience. Two pieces of tactile wearable interfaces (a vest and a glove), made in soft material, enabled users to easily move around them. The glove consists of three sensors that send signals to the computer device and allows for artificial sensory augmentations. The glove allows the user to browse, select, and move augmented visual objects while transferring vibrations to reach the fingers via three vibrotactile actuators. Two other devices for stimulating the touch sense are in [57]. They are the "RingU," a wearable finger device, and "Kissenger," which is a sensor type of plug-in for mobile devices to be laid on top of the display, and to then be interacted with by the user.

Research addressed to technological possibilities of stimulating the chemical senses, comprising smell and taste, has been done too. One example is [63] who researched on the trigeminal sense: it is related to humans' ability to sense heat and cold through the skin. There exist two ways of digitally stimulating the chemical senses: a) with the release of a chemical substance, triggered by a digital component such as AR and by stimulating the taste buds directly through an electrical component, and b) the other happens electrically without any chemical substances being released, and triggered digitally by AR too. However, several limitations come along the abovementioned solutions. For instance, when it comes to taste, then some taste senses are easier to target than others, and also some people are more sensitive than others; therefore, it is difficult to make a universal AR taste stimulation [63].

The last decade also witnessed a significant growth in the field of olfactory research [47]. Cultural Heritage is probably a domain in which the application of this kind of approach is easier. As a matter of fact, the problem related with olfactory reproduction is focused on the entire smell transmission pipeline [40]: the three phases of capturing and reading the smell, formalizing it in a generalized code, and reproducing it for the end user is really challenging. Nevertheless, in context of ancient landscape reconstructions, the effort could be limited to the last step, dealing with artificially defined smells. The simulation of scents and smells has to face issues related to odor reproduction end their perception [72, 73]. Limiting the analysis to the off-line and on-site applications, a museum can arrange quite easily an olfactory VR/AR experience, as has been done for some projects (Faruolo F., “Scented theatre: profumi in scena”, www.smellfestival.it/2019/12/27/scented-theatre-profumi-in-scena).

We are probably at the beginning of a long way to be done in this direction, as a series of theoretical questions are still to be solved, such as “Which dimension do we want to give to past scents and smells?”, “What emotions we intend to stimulate in users?”, and “How much comparable are ancient smells with existing and reproducible ones?”. We have not gone so far answering these questions, but they will probably bring us toward new meanings of multisensorial impact of virtual reconstructions in the Cultural Heritage domain.

The technological advancements within AR will naturally entail looking at this in light of the different types of devices that are needed to make possible the enhancements of **all five senses**.

16.6 Efficacy and Effectiveness of AR

As presented in Sects. 16.3 and 16.4, AR differs from VR because the focus of the interaction of the performed tasks leverages on the interplay between real and virtual contexts. AR therefore offers the opportunity of a digital overlay onto a real or virtual environment, proving users of extra layers of visualization (and further understanding of the Cultural Heritage objects and sites). For the user, virtual and physical layers are blended in such a way that an immersive, interactive environment is experienced. An intentional design of the user experience is of primary importance because there is no, technology better than another, but what is fundamental is to evaluate the context, the environment, the scenario of use, the type of target, what we want to communicate, and the sustainability of such technology. On this base, we move toward the most conscious and harmonious choice. All content will be modeled accordingly. Nevertheless, how to design the best user experience for AR applications? Which are the issues to be taken into account for a fruitful and efficient cultural “augmented” experience?

16.6.1 The Importance of User Experience Design for AR Systems

The experience of AR applications must necessarily take into account:

1. The type and structure of the content
2. The times of fruition
3. The modalities of fruition
4. The place of fruition
5. The conditions of fruition

All these variables obviously influence the choice of technology that a developer might decide to use from time to time. In general, it is clear that the design of the user experience is strategic not only for the purpose (a) of the aesthetic enjoyment of the user in front of the real cultural asset reconstructed in 3D but also (b) for the stimulation of curiosity toward the story being told through AR systems and finally, (c) for the comprehensibility of the information, values and meanings referable to each storytelling of the Past, being them landscapes, historical characters, and artifacts. The design of the user experience in AR therefore has its focus on:

1. The **solicitation of a feeling or emotion** in those who observe the scenario reconstructed in 3D, made up of characters, objects, and a particular setting as well as the rhythm of the story;
2. **Stimulating users to actively participate** in the AR experience and allowing them to feel comfortable while facing the multimedia application so that interaction can take place;
3. **Triggering an indirect but equally effective educational mechanism** as regards the understanding, storage, and elaboration of the content experienced during the use of the multimedia application.

From a cognitive point of view, the design of AR applications must be aimed at facilitating that step of the learning process which is usually activated when observing something that refers to the Past: abstraction [22] (Fig. 16.9).

During numerous surveys conducted on visitors to cultural sites [52, 53, 55, 56], it has been observed that a prevalent part of the visitors is initially very intrigued by new technologies, especially AR, and wants to try them. However, once the surprise or discovery effect aroused by technology is exhausted, attention tends to decline very quickly if any attractive, meaningful contents are not presented. In order to maintain a high level of attention, involvement, and motivation, working on content is essential: experimentation on narratives, representation and dramatization forms, and new storytelling techniques are advisable in the Cultural Heritage field. Cultural content, when experienced in coexistence with artifacts and sites, must be designed to continuously encourage and stimulate users to make comparisons between today's situation and the past one, in a diachronic and multilevel vision, so to "read," "translate," and finally "understand" the heritage.

16.6.2 Efficacy and Effectiveness of AR Systems in Relation to the Context of Use

The great effectiveness of Mixed Reality consists in the fact that Real and Virtual no longer exist in separate spaces and times but coexist conceptually and "physically" as a "continuum" [45], becoming synergistic components of our lived space. Thus, the use of the Virtual as an enhancement of the reality and multiplication of its communicative potentiality finds its maximum effectiveness in mixed reality experiences. The environmental and logistical variables largely influence the condition of use of such systems and bring designers and developers to very different choices in the conception of applications and in the adoption of specific technological solutions. Reality, thanks to the coherent superimposition of "virtuality," becomes multilayered, stratified, in continuous potential transformation, a "canvas" on which forms and meanings can be built or found, and a series of possible and potential representative, interpretative, and imaginative levels. This potentiality can be expressed both in the artistic

16

Fig. 16.9 Personal elaboration of the experiential model of [32] and [17]

domain and in the field of study and valorization of a cultural site, adhering to its original meaning. Basically, a new convergence of sciences, arts, languages, and technologies is possible today, as well as the fusion between real and virtual worlds, real actors, digital characters, theatrical techniques, and gaming paradigms. In a mixed reality environment, the accuracy and consistency of the overlapping between real and virtual content are crucial conditions to guarantee the intelligibility and "credibility" of the experience. A misalignment or latency would be immediately perceived by the user and would be disorienting and annoying. Here are some basic criteria in order to guarantee a correct overlapping of real and virtual content.

1. **3D graphic content** produces the most powerful impact in a mixed reality experience. Thus, a preliminary topographic and volumetric survey of the real object on which the virtual projection will intervene cannot be ignored. Starting from the 3D model of the real object, the virtual contents will be shaped in order to obtain a perfect correspondence. The superimposition of the real and virtual views is achieved by matching the orientation of the virtual camera with the user's point of view, thanks to appropriate tracking systems;
2. The virtual contents must be superimposed on the real space so they must be placed on a transparent or black **background**;
3. The **scale** of the virtual projection must be correctly proportioned to create a precise match between real and virtual content;
4. **Camera constraints**: due to the overlapping between real and virtual worlds and the sense of presence, the user should live within the mixed reality, and the virtual camera position must match the position of the visitor's eyes. A multi-camera could take the visitor into a different paradigm, more related to cinema;
5. **Field of view (FOV) and depth of field**: the fields of view of real and virtual must be coherent, and everything in the scene must be in focus, due to the illusion of reality;
6. **Matching lights and colors**: it is very important that the virtual lights and shadows correspond to the real ones in terms of kelvin degree, intensity, and color, at least if the aim is to produce an effect of realism.
7. **Sound spatialization**: 3D soundscape is fundamental to create perceptual credibility, coherently with visual perception and FOV.

There are different technical approaches and technologies, more or less immersive, implementing such kind of "augmented" experience [45]. The immersive vision through HMD, enhancing the sense of presence and embodiment [68], brings the user into a completely new way of experiencing artificial environments, in comparison with any other VR desktop system [8]. Movement and tracking have to deal with several factors such as stereo vision, binocular distance, position of objects in 3D space, scale factor, user's speed of movement in the space, proprioception, and sickness [55]. For this reason, their implementation has to be very well designed in relation to the context of use [51].

In general, the context of use is a fundamental variable that must be taken into account when planning the kind of technology to be adopted, the complexity of interaction, the audio-visual languages, the duration of the contents, and the rhythm of multimedia communication.

If the application is used on the cultural heritage site, standing still in a fixed point and the user simply turns his eyes around, it is sufficient to have an efficient geo-location and orientation system. In this case, motion tracking is an option that can enrich the sense of immersion and presence, but especially when applied to an immersive display in outdoor spaces, it is renounceable. Indeed, not all users wearing immersive display react positively to motion tracking (it also depends on the technical and perceptive criteria followed in the implementation; For some of them the experience can even cause sickness, loss of balance and vertigo. HMDs that currently support motion tracking are more sophisticated than those ones that do not. They need maintenance and they can create management problems in outdoor public places (battery charging and overheating, connection to a PC by means of cables in case of 3D models with high resolution). However, in this domain, technology evolves very rapidly and this critical aspect will be overcome in the short term. If the MR experience takes place while the user is moving (walking, driving, etc.), immersive display is not recommended for safety and cognitive reasons. During the movement, the experience will be preferably accessed through a display monitor (Window on the World – WoW). Besides, at present times, tracking features are difficult to be managed in mobility, for the following technical reasons: if the digital content consists in written captions or "labels," the need for a perfect collimation is less stringent, but if the virtual content is a "reconstructive" graphic, the problem becomes more serious.

From a contextual point of view, being along a visit path (archaeological site, monument, museum) requires the users to stop and walk in a space that is already plenty of contents and *stimuli*; visitors are inevitably conditioned by environmental factors such as light, air, temperature, crowding, noise, etc.; the visiting time may be limited or influenced by being in a group; and objects in their hands can inhibit certain movements. The experience is therefore mainly real; it is "augmented" through the superimposition of reconstructive and narrative virtual contents that helps to understand what user is observing. In these cases, monitor-based systems like tablet or smartphone (WoW) can be preferable, more flexible in their usage, shareable, and more relaxing.

In conclusion, immersive displays can provide efficacious, and maybe more powerful, experiences in terms of perception, but not while moving in a cultural space and with specific constraints. A good solution is to highlight, in the real space, points of interest where the visitor is invited to stop and enjoy AR/MR experiences. The enjoyment of AR/MR systems along a visit path, using both monitor-based systems and HMDs, requires the creation of micro-experiences, to be lived with immediacy, distributed in the physical space. Interaction should be limited only to the orientation in the 360° space. Any type of narrative must be conceived in the form of "pills," meaningful but short.

In the case of projection mapping (projector-based augmented reality), images projected directly onto architectural structures or objects are very effective, both in terms of immediacy of understanding and of spectacular impact. This solution let audience not to wear anything and not to physically interact with technological devices. Therefore, usability is maximized. The virtual images overlap one to another interacting with the physical volumes and their material consistency and texture, highlighting scantly perceptible traces, giving back to the naked surfaces their ancient decorations, and bringing the invisible to light. It is a kind of restoration of the light through the light. The latter can guide perception of fragmentary contexts, if wisely used and dosed. In this kind of applications, professional projectors are needed, with good brightness and durability and good resistance to non optimal environmental conditions (humidity is usually the main problem). Thus, the sustainability of these technologies must be carefully assessed, especially when they are installed for long or permanent exhibitions.

16.7 Directions and Future Perspectives

Communicating Cultural Heritage, making it readable and comprehensible to a wider and differentiated audience, is the result of a profound stratification of knowledge which cannot rely only upon technological employment. Through a well-planned communication strategy for Cultural Heritage is it possible to bring users closer to their territories and artifacts, to stimulate a sense of belonging, enriching everyone's cultural background in a transparent and immediate way.

The experience of the Past becomes a story to be told that in turn becomes the "new" heritage of the community – and therefore a new cultural element. How the use of AR can facilitate this transformation process? By studying the visitor's needs, behaviors, and the contexts of technological and cultural fruition.

In general, the analysis of the entire AR experience becomes a fundamental prerequisite for a policy of fruition of Cultural Heritage that does not consider the public a passive subject, but a lively agent with which to interact, a protagonist in the production of "new" Culture and a vehicle for its dissemination. Specifically, we envision for the next futures four possible scenarios of enhancement for AR applications:

A. **Communication.** With technological evolution of digital devices and "augmenting" techniques, content creation will need to adjust and update its own paradigms and communication strategies too. This will also enable the elaboration of new format and standards for AR storytelling and for designing new graphic user interfaces (GUIs) able to attract visitors' attention and powerfully convey the cultural message. Media hybridization is rapidly evolving: virtual reality, cinematography, theatrical paradigms, gaming rules, artificial intelligence, artificial life, digital interfaces working with sensors, sound design, and haptic systems will be more and more combined, to create novel kinds of experiences where virtual and real work as "a continuum." The concept of "authenticity" will evolve as well, no more, or not only, connected to the materiality of the surrounding world.

B. **Convergence of AR and VR.** See-through AR head-mounted displays dealing with stereoscopic content, such as HoloLens 2 or Magic Leap, are showing an incredible potential when properly used. Despite their current limitations (see Sect. 16.4), a profound integration between AR and VR segments – including revolutionized form factors and display technologies – can be foreseen during the upcoming years. The introduction of open standards like OpenXR (https://www.khronos.org/openxr) for desktop applications [44] and WebXR (https://www.w3.org/TR/webxr/) for web applications [39] represents concrete efforts towards the complete unification between the two worlds. The introduction of such standards has two main objectives: (a) extend the range of supported input including VUI (voice-user interface) and gesture-based interaction and (b) create a rock-solid technical ground for developers to craft XR experiences, allowing content creators to integrate real-world media with contextual overlays.

C. **Artificial Intelligence (AI).** Another active research direction for AR during the last couple of years is represented by *artificial intelligence* (AI): new solutions related to object classification and estimation of lighting conditions of real contexts (see limitations in Sect. 16.4). A strong interest, indeed relies on *deep learning* approaches [34], exploited to deal with complex and dynamic environments (sudden changes of lighting conditions, etc.) and to create a more realistic, seamlessly integrated and aesthetically pleasing rendering of the virtual layer. Recent studies [36, 41] are investigating such direction for AR/MR segment, in order to estimate lighting models from limited field of views (see Sect. 16.4).

D. **Internet of Things (IoT).** An interesting perspective seems to be using AR for making visible – through optical metaphors – the data obtained by sensors in an *Internet of Things* (IoT) context. For instance, during a technical survey, technicians will be able to see directly on the building walls the structural problem and potential damages reported by sensors in real time.

References

1. Agusanto, K., Li, L., Chuangui, Z., Sing, N.W.: Photorealistic rendering for augmented reality using environment illumination. In: Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 208–216. IEEE, Piscataway (2003)
2. Aittala, M.: Inverse lighting and photorealistic rendering for augmented reality. Vis. Comput. **26**(6–8), 669–678 (2010)
3. Akşit, K., Lopes, W., Kim, J., Shirley, P., Luebke, D.: Near-eye varifocal augmented reality display using see-through screens. ACM Trans. Graph. **36**(6), 1–13 (2017)
4. Angelopoulou, A., Economou, D., Bouki, V., Psarrou, A., Jin, L., Pritchard, C., Kolyda, F.: Mobile augmented reality for cultural heritage. In: International Conference on Mobile Wireless Middleware, Operating Systems, and Applications, pp. 15–22. Springer, Berlin (2011)
5. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. **6**(4), 355–385 (1997)
6. Azuma, R.T.: The most important challenge facing augmented reality. Presence Teleop. Virt. **25**(3), 234–238 (2016)
7. Barsanti, S.G., Malatesta, S.G., Lella, F., Fanini, B., Sala, F., Dodero, E., Petacco, L.: The winckelmann300 project: dissemination of culture with virtual reality at the capitoline museum in rome. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. **XLII-2**, 371–378 (2018). https://doi.org/10.5194/isprs-archives-XLII-2--371-2018
8. Bekele, M.K., Pierdicca, R., Frontoni, E., Malinverni, E.S., Gain, J.: A survey of augmented, virtual, and mixed reality for cultural heritage. J. Comput. Cult. Herit. **11**, 7:1–7:36 (2018)
9. Bilbow, S.: Developing multisensory augmented reality as a medium for computational artists. In: Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '21. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3430524.3443690
10. Billinghurst, M., Duenser, A.: Augmented reality in the classroom. Computer **45**(7), 56–63 (2012)
11. Boi, M., Boi, P.: Ingegneria elevato(n). Ingegneria del futuro o futuro dell'ingegneria? Dei Merangoli Editrice (2017). https://books.google.it/books?id=e_brtAEACAAJ
12. Borghini, S., D'Alessio, A., Scoccianti, M.: Aureo filo. La prima reggia di Nerone sul Palatino. Soprintendenza archeologica di Roma. Mondadori Electa (2019). https://books.google.it/books?id=92K1xAEACAAJ
13. Bruno, F., Barbieri, L., Mangeruga, M., Cozza, M., Lagudi, A., Čejka, J., Liarokapis, F., Skarlatos, D.: Underwater augmented reality for improving the diving experience in submerged archaeological sites. Ocean Eng. **190**, 106487 (2019)
14. Caudell, T., Mizell, D.: Augmented Reality: An Application of Heads-Up Display Technology to Manual Manufacturing Processes, vol. 2, pp. 659–669. IEEE, Piscataway (1992). https://doi.org/10.1109/HICSS.1992.183317
15. Choudary, O., Charvillat, V., Grigoras, R., Gurdjos, P.: March: mobile augmented reality for cultural heritage. In: Proceedings of the 17th ACM International Conference on Multimedia, pp. 1023–1024 (2009)
16. Covaci, A., Zou, L., Tal, I., Muntean, G.M., Ghinea, G.: Is multimedia multisensorial? – a review of mulsemedia systems. ACM Comput. Surv. **51**(5) (2018). https://doi.org/10.1145/3233774
17. De Freitas, S., Jameson, J.: Collaborative e-support for lifelong learning. Br. J. Educ. Technol. **37**(6), 817–824 (2006)
18. Del Verme, L.: Un itinerario 'ad patres'. il cimitero delle fontanelle a napoli tra tecnologia e creatività. Archeologia e Calcolatori **30**, 405–421 (2019)
19. Dewey, J.: Art as Experience. Penguin, New York (2005)
20. Du, C., Chen, Y.L., Ye, M., Ren, L.: Edge snapping-based depth enhancement for dynamic occlusion handling in augmented reality. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 54–62. IEEE, Piscataway (2016)
21. Dunn, D., Chakravarthula, P., Dong, Q., Fuchs, H.: Mitigating vergence-accommodation conflict for near-eye displays via deformable beamsplitters. In: Digital Optics for Immersive Displays, vol. 10676, p. 106760U. International Society for Optics and Photonics, Bellingham (2018)
22. Fedeli, M., Frontani, L., Mengato, L.: Experiential learning. Metodi, tecniche e strumenti per il debriefing: Metodi, tecniche e strumenti per il debriefing. Cubo Rosso. Franco Angeli Edizioni (2015). https://books.google.it/books?id=JYr-CQAAQBAJ
23. Feng, Q., Nozawa, T., Shum, H.P., Morishima, S.: Occlusion for 3d object manipulation with hands in augmented reality. In: VRST '18: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (2018)
24. Fritz, F., Susperregui, A., Linaza, M.T.: Enhancing cultural tourism experiences with augmented reality technologies. In: 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST) (2005)
25. Furht, B.: Handbook of Augmented Reality. Springer Science & Business Media, Berlin (2011)
26. Genç, Ç., Soomro, S., Duyan, Y., Ölçer, S., Balcı, F., Ürey, H., Özcan, O.: Head mounted projection display & visual attention: visual attentional processing of head referenced static and dynamic displays while in motion and standing. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 1538–1547 (2016)
27. Guidazzoli, A., Liguori, M.C.: Realtà virtuale e beni culturali: una relazione in evoluzione vista attraverso i progetti sviluppati presso il cineca (2011)
28. Holynski, A., Kopf, J.: Fast depth densification for occlusion-aware augmented reality. ACM Trans. Graph. **37**(6), 1–11 (2018)
29. Jesionkowska, J., Wild, F., Deval, Y.: Active learning augmented reality for steam education—a case study. Educ. Sci. **10**(8), 198 (2020)
30. Jung, T., Tom Dieck, M.C.: Augmented reality and virtual reality. In: Empowering Human, Place and Business. Springer International Publishing, Cham (2018)
31. Kohn, V., Harborth, D.: Augmented reality – a game changing technology for manufacturing processes? In: Bednar, P.M., Frank, U., Kautz, K. (eds.) 26th European Conference on Information Systems: Beyond Digitization—Facets of Socio-Technical Change, ECIS 2018, Portsmouth, June 23–28, p. 111 (2018). https://aisel.aisnet.org/ecis2018_rp/111
32. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development. FT Press, Upper Saddle River (2014)
33. Kruijff, E., Swan, J.E., Feiner, S.: Perceptual issues in augmented reality revisited. In: 2010 IEEE International Symposium on Mixed and Augmented Reality, pp. 3–12. IEEE, Piscataway (2010)

34. Lalonde, J.F.: Deep learning for augmented reality. In: 2018 17th Workshop on Information Optics (WIO), pp. 1–3. IEEE, Piscataway (2018)
35. Ledermann, F., Schmalstieg, D.: Presenting past and present of an archaeological site in the virtual showcase. In: International Symposium on Virtual Reality, Archaeology, and Intelligent Cultural Heritage, pp. 119–126 (2003)
36. LeGendre, C., Ma, W.C., Fyffe, G., Flynn, J., Charbonnel, L., Busch, J., Debevec, P.: Deeplight: learning illumination for unconstrained mobile mixed reality. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5918–5928 (2019)
37. Li, X., Yi, W., Chi, H.L., Wang, X., Chan, A.P.: A critical review of virtual and augmented reality (VR/AR) applications in construction safety. Autom. Constr. **86**, 150–162 (2018)
38. Lopes, P., You, S., Ion, A., Baudisch, P.: Adding force feedback to mixed reality experiences and games using electrical muscle stimulation. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2018)
39. MacIntyre, B., Smith, T.F.: Thoughts on the future of webxr and the immersive web. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 338–342. IEEE, Piscataway (2018)
40. Marco, S., Gutierrez-Galvez, A., Lansner, A., Martinez, D., Rospars, J.P., Beccherelli, R., Perera, A., Pearce, T., Verschure, P., Persaud, K.: A biomimetic approach to machine olfaction, featuring a very large-scale chemical sensor array and embedded neuro-bio-inspired computation. Microsyst. Technol. **20** (2014). https://doi.org/10.1007/s00542-013-2020-8
41. Marques, B.A., Drumond, R.R., Vasconcelos, C.N., Clua, E.: Deep light source estimation for mixed reality. In: Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018), pp. 303–311 (2018)
42. Marto, A., Melo, M., Gonçalves, A., Bessa, M.: Development and evaluation of an outdoor multisensory ar system for cultural heritage. IEEE Access **9**, 16419–16434 (2021). https://doi.org/10.1109/ACCESS.2021.3050974
43. Masneri, S., Domínguez, A., Wild, F., Pronk, J., Heintz, M., Tiede, J., Nistor, A., Chiazzese, G., Mangina, E.: Work-in-progress–arete-an interactive educational system using augmented reality. In: 2020 6th International Conference of the Immersive Learning Research Network (iLRN), pp. 283–286. IEEE, Piscataway (2020)
44. McGrath, O.: Ar/vr strategy considerations for academic computing services. In: Proceedings of the 2019 ACM SIGUCCS Annual Conference, pp. 15–18 (2019)
45. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **77**(12), 1321–1329 (1994)
46. Mourkoussis, N., Liarokapis, F., Darcy, J., Pettersson, M., Petridis, P., Lister, P., White, M.: Virtual and augmented reality applied to educational and cultural heritage domains. In: Proceedings of Business Applications of Virtual Reality, Workshop (2002)
47. Nakamoto, T.: Human Olfactory Displays and Interfaces: Odor Sensing and Presentation (2012). https://doi.org/10.4018/978-1-4666-2521-1
48. Noh, Z., Sunar, M.S., Pan, Z.: A review on augmented reality for virtual heritage system. In: International Conference on Technologies for E-Learning and Digital Entertainment, pp. 50–61. Springer, Berlin (2009)
49. Nurminen, A.: Mobile multisensory augmentations with the cultar platform. In: SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications, SA '15. Association for Computing Machinery, New York (2015). https://doi.org/10.1145/2818427.2818457
50. Papagiannakis, G., Ponder, M., Molet, T., Kshirsagar, S., Cordier, F., Magnenat-Thalmann, M., Thalmann, D.: Lifeplus: revival of life in ancient pompeii, virtual systems and multimedia. Technical report (2002)
51. Pietroni, E.: Experience design, virtual reality and media hybridization for the digital communication inside museums. Appl. Syst. Innov. **2**(4) (2019). https://www.mdpi.com/2571-5577/2/4/35
52. Pietroni, E., Pagano, A., Amadei, M., Galiffa, F.: Livia's villa reloaded virtual museum: user experience evaluation. In: Proceedings of the 9th Annual International Conference of Education, Research and Innovation (2016)
53. Pietroni, E., Pagano, A., Poli, C.: Tiber valley virtual museum: user experience evaluation in the national etruscan museum of villa giulia (2016)
54. Pietroni, E., D'Annibale, E., Pagano, A.: The use of holographic showcases inside the museum's context. Towards an advanced museology creating a dramaturgy around the exhibited objects (2018)
55. Pietroni, E., Pagano, A., Fanini, B.: Ux designer and software developer at the mirror: assessing sensory immersion and emotional involvement in virtual museums. Studies Digital Heritage **2**(1), 13–41 (2018)
56. Pietroni, E., Ferdani, D., Forlani, M., Pagano, A., Rufa, C.: Bringing the illusion of reality inside museums—a methodological proposal for an advanced museology using holographic showcases. In: Informatics, vol. 6, p. 2. Multidisciplinary Digital Publishing Institute, Basel (2019)
57. Pradana, G.A., Zhang, E.Y., Cheok, A.D., Morisawa, Y.: Delivering haptic sensations in mobile marketing. In: Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology, ACE '15. Association for Computing Machinery, New York (2015). https://doi.org/10.1145/2832932.2856223
58. Rabbi, I., Ullah, S.: A survey on augmented reality challenges and tracking. Acta graphica: znanstveni časopis za tiskarstvo i grafičke komunikacije **24**(1–2), 29–46 (2013)
59. Ridel, B., Reuter, P., Laviole, J., Mellado, N., Couture, N., Granier, X.: The revealing flashlight: interactive spatial augmented reality for detail exploration of cultural heritage artifacts. J. Comput. Cult. Herit. **7**(2), 1–18 (2014)
60. Rodrigues, J., Ramos, C., Pereira, J., Sardo, J., Cardoso, P.: Mobile five senses augmented reality system: technology acceptance study. IEEE Access **PP**, 1–1 (2019). https://doi.org/10.1109/ACCESS.2019.2953003
61. Rodrigues, J., Cardoso, P., Monteiro, J., Ramos, C.: Smart Systems Design, Applications, and Challenges. Engineering Science Reference (2020). https://books.google.it/books?id=iTstzQEACAAJ
62. Speicher, M., Hall, B.D., Nebeling, M.: What is mixed reality? In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2019)
63. Spence, C., Obrist, M., Velasco, C., Ranasinghe, N.: Digitizing the chemical senses. Int. J. Hum.-Comput. Stud. **107**(C), 62–74 (2017). https://doi.org/10.1016/j.ijhcs.2017.06.003
64. Stone, R.T., Bisantz, A.M., Llinas, J., Paquet, V.: Augmented multisensory interface design (amid): a human-centric approach to unisensory and multisensory augmented reality design. J. Cognit. Eng. Decis. Making **3**(4), 362–388 (2009)
65. Sutherland, I.E.: The ultimate display. In: Multimedia: From Wagner to Virtual Reality, vol. 1 (1965)
66. Tscheu, F., Buhalis, D.: Augmented reality at cultural heritage sites. In: Information and Communication Technologies in Tourism 2016, pp. 607–619. Springer, Berlin (2016)
67. Van Krevelen, D., Poelman, R.: A survey of augmented reality technologies, applications and limitations. Int. J. Virtual Real. **9**(2), 1–20 (2010)
68. Varela, F.J., Thompson, E., Rosch, E.: The embodied mind: Cognitive science and human experience. MIT Press, Cambridge (2016)

69. Vienne, C., Sorin, L., Blondé, L., Huynh-Thu, Q., Mamassian, P.: Effect of the accommodation-vergence conflict on vergence eye movements. Vis. Res. **100**, 124–133 (2014)
70. Vlahakis, V., Karigiannis, J., Tsotros, M., Gounaris, M., Almeida, L., Stricker, D., Gleue, T., Christou, I.T., Carlucci, R., Ioannidis, N., et al.: Archeoguide: first results of an augmented reality, mobile computing system in cultural heritage sites. Virtual Reality Archeology Cultural Heritage **9**(10.1145), 584993–585015 (2001)
71. Wloka, M.M., Anderson, B.G.: Resolving occlusion in augmented reality. In: Proceedings of the 1995 symposium on Interactive 3D graphics, pp. 5–12 (1995)
72. Yanagida, Y.: A Survey of Olfactory Displays: Making and Delivering Scents, pp. 1–4. IEEE, Piscataway (2012). https://doi.org/10.1109/ICSENS.2012.6411380
73. Yanagida, Y., Nakano, T., Watanabe, K.: Towards precise spatio-temporal control of scents and air for olfactory augmented reality. In: 2019 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), pp. 1–4 (2019). https://doi.org/10.1109/ISOEN.2019.8823180
74. Yue, Y.T., Yang, Y.L., Ren, G., Wang, W.: Scenectrl: mixed reality enhancement via efficient scene editing. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 427–436 (2017)
75. Zoellner, M., Keil, J., Wuest, H., Pletinckx, D.: An augmented reality presentation system for remote cultural heritage sites. In: Proceedings of the 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, pp. 112–116. Citeseer (2009)

Emanuel Demetrescu PhD in Archaeology, is a researcher at the Institute of Heritage Sciences – National Research Council, Italy. A specialist in digital archaeology, building archaeology, and 3D source-based modelling, his main research regards the creation of theoretical and methodological links between scientific record in Cultural Heritage and creative industries (3D digital libraries, virtual museums, open world games, and VR experiences).

Bruno Fanini Computer scientist, PhD – focuses his research on interactive 3D graphics, immersive VR, interaction design, and 3D user interfaces at the Institute for Heritage Sciences of the National Research Council (CNR ISPC). He developed software tools, serious games, Web3D/WebXR applications, and services within national and international projects.

Daniele Ferdani PhD in Archaeology, is a researcher at the Institute of Heritage Sciences – National Research Council, Italy. His research activities deal with digital archaeology and are particularly oriented to the definition and experimentation of methodologies and integrated digital technologies for the acquisition, interpretation, 3D modeling, and visualization of cultural heritage and dissemination through virtual reality systems.

Alfonsina Pagano is currently a research fellow at the National Research Council, Institute of Heritage Sciences, in Rome, working as user experience researcher. She also works for several companies as multimedia consultant, communication assistant, and user interface evaluator. Her research interests include transmedia projects with the goal of investigating the user experience design of digital applications, digital storytelling strategies, and cognitive parameters applied to the virtual museums field.

Augusto Palombini PhD, is a researcher at the Institute of Heritage Sciences – National Research Council, Italy. His research activity on digital archaeology is focused on landscape archaeology, virtual museums, digital storytelling, computer science, and multimedia solutions applied to Cultural Heritage and topography, particularly dealing with open-source software. He has been involved in six EU research projects, between VIIFP and H20202.

Eva Pietroni is an expert in CH preservation, an art historian and musician, and a researcher at ISPC-CNR. Her activity focuses on "virtual" museums, from digitization to 3D reconstruction and communication. The creation of new narrative forms and the hybridization of media (VR, theatre and cinema, mixed reality, videogames, music, holography, documentaries) and the experimentation of new interaction interfaces are privileged themes in her research.

17 Augmented Reality in Holocaust Museums and Memorials

Jennifer Challenor and Minhua Ma

Contents

17.1 Introduction 413

17.2 Traditional Education of the Holocaust 414
17.2.1 Ethical Considerations for Education of the Holocaust 415

17.3 Multimedia Content for Holocaust Education 417

17.4 Augmented Reality for Holocaust Education 420

17.5 Ethics of Developing Augmented Reality Apps for Holocaust Museums and Memorials 422

17.6 Visualization Methods for Digital Reconstruction 424

17.7 Potential Applications of Augmented Reality for Holocaust Education 427

17.8 Conclusion 430

References 430

Abstract

Augmented reality (AR) is a new medium with the potential to revolutionize education in both schools and museums by offering methods of immersion and engagement that would not be attainable without technology. Utilizing augmented reality, museums have the capability to combine the atmosphere of their buildings and exhibits with interactive applications to create an immersive environment and change the way that audiences experience them and therefore providing the ability to perform additional historical perspective taking. Holocaust museums and memorials are candidates for augmented reality exhibits; however, using this technology for them is not without concerns due to the sensitive nature of the subject. Ethically, should audiences be immersed in a setting like the Holocaust? How is augmented reality currently being used within Holocaust museums and memorials? What measures should be taken to ensure that augmented reality experiences are purely educational and neither disrespectful to the victims nor cause secondary trauma? These are the questions that this chapter will seek to answer in order to further develop the field of augmented reality for Holocaust education. To achieve this, previous AR apps in Holocaust museums and memorials have been reviewed, and a series of studies on the usage of AR for Holocaust education have been examined to identify the ethical considerations that must be made and the ramifications of utilizing AR technology to recreate tragic periods of history.

Keywords

Augmented reality · Mixed reality · Technology-enhanced learning · Ethics · Holocaust · Museum · Memorial

17.1 Introduction

Museums and heritage sites historically faced the near unachievable mandate to provide education of historical events to new generations in a manner that simultaneously informs accurately about the past while also being presented in ways that are appropriate for the context of the topic. With the rise of technology-enhanced learning and the addition of educational computer applications within these environments, this challenge has become exponentially more complex. Both virtual and augmented realities are examples of technologies being developed for museums and heritage sites, each offering a new medium of exploring and interacting with elements of history that are otherwise unavailable and also allowing users a personal experience. However, while it is theoretically possible to create an accurate simulation

J. Challenor (✉)
Games & Visual Effects, Staffordshire University, Stoke-On-Trent, UK
e-mail: JC62@Staff.staffs.ac.uk

M. Ma
Falmouth University, Penryn, Cornwall, UK
e-mail: M.Ma@falmouth.ac.uk

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_17

of events within a virtual environment, there are concerns about the ethics of presenting history in this fashion. This raises questions, such as what should be allowed to be depicted within the confines of digital representation? Should audiences be exposed to traumatic historical events? How does an appropriate representation differ from a disrespectful one? How is this technology currently being used for history education and what measures were taken to ensure they were appropriate? These are the questions this chapter seeks answers to, with the intention of being used as ethical guidelines to create an accurate but appropriate virtual and augmented reality application for Holocaust museums and memorials.

17.2 Traditional Education of the Holocaust

Before exploring the ethics of using technology for history education, it is first prudent to analyze how history is or was educated using non-computerized methods to establish the ethical considerations made prior to the adaptation of technology for educational purposes and how they may be applicable to emerging digital technologies such as virtual and augmented realities. Historical events and teaching about them vary culturally, with each country taking a different approach to the subject depending on cultural values and the involvement of the nation involved. This was discussed by Eckmann et al. [1] on teaching about the Holocaust, which talks about how certain countries prepare their schools history curriculum in order to address the topic. For example, Germany takes a very direct stance to the Holocaust and its role within it, Poland teaches it with a focus on the German invasion and how it impacted upon its citizens, and Norway does not include it within its curriculum at all. Educators in Israel undertake special training on how to teach the Holocaust due to its significance on their state religion, but this training is rarely undertaken by educators in other countries. The significance of this is that a single set of ethical guidelines cannot be established that will be appropriate for every nation as cultural differences and historical importance of an event will determine how it is taught.

Cultural differences in educational standards are also impacted by religious schools that teach both faith and culture of the religion alongside the state curriculum, such as Jewish Day schools. These schools have the freedom to deviate from the curriculum designated for public schools when it comes to matters of importance to the faith of the institution including which areas of history or sociology to focus upon. An example of this is within the schooling in the United States, where Ellison [2] examined the differences in how the Holocaust is taught in public schools when compared to Jewish Day schools. This study was performed with surveys distributed to Jewish Day schools but did not do the same to public schools, instead relying upon the data from a study from 2004, which does question the validity of the findings as they rely upon potentially obsolete data. Regardless, the study found that teachers of the Holocaust held similar rates of qualifications regardless of institutional type and that the topic was covered in all regions of the United States. Both types of school used multimedia content as their preferred educational resource; however, not all components of teaching were similar, with differences including elements such as the time allocated to teaching the Holocaust. While not a striking disparity, Jewish Day schools on average allocated over a month of time to teaching this topic, while public schools averaged around three weeks, which demonstrates a difference in priority of the subject based upon religious background. The study also notes that the motivation for teaching the Holocaust varied based upon school type; Jewish Day schools educate about the Holocaust as a way of teaching the history of antisemitism, whereas public schools teach the subject to educate about the dangers of prejudices and the horrors they can lead to when nurtured rather than challenged. Ellison concludes on a critical note regarding these presentations of what the Holocaust truly was at its core, as he asserts that the event was the culmination of anti-Semitism and should be presented as such rather than the public-school approach of utilizing it as a metaphor for all forms of prejudice and stereotyping. Public schools may have difficulty allocating sufficient time to cover both the Holocaust and other forms of prejudices, so it is not unreasonable that they have previously attempted to combine the two topics; however, for a Holocaust museum, there is no doubt that the focus should be on the discrimination that targeted the victims.

Educators as a collective, both in schools and other learning institutions, must approach the Holocaust and other tragedies delicately to balance the factual education with the moral lessons involved. Suissa [3] discussed this in her paper on Holocaust education, which she describes as *complex and difficult*. Suissa analyzed several authors' stances on Holocaust education and the lessons they chose to emphasize during their teaching, including Tutu [4] who focuses upon the immorality of genocide and preventing it in the future, Langer [5] who takes a more philosophical approach into the meta-ethical aspects of the concept of evil and Lang [6] who places significance on the steps that lead to the genocide. As a philosopher, Suissa's focus was on morality and epistemic justice, but her analysis of the ethical lessons demonstrates that educating about the Holocaust or other tragedies is not as simple as discussing the facts and figures with learners, instead educators have the additional responsibility of explaining complicated concepts regarding morality and society and why events like these must not be permitted to occur again.

English secondary schools have been the focus of several studies on pedagogies for Holocaust education, which is a topic covered across multiple classes including history, religious education, English, citizenship, and PSHE (personal, social, and health education). Pettigrew et al. [7] conducted an extensive survey with over 2,000 teachers from secondary schools all throughout England with the intention of discovering how the Holocaust was being taught to students. The data was gathered via an extensive 54-page length online survey and followed up with interviews to the limited number of teachers who were available and willing to participate. Some of the data collected pertained to the subjects that addressed the Holocaust or the number of hours dedicated to teaching it; however, the study also discovered topics that were and were not covered within the classroom. The five most commonly addressed topics were testimonies of survivors who were persecuted, the Auschwitz-Birkenau concentration camp, propaganda/stereotyping, Kristallnacht, and the choices of bystanders. Four topics that approximately 60% of respondents claimed were not likely to be taught included the Nuremberg laws, the actions of rescuers, the rise of Hitler and the Nazi Party, and combating modern-day racist ideology. One area that was covered by exceptionally few teachers was Jewish life and culture prior to the war, which experts believe is essential for making students understand the impact of the Holocaust and how ruinous it had been for the Jewish community. The study also inquired as to teaching aims and objectives toward teaching about the Holocaust and what educators hoped to achieve with their lessons, to which responses leaned toward a preventative lesson; by exposing their students to the harsh reality of what human brutality can accomplish, many teachers expressed that they wanted students to understand how they could be individually important at preventing further tragedies.

Teaching the Holocaust at English schools has been a point of contention too, with views being expressed about how much of the curriculum is focused around the topic. One such criticism was voiced by Sir Schama [8] in an interview about the British educational system, specifically regarding how Jewish history is taught within UK schools. Sir Schama's concerns regard how Jewish history ranges over several thousand years with countless events and periods that could be taught, but schools shun them all in favor of making the Holocaust the focus of the curriculum for many subjects, which results in what Sir Schama refers to as *Holocaust fatigue*. While a valid point, this may not be applicable for a Holocaust museum or heritage site environment where the focus of the exhibits is based specifically around this singular period of Jewish history; however, Sir Schama also commented that pedagogical methodologies for teaching the Holocaust are in need of modernization to utilize digital tools to mitigate or bypass the effects of Holocaust fatigue. As a Professor of history with teaching experience and a specialization in Jewish history, Sir Schama's opinions regarding the development of pedagogy must be considered for the future of education along with how technologies such as augmented reality can be utilized to effectively evolve teaching methodologies. This is especially important for expanding educational capacity beyond just a single topic such as the Holocaust, as providing students with the digital tools to explore more of Jewish history would assist with learning more about the cultural heritage of the Jewish people and combat anti-Semitic notions, which Sir Schama claims are growing across both sides of the political spectrum.

17.2.1 Ethical Considerations for Education of the Holocaust

Given the nature of the Holocaust and its unprecedented levels of hostility toward the Jewish community and the other targeted groups such as the Romani, teaching the subject requires educators to ensure that it is handled with the correct sensitivity and is not construed in a manner that is offensive, inappropriate, or needlessly distressing. This includes both teaching the subject in a manner that does not cause unnecessary upset to students and addresses the victims of the Holocaust in a manner that is respectful. For this reason, it is important to consider ethical issues when teaching the Holocaust.

Education of the Holocaust within English and Welsh schools did not begin until 1991, when an analysis of the curriculum revealed that neither the World War II nor the Holocaust were included at any stage of learning. According to a study by Geoffrey Short [9], a large challenge faced when introducing the concept of the Holocaust to the schoolchildren of the time was getting them to comprehend the cultural identity of the Jewish people. This is because of two potential factors, which could alienate perception of Jewish culture, including inherited anti-Semitic prejudices gained from parents and the media or bad teaching that does not encourage students to sympathize with the Jewish people. This can include failure to communicate that Judaism is a monotheistic religion akin to Christianity or Islam rather than an ethnic group, the distinction between which may determine if a student views a Jewish person as a follower of a faith or an alien entity. The study performed in this research involved interviewing schoolchildren in groups of two and asking them questions about Judaism, which varied between methods of worship, tenets of the faith, the lifestyle requirements such as kosher tradition, even questions about the student's personal identity, and how they felt their life would differ if they had been born into or converted to Judaism. Responses from these interviews showed misunderstanding about the religion, inability to discern between Judaism and Islam, acknowledgment of societal prejudice against the Judaism, demonstration of anti-Semitic stereotypes, and some level of tolerance and sympathy. The author then discusses adjustments that are required to the teaching policy in order to combat misconceptions, including ensuring that religious education teachers clearly differentiate between Judaism and Islam, clarifying the nature of Judaism as a faith rather than an ethnicity, reinforcing the origins of Christianity with Judaism and challenging anti-Semitic stereotypes. Finally, it is also of exceptional importance that students can discern the difference between Jewish identity and dedication to the Jewish faith as the Holocaust did not discriminate between the two. Within the context of ethical considerations, the significance for teaching this study is less on how discussing the Holocaust should be approached but rather on how to teach about Jewish culture and cultivate understanding of the systematic prejudice that targeted individuals regardless of how devout they may or may not have been.

Geoffrey Short has also written about the ethical and pedagogic issues of teaching about the Holocaust in another paper [10] that sought to further analyze how it was being taught within England. This study was based upon the same surveys and more recent ones that were directed at secondary school teachers within England regarding their continued curriculum development and how they were adapting their teaching of the Holocaust. While this paper gives mention to similar points as the previous one, it also raises three ethical concerns that must be addressed: freedom of speech, Holocaust denial, and infliction of pain.

- Freedom of speech does not refer to a constitutional or human right, but rather how the teacher allows for open discussion within their classroom as schoolchildren can form or repeat anti-Semitic sentiments, and how the teacher chooses to handle such comments is a delicate matter. While some would advocate for the classroom to be a completely open forum, the paper argues that should a teacher choose to allow a platform for discussion, they must be prepared to challenge anti-Semitic views or other forms of prejudice and provide an explanation for why these views are either incorrect or unacceptable.
- Holocaust denial is a false version of events proposed by revisionist historians with anti-Semitic intent and could be disruptive to the learning process if addressed within the classroom. The paper acknowledges that by informing a class about the views expressed by Holocaust deniers, an educator risks giving them credibility or derailing their lesson plan by having to confront them. Should this topic arise during a class, the paper recommends demonstrating the evidence that shows that the Holocaust really did happen and to open a dialog with the students about the motivations the revisionists would have for denying it.
- Infliction of pain regards the situations when a person intentionally causes harm to another. Schools of philosophy have discussed this topic at great lengths and from multiple aspects, but these considerations are far beyond schoolchildren who are too young, naïve, or inexperienced to comprehend such concepts, and for whom the topic of the Holocaust may be distressing. The paper directly condemns overly distressing content, claiming that while the topic will always be a painful one for new learners, it is wrong for an educator to make it more painful than is required to effectively teach it.

While promoting discussion is the responsibility of an educator, providing a neutral platform for views of prejudice or hate is ill-advised as it risks allowing the views of other students to potentially be influenced. However, silencing dissenting viewpoints will only seek to remove them from the conversation, instead an educator should seek to challenge these positions whenever possible to make students understand why anti-Semitism is wrong. Educators should also be selective with the content of their lessons to ensure that the material is not overly traumatic or distressing for students. This may be challenging at times due to the nature of the Holocaust, but callously causing emotional anguish in learners for the sake of educational benefit is both unethical and detrimental to the learning process.

Approaching this challenge may be a daunting prospect, but support is readily provided by specialist institutions who can offer advice and lesson content for educators. One such institution is the International Holocaust Remembrance Alliance, which offers resources and guidelines [11] for teaching the subject to schoolchildren in a manner that is accurate and appropriate to younger sensibilities. Details of this can be found on the institution's website, which contains educational materials and supporting advice for approaching individual topics, including how to effectively utilize witness testimonies, the type of language to use to dispel myths or misconceptions, and how to respond to the concerns of students. From an ethical perspective, guidelines are also provided for the topics that require a considerate approach such as the usage of graphic imagery, something which should not be used with the intent of being shocking as this is distressing for students and grossly degrading to the victims while not providing any appropriate educational value. Depiction of the Nazis is also important as the guidelines advise against any attempts to dehumanize them or present them as being monsters, instead claiming that the educator should take every opportunity to reinforce that the perpetrators of the Holocaust were ordinary people and how anyone can be capable of performing atrocities under certain circumstances. Students should also not be tasked with any assignment that requires them to align with the perspective of a perpetrator or victim because of the danger of falsely empathizing with their designated role; creative writing or roleplaying from these perspectives will only achieve a level of depth that the guidelines consider to be superficial while also being immensely disrespectful to the suffering that the victims really went through.

Other institutions have offered similar advice on how to teach about the Holocaust, but there is some variance in the guidelines provided depending upon the organization. For instance, the United States Holocaust Memorial and Museum [12] provides their own instructions with some new points raised. An example of these points would be teaching on how the Holocaust happened while condemning that it was allowed to; the event should not be presented as an inevitable occurrence but as a travesty that could have been avoided. Preventing tragedies like the Holocaust happening again is reliant upon enforcing how the actions or inactions of individuals, political parties, and nations were all pivotal in allowing the escalation that lead to genocide occurring. Similarly in regard to the actions of individuals or groups, history should not be romanticized in any way; although there are incredible stories of those who risked their lives to rescue or aid those persecuted under the Nazi regime, this only accounted for a small number of people, and placing too much emphasis upon them will skew perceptions of history. The guidelines give further mention to the victims and how they should be presented as people rather than statistics, because the act of simply presenting numbers does little to convey the scale of the atrocity. Each person who suffered under the Nazi regime had a family and friends, a school or a job, hobbies, interests, and many other distinctive characteristics; presenting stories or testimonies is greatly more beneficial for education purposes to make students understand just quite how malicious the Holocaust was.

17.3 Multimedia Content for Holocaust Education

Multimedia formats such as television, films, and computer games have very rapidly become an intrinsic part of both society and culture within the last century, each becoming more prevalent with the progression of technology and the availability of consumer grade electronics such as televisions, computers, game consoles, and smartphones. While multimedia content has mostly seen its use for recreational purposes, there is a great deal of potential for educational value too, as was claimed by Hammer and Kellner [13] in their study on a multimedia pedagogy. This study aimed to analyze the usage of multimedia technology intended to teach historical events such as the Holocaust. One such example of this is the Shoah Project, an educational program created by the USC Shoah Foundation that intended to capture high-quality recordings of Holocaust survivors in both video and audio format and utilize it within a computer application. This application would allow additional content to be presented alongside the testimonies which included maps, historical footage, music, sound effects, and any other relevant multimedia content. Hammer and Kellner claim that an immediate advantage of using this type of multimedia content for learning is that it immediately dispels myths or misconceptions that could form within the classroom, such as the infamous *led to the slaughter* metaphor: the false assumption that the victims of the Holocaust gave no resistance. By making use of multimedia content, the creators were instead able to present the interviewees at the end of each segment and allow them to introduce their families, show pictures and news clippings, read letters or journals, or present any other material available to them to convey their experiences, which consequently would assist with overcoming prior assumptions made by learners.

Multimedia methods of educating about the Holocaust are not without criticisms, and some have even argued that pedagogies including them have suffered for it. Films, television shows, books, and games made with the Holocaust as the central narrative theme are all subject to elements of revisionism or alteration at the whims of the author and would be detrimental to learning and a challenge to educators who would have to revise the topic to correct historical inaccuracies. A film that perpetuated this problem is *The Boy in the Striped Pyjamas* [14], a film based upon a book of the same name [15] about two young boys on opposing sides of the Holocaust. An essay by Michael Gray [16] analyzes the film adaptation and its usage within education in the aftermath of a campaign by the distributor of the film to advertise it to schools as a valuable education resource. The distributor also launched a survey to test what schoolchildren knew of the Holocaust as part of this campaign and published the findings. The schoolchildren who participated in the survey displayed some concerning misconceptions regarding the Holocaust, such as falsely believing that Jewish citizens went voluntarily to concentration camps, believing that the prison uniforms really were pyjamas or that Auschwitz was a small and isolated camp far away from society. Historically, the film is flawed and full of errors, such as the Holocaust being treated as a guarded secret even to the wife of a camp's commandant or the main character mysteriously having avoided the extreme societal prejudice against the Jewish community that would have been taught to him at school. An educator could potentially disregard the historical inaccuracies of the film in favor of the moral lessons of the film, but Gray also chastises these due to the flaws that emerge when the story is subjected to scrutiny, such as how the tragedy of the story tries to influence the audience into sympathizing with a Nazi commandant due to the loss of his son over the deaths of the Jewish victims. The author concludes that the film's educational benefit stems only to Holocaust awareness, and prior points made suggest that using this film to teach the Holocaust would mislead students with a confusing moral message. The significance of this study is a clear demonstration that misinformation can easily be spread through multimedia content when the creative freedom of the author interferes with historical accuracy, and so for a multimedia text to be of educational benefit, it must endeavor to be as historically accurate as possible.

Maintaining complete historical accuracy directly contrasts with the vision of the auteur as many directors attempt to influence the visual or narrative components of their films to conform to their own conventions or style. For instance, the director Tim Burton is renowned for making films with overt gothic aesthetics, Michael Bay infamously includes a large number of explosions within his movies, and M. Night Shyamalan has a reputation for making films with a narrative twist, yet where does the line get drawn between the will of the creator and acceptable deviation from history? The famous example of this would be *Schindler's List* [17], an Oscar winning film about Oskar Schindler, a businessman and member of the Nazi Party who saved the lives of 1,200 Jewish citizens during the Holocaust. Classen and Wächter [18] discussed this film in their paper on its authenticity and believability for audiences, in which they discuss the elements of production that assisted with its historical accuracy. A majority of this paper is an analysis on the elements of the film that evoked a sense of realism from the audience. However, it does also detail the measures taken by the director and production crew to ensure that the film was made to be as accurate to history as physically possible. The authors discuss the unyielding nature of the film's director Steven Spielberg, who refused to compromise on any details such as filming in original locations including the Auschwitz-Birkenau camp and the Kraków Jewish Ghetto even in spite of resistance from the relevant authorities to allow permission, all to guarantee a feeling of authenticity beyond that of a superficial glance at a constructed set. The narrative of the film was an element that faced criticism due to the plot structure being fairly conventional for a Hollywood movie with the protagonist getting a happy ending; however, this was also true for the real Oskar Schindler who did successfully escape prosecution for war crime charges on the grounds that he had profiteered from slave labor. Another factor that the authors claim contributed to the feeling of authenticity was the aesthetic decision by Spielberg to shoot the film in black-and-white, a decision that was unconventional for a Hollywood film at the time but served to mimic the effect of archival footage captured during either the end of the war or the liberation of the camps. Finally, the film closes with the real survivors from Schindler's factory attending his grave along with the actors who portrayed them as a final message to the audience that the story was real and that the Holocaust did truly happen outside the scope of the film's narrative, an element that the authors praise due to its powerful message. Looking back to the original question of how much autonomy a director has when making an educational film, Spielberg has established a very sensible benchmark for comparing future works via the unwritten rules that can be extrapolated from this piece. These rules can be described as follows: Aspire for complete accuracy whenever possible, do not compromise on details even when they are detrimental to production, make aesthetic variations only when they serve to enhance the piece, and emphasize the reality of the story at every stage of the narrative.

Concentration camps tend to be the focus of film depictions of the Holocaust, which stands to reason as the forced labor or systematic murder makes for emotionally evocative material but does not tell the full tale nor educate audiences on the social, economic, or geopolitical circumstances that lead to the genocide occurrence. Roman Polanski challenged this trend with his film *The Pianist* [19], a drama film based on the autobiography of Władysław Szpilman, a Jewish man from Poland who narrowly survived the Holocaust. As a director, Polanski was exceptionally qualified to make a film about this period as he himself had survived the Holocaust as a child and had first-hand experience with life in Poland under the Nazi regime. The film follows Szpilman and the struggles endured by his family from the initial invasion by the German military, beginning with the family being targeted by the German soldiers with less severe forms of harassment such as restrictions upon finances or being banned from certain establishments. Over the course of the first act, the restrictions become increasingly harsh and expand to being forced to wear star of David armbands, getting forcefully relocated into a ghetto and ultimately being deported to the Treblinka extermination camp, although Szpilman escaped this fate after being rescued prior to boarding the train and spends the remainder of the film running, hiding, and evading capture at the hands of the German soldiers. This representation shows a different side to the Holocaust: life for the Jewish citizens of Poland outside of the concentration camps and the struggle to stay alive under the oppressive regime. Polanski's own experiences coupled with a meticulous effort to follow Szpilman's autobiography as closely as possible created an excellent educational resource that accurately conveys the living conditions and prejudices that severely impacted upon the lives of the Polish Jewry. The film also depicts an event not commonly shown in the media: The Warsaw Ghetto Uprising, a movement by the Jewish resistance to fight back against the Nazi soldiers who were attempting to deport them to death camps. This depiction of the uprising is important as an educational tool to dispel the false myth of *led to the slaughter*, instead showing that the Jewish people made a stoic stand against their oppressors, even if it was an ultimately doomed effort. Once more, accuracy is shown as being the key to creating an effective educational resource, does not attempt to dramatize or censor details but depicts events as closely as possible to how they really occurred.

Films are not the only form of multimedia through which the Holocaust has been portrayed, but television has also been used as a medium to address this period of history including another work of Spielberg's, the miniseries *Band of Brothers* [20]. As a drama series about World War II, the show revolved around a company of US soldiers as they progressed through the war but also featured an episode in which the main characters are present during the liberation of the Kaufering IV concentration camp. During this episode the soldiers discover and enter the camp shortly after it had been abandoned and learn of the events that had transpired from a heavily emaciated prisoner before deploying relief efforts to assist the surviving victims, each relieved at the sight of the approaching allied forces and rushing to show their gratitude, while the soldiers look around in horror. Controversially, the episode ended with a commanding officer ordering the victims to be kept locked inside the camp, a potential deviation from actual history which was called into question by Mendes [20] in her article on Holocaust liberation efforts. Mendes argues that this was a dramatization of events when compared to a documentation from real history; there is evidence to suggest reasons why allied forces may have wanted to keep the inmates contained within the camps but no proof to suggest that they were ever forcibly locked within. *Band of Brothers* suggests that the prisoners needed to be contained for medical personnel to monitor their recovery, a concern that was valid due to epidemics that were prevalent within the concentration camps including typhoid fever and typhus, the disease which infamously killed Anne Frank. Victims were also severely malnourished to varying degrees, and the act of eating food provided by the soldiers was enough to kill them as they were unable to digest it, resulting in medical staff requiring nutrient-rich liquid foods to save them. Despite this, however, an investigation committee created by President Harry Truman wrote a report on the treatment of concentration camp prisoners post liberation, which found that victims were not being kept in their original camps but in newer resettlement camps established by the US military for humanitarian purposes. This conflicts with the representation depicted by *Band of Brothers*, which seems to have been aiming for dramatic impact rather than an accurate recreation of events and also contrasts with the previous work of Spielberg who had strived for complete accuracy. Nonetheless, as a representation, there is still educational benefit for viewers for the portrayal of the liberation and the severe emaciation of the victims at the end of the Holocaust, a tragic occurrence that relies upon visual media to accurately convey how starved and mistreated the victims had been.

Computer games are another field of multimedia which have depicted the Holocaust, albeit to varying levels of responses from audiences. One such game is *Call of Duty WWII* [sic] [21], a World War II-themed action game, in which the player takes the role of an American soldier in the European theatre. The story of the game begins with the player participating in the battle of Normandy and ultimately leads to the player walking through the Berga concentration camp shortly after it had been abandoned. During this segment of the game, there are no gameplay mechanics or interactions, the player can only explore the camp, while the non-playable characters of the player's squad remark upon the horrific conditions that the inmates of the camp had been subjected to. However, there is no mention given to the Holocaust nor the systematic targeting of the Jewish people by the Nazi party, something which was criticized by Rosenberg [22] in his article upon the depiction of the Holocaust within this game. Rosenberg argues that the depiction of the Holocaust within the game is a very weak portrayal of the event because of how little is mentioned or shown within the level as the graphic content consists of the miniscule living spaces and two corpses of inmates that were executed by firing squad, something which the player would likely not consider to be shocking after the much more extreme depictions of graphic violence previously encountered in the game. Rosenberg compares this to the education they received about the Holocaust at a Hebrew school and how they were exposed to documentaries, camera footage, photographs and other graphic content, all of which served to demonstrate the genuine and brutal nature of the event instead of attempting to censor it to cater to the sensibilities of audiences. Despite this, Rosenberg does also concede that while they considered the games depiction of the Holocaust to be shallow, it would be challenging to create a concentration camp game level in an appropriate fashion and that although the representation may have been poor, it was still an important stride that the event be acknowledged as other computer game adaptations of World War II fail to give mention to the Holocaust at all.

Representation of the Holocaust is a difficult subject within computer games as developers do not acknowledge the topic nor attempt to create a game centered around it. This changed in 2018 when Alien Games, a Ukrainian game developer, announced an upcoming title called *Cost of Freedom* [23], a computer game in which the player assumes the role of either a Jewish prisoner or a Nazi guard. Immediately after announcement, the game was lambasted by the media due to the overtly shocking gameplay mechanics depicted in the trailer, which included the player being given the choice to order the execution of a group of inmates. Harsh criticism of the concept of the game was unanimous among journalists who wrote about the game including an article by Pepper [24] that condemned the game for disregarding human empathy and portraying victims of the Holocaust as though they were playthings. The author was also critical of the very concept of a Holocaust computer game, and how it has no right to exist as a virtual playground as the very concept is immensely disrespectful to the millions who died during it. The game was very quickly banned in

Poland and was subsequently cancelled by the developers due to the overwhelmingly negative public response, but this does draw into question how the Holocaust can be represented within future computer games. Rosenberg [22] did not approve of its representation within *Call of Duty WWII* because he felt that it was too heavily censored and did not accurately convey the brutal oppression of the victims, and yet *Cost of Freedom* took no measures to censor any element of the Holocaust but was criticized for its portrayal; so what type of representation is appropriate? The answer seems to lurk within the involvement of the player and their role within the game, as *Cost of Freedom* allowed the player to be an active participant within genocide as where *Call of Duty WWII* only allows the player to be a spectator of the aftermath. A game where the player is able to witness an unaltered depiction of the Holocaust that is constructed to be accurate but has no gameplay features or player involvement would seem to be the acceptable compromise; however, this does not align with typical game development as the purpose of such a level may be construed as being forced upon the player as opposed to being an organic component within the narrative structure. Any game developer who attempts to include the Holocaust within their product certainly has a very large challenge ahead of them in order to meet these requirements.

Gameplay within a concentration camp is another component of representation, which has been avoided by computer game developers too, most likely due to the aforementioned challenges and also because of potential backlash by the media or public opinion. Despite this, a concentration camp level was included inside the action shooter game *Wolfenstein: The New Order* [25]. The Wolfenstein franchise is a series of shooter games set in an alternate version of World War II with overt science fiction themes in which the player engages in gratuitous violence against the Nazi regime. In the concentration camp level, the player infiltrates the fictional Camp Belica to rescue a Jewish scientist but does so under the guise of a prisoner. Upon entry into the camp, the player witnesses as newly admitted prisoners are divided into two lines: prisoners suitable for work and those who will be immediately killed. Following this, the player has their arm tattooed with an identification tattoo and is sent to operate a concrete manufacturing machine. During the brief period that the player acts as a prisoner, the majority of inmates they interacted with are from various African countries, and no mention is directly given to either the Holocaust, the persecution of the Jewish people by the Nazis or the final solution, although the player's character introspectively acknowledges the Auschwitz and Buchenwald camps as existing. Exploring this representation of the Holocaust does not last long however, as the player quickly finds a weapon and leads a violent uprising against the Nazi soldiers, differentiating from actual history by merit that the player's character is still in peak physical condition and has not suffered the effects of emaciation from being starved and abused as the real victims did. *Wolfenstein: The New Order* was able to depict the Holocaust in such a direct manner due to the degrees of separation from reality that its narrative maintained; this is an alternate version of history in which the Nazis had science fiction weapons and the player is not engaging with what is intended to be a realistic depiction of history. Whatever actions occurring in this fictional version of reality may do so without any connotations regarding the real world, and because of that stipulation, the developers held a greater deal of freedom to explore the topic of the Holocaust with lessened risk of causing offense to the real victims. While this was a viable method for a computer game, it does not translate to representations outside the boundaries of entertainment media as attempting to portray the real-world Holocaust in a similar manner would be grossly disrespectful and exceptionally inaccurate with little or no educational value. The elements of the Holocaust depicted within this level were not poised as being educational nor were they explained, instead they were left at the mercy of the players' understanding, and a player with no prior knowledge of the Holocaust would only be able to assume that they were part of the fictional narrative, thus rendering this game inappropriate for educational purposes.

17.4 Augmented Reality for Holocaust Education

Within museum environments, augmented reality applications have become a popular addition to many exhibits due to their ability to enhance existing exhibits and provide both additional information and visualize concepts that may otherwise be unavailable. An example of this would be at the National Holocaust Centre and Museum in the United Kingdom, within which a study was performed by Ma et al. [26] to create an augmented reality exhibit of a virtual survivor. This research sought to digitally preserve the testimony of a Holocaust survivor as survivors are becoming increasingly scarce with the passage of time, and within a manner of decades, there will be no living people left with a first-hand account of this event. As a technical investigation, the intention of the research was to create and evaluate an interactive narrative system, utilizing extensive stereoscopic video footage of a survivor as he recounted his stories and experiences that were then combined with a 3D character model. Visitors to the museum could interact with this model by asking it questions using a microphone, at which point the system would then select a predetermined response and play the appropriate video segment to answer the question. While this study does not publish results, it still demonstrates that this technology is being utilized for Holocaust museums.

Narrative-driven experiences are a seemingly popular implementation of augmented reality technology, having seen its use in multiple studies. Another study to explore this area was a research performed by Stapleton and Davies [27] with the Maitland Holocaust Museum, utilizing augmented reality to create a storytelling experience that followed the life of a child. The story begins in the 1930s in a middle-class family home and progresses to both a ghetto apartment and ultimately a concentration camp, using a child's doll as the only consistent element between all locations. At the beginning of the story, the doll is new; at the second location, it has become disheveled; and at the third, the doll is replaced by one fashioned from objects from within a concentration camp. The authors discussed how the doll artifact was enhanced with the sound of a child playing rather than any visual representations, which served to manipulate the audience's imagination in a fashion they claimed was more powerful than any archival footage. This study was also a test into technical feasibility rather than testing a hypothesis but does raise a point regarding Holocaust education: audience emotions. Due to the ability of virtual representations to create a visceral experience, it is possible to cause emotional distress to audiences, with some audiences from this study being upset to the point of crying. Paradoxically it may be considered both optimal and suboptimal to reach this reaction; eliciting an emotional response of this nature is evidence that the audience successfully comprehends the horrors of the Holocaust but could also mean that the experience has been too traumatic to have caused this level of distress.

Storytelling of the Holocaust can be a compelling experience for viewers with a basic understanding of the period and knowledge of the perpetual dangers that the characters are in, something which can be even more tragic when the audience has the advantage of knowing how history actually played out and therefore the peril that lays ahead. Such an experience is being created by Jin [28] as part of his PhD research into narrative experiences for augmented reality, in which he is creating a story of a fictional family enduring prejudice at an early stage of the Holocaust circa 1939. This story is depicted in the National Holocaust Centre and Museum within the United Kingdom and is built for the Microsoft HoloLens. Within this story, the fictional family is facing hardships due to the growing anti-Semitic sentiments reverberating throughout Germany and must make decisions regarding their son's education, which the audience get to make via the interactivity along with the freedom to engage with virtual props that exist within augmented space. At this time, Jin has not published any work on his project; however, his technical demonstration video serves to demonstrate a proof-of-concept that narrative experiences using augmented reality are becoming a focal point for research projects, speculatively suggesting that this technology may soon become an intrinsic component of museum exhibits in the near future.

Creation of narrative scenes or any other variation of visualization project for the Holocaust is dependent upon accurate reconstruction of the historical location to provide correct context for the audience as well as for the purpose of immersion. Building an accurate reconstruction of any location is often a daunting challenge in itself but is exacerbated by the additional difficulties of the fragmented data available for the Holocaust. While some of the concentration camps were meticulously documented or preserved, others were converted or destroyed in the aftermath of the war with minimal records kept regarding their construction or designs, which has resulted in difficulties or compromises for those who have sought to reconstruct these camps for heritage visualization. An example of this would be a study by Kerti et al. [29] on the Lager Sylt concentration camp, a site that was almost entirely destroyed by its own commandant prior to the arrival of British forces at the end of the war. Due to incredibly little remaining of the original location and scarce availability of historical records, it is not possible to recreate an accurate photorealistic 3D visualization of the site. Simplified geometries were used instead to represent the locations and size of buildings, and each geometry is color coded instead of textures to indicate historical accuracy. Buildings that were documented and could be accurately represented both in location and dimensions were marked as white, while others were marked as gray to indicate that there was data (both marginal or extrapolated), suggesting that a building of that size existed at the location, but could not be guaranteed. While an elegant and informative solution to the problem, it also lacks a potential compromise for environments that are intended to be immersive, as audiences may be distracted by brightly colored buildings that lack textures or consist only of simple geometry. Within this context, is it acceptable to deviate from recorded history to create an aesthetically pleasing but potentially incorrect representation, or should accuracy be the highest priority even at the cost of the visuals?

Representing structures comes with challenges other than their appearance and audience perception, as the other factor to consider is the accuracy of their locations. Some of the larger or more infamous camps were extensively photographed in the aftermath of the war, which serves as a historical basis for accurate reconstruction as there is visual data to demonstrate where structures once stood. One such camp would be the Bergen-Belsen concentration camp, which was also destroyed but had been documented by the British and Canadian forces before they burned it down shortly after its liberation. This location was the subject of a study performed by Pacheco et al. [30] on location-based augmented reality using historical datasets, in which the authors digitally reconstructed the camp using spatial data. Participants in this study were provided with a handheld device that would render simple structures in the locations of buildings at the site to show where they once stood. Results collected from this study were focused toward spatial memory development and found that participants who were given free rein to explore had a better spatial understanding of the site than those who had been taken on a guided tour; however, this research was still significant as a technical feasibility test, as the technology to correctly and consistently display the locations of destroyed landmarks in augmented reality is a key factor toward maintaining historical accuracy.

17.5 Ethics of Developing Augmented Reality Apps for Holocaust Museums and Memorials

Simulating the Holocaust is a technically feasible concept, with a surplus of resources available to construct an accurate environment and create a narrative experience that would be as close to the real event as physically as possible both within the confines of virtual or augmented reality. The United States Holocaust Memorial Museum [12] advises against this concept however, instead condemning any simulation exercises to be pedagogically unsound for several reasons. Firstly, the experience may serve to engage audiences but runs the risk that the intended lessons of the simulation could get lost upon delivery and learners may get distracted with the less important elements of the application such as gameplay features. The other concern is that a learner could walk away from the simulation with a false sense of empathy and understanding of what it would have been like to have been a victim or even a perpetrator of the Holocaust depending upon the design of the simulation. All simulations have the potential for this risk, and a learner could experience a false sense of understanding from any period of history; however, the heavily prejudiced nature of the Holocaust makes this possibility far more insidious than that of other simulations. For instance, there are no ethical concerns related to a hypothetical simulation of ancient Rome, a city rather culturally progressive for its time that was home to a diverse population. Because the player would only be walking through the streets or performing interactions based upon recorded history, the player would only have the potential of believing they have experienced life as a Roman citizen and not as a modern Italian. With no content aimed at harming or oppressing singular group of people, the possibility of witnessing or engaging with content that could be construed as offensive to modern-day citizens of Rome or Italy is minimal, but the same would not be possible for the Holocaust.

Empathy is an aspect of history education that is difficult to approach because part of teaching any area of history requires students to understand the topic and even sympathize with it, but as aforementioned there is always a concern that learners may falsely correlate their acquired knowledge of the topic to an understanding of what it would have been like to have experienced history themselves. Bathrick et al. [31] addressed this concern on visualizing the Holocaust and referred to the process as *empathetic substitution*, something which the authors claim is more common in teenage students, and any attempts by the educator to encourage it among learners should be scrutinized. This process is a natural development and something that many will do when presented with historical occurrences or even fictional ones due to the innate curiosity that makes a person ask, *What would I do if I were in this situation?* Adolescents particularly will have a distorted view of this question with the firm belief that they could act heroically and change the outcome of events without calling into question how their version of events would depend entirely upon their knowledge of the original outcome, something which they would not know if they were experiencing the event first-hand. To avoid this, the authors address that there is a very fine line that exists between empathy and identification, one that serves as a challenge for any learning experience because the creator must tailor their content in a fashion that makes learners consider the victims without superimposing themselves into the scenario. An example of this is at the United States Holocaust Memorial Museum [30], where upon entry visitors are provided with an identification card for one of the victims that includes biographical information about that person, a simple but effective technique to iterate the tragedy of the event that works by influencing the visitor to empathize with that person but not to identify as them. This method creates a degree of separation between the visitor and the Holocaust, allowing them to consider how events impacted upon a specific victim and using that person as the lens through which their comprehension forms. The difficulty with this approach is that there is no generic solution that will be applicable to every visualization; however, the core concept of creating a focal point for the audience is a rather sensible approach to minimize the possibility of learners attempting to insert themselves into the situation. Within augmented reality, this could take the form of a character, such as in Jin's [24] narrative project where the audience does not directly participate within the story but rather influences the fictional character Leo; the learners occupy this space as a non-diegetic entity that empathizes through Leo instead of themselves. By doing so, the learner experiences history from a second-hand perspective as a spectator rather than a participant and therefore mitigating the risk of empathetic substitution.

Bathrick et al. [31] also discussed how representations of the Holocaust are affected by subsequent generations as teaching the topic to newer generations has differentiated from how it was first taught by the generation it impacted. This pedagogical shift was triggered due to the aging demographic of the Holocaust survivors and subsequently how the topic of trauma was going to be handled when teaching new audiences of the event. To this end, Bathrick et al. [31] invoked the concept of secondary witnessing: the understanding gained from the children of the surviving generation and their methodologies for teaching about the Holocaust based upon what their parents taught to them. From this concept, two theorists are identified: The first is Libeskind [32], an architect who designed the Jewish Museum in Berlin. The design of the museum is part of the experience with rooms designed to be claustrophobic, contain sharp angles, be dimly lit, consist of empty spaces, and provide an unsettling feeling to visitors with the intention of making the exploration of the museum a traumatic and thought-provoking experience. This stance on experiencing trauma is one detached from the Holocaust itself: The visitors are not made to feel as though their experience was similar to that of the Holocaust but rather a different sensation of trauma, one conjured from the oppressive atmosphere of the museum and its visual symbolism rather than one of empathetic substitution. Alternatively, the opposing theorist Hirsch [33] coined the concept of postmemory: how the experiences of a generation are taught by their subsequent generations. Hirsch argues that the understanding of the children of the surviving generation translates into different forms of artistic representation. Bathrick et al. [31] then argued that under postmemory audiences come dangerously close to empathetic substitution, but under Hirsch's theory, the narrative is controlled from the perspectives of secondary witnesses rather than the surviving generation, and consequently the impact of the trauma is not pressed onto newer audiences from a perspective through which they can attempt empathetic substitution.

Postmemory as a concept was first mentioned by Hirsch when he commented upon Art Spiegelman's graphic novel Maus [34], which is perhaps the best example of its implementation. The story of Maus is a visualized depiction of the author's father, Vladek Spiegelman, a Holocaust survivor from Poland who recounts his life experiences and his treatment under the Nazi regime. Narratively, Maus is told as a recounting of interviews the author had with his father, each recreated in graphic art form to tell the story as a continuing series of interviews from Spiegelman's own perspective rather than his fathers. This choice leaves another degree of separation between the reader and the events of the Holocaust because the reader is not interpreting from Vladek's perspective, instead they are learning from Spiegelman's understanding of Vladek's recollection. Spiegelman's recreation of events is told entirely as his interpretation of his father's memories that depicts events with an incredible degree of accuracy, even going as far as to convey his father's accent within his use of English and include elements that Vladek asks (within the dialog) to be omitted. Maus is an exemplary case of postmemory visualizations of the Holocaust because the narrative structure is dictated to the reader; there is no room for the reader to even attempt empathetic substitution because of how the author controls the narrative advancement of the plot along with how sporadically the narrative alternates between the past to the present day, an effect that leaves the reader wanting more without pausing to consider how they themselves would react in the same scenario.

Visualizing the Holocaust through any medium, not just augmented reality, has the potential to be a traumatic experience to the audience regardless of the author's intention simply due to the visceral nature of the subject, but by adhering to the concepts presented by Libeskind and Hirsch, the effect can be contained and channeled into a positive learning experience that does not needlessly traumatize the learner and maintains respect toward the victims. Libeskind's approach was that of distraction: creating an oppressive and unsettling atmosphere for audiences that causes distress for a reason other than the educational topic to force learners to focus upon their own discomfort rather than projecting themselves into the Holocaust. For architecture this is a sound concept; however, a designer for augmented reality will be limited by the environment of the museum or heritage site they are creating for, and reshaping it to fit this goal is an unlikely outcome. While theoretically possible to create an equally oppressive atmosphere with augmented content, it must be tailored to enhance the surrounding environment rather than conveyed using digital characters as this would once again create a precedent for empathetic substitution. Hirsch's approach relies upon a narrative figure detached from the events of the Holocaust but still connected to it, usually by a parental figure. This may be difficult to implement for an augmented reality scene because the concept relies upon a controlled narrative that allows audiences to comprehend the trauma but not falsely believe they have experienced it themselves, a concept which is only possible if a narrative element can be employed within the scene. Any attempts to visualize the Holocaust for augmented reality must carefully analyze their design methodology to employ a theorist's model when applicable and remain vigilant to prevent empathetic substitution from taking place.

Visualizations that are not created by secondary witnesses may be restricted with a lesser degree of artistic freedom than those such as Maus. Due to the Maus being a graphic novel, it is a non-multimedia visualization that takes curious artistic liberties with its portrayal of different groups as anthropomorphized animal people; the Jewish characters are mice, and the German characters are cats. As a visual metaphor, this sends a powerful image to the reader regarding the social hierarchy of German-occupied Poland but is only appropriate due to the author's narrative structure and his relationship to a survivor, because should this metaphor be attempted by anyone else it could be perceived as an offensive depiction. For instance, throughout history, mice have been considered vermin, and cats were kept as pets to hunt them down and keep them out of ones dwelling to prevent them from stealing food or spreading disease. If a non-secondary witness attempted to create the same visual metaphor as Spiegelman, it could potentially be interpreted as an insult toward the Jewish people by likening them to rodents. Spiegelman's connection to the Holocaust through his father allowed his interpretation to be presented without bias nor prejudice, but without a similar connection, another visualization in a similar manner could be read as propagating the *led to the slaughter* metaphor, and for this reason, visualizations for educational purposes should be kept to realistic depictions instead of stylized ones.

Augmented reality stems beyond mere visualizations with the technology allowing for interactions from the user that can stem from limited engagements to full game mechanics if the creator sees fit to include them, but this may not be an appropriate usage of the technology. As augmented reality for Holocaust education is still a relatively new field, there are not many examples to draw upon; however, visualizations for educational purposes are a much more established field that have attempted many interaction methods that have achieved mixed results. An example of this would be Playing History 2: Slave Trade [35], an interactive computer game designed to be an educational experience to teach children about the Atlantic slave trade of the eighteenth century. The player takes the role of a young slave serving as a steward aboard a slave trading vessel and interacts with a variety of characters to learn about the slave trade along with how the practice had impacted or influenced their lives. Part of the gameplay included minigames that were intended to teach about specific topics, but one of these minigames infamously caused outrage on the internet and in the media due to how grossly insensitive it was toward the subject matter: *slave Tetris*. Intended as a gameplay interaction to depict how tightly slaves were kept within slave trading vessels, the player had to play a game of Tetris with slaves as the blocks in an attempt to fit as many as possible into the storage space of the ship. During an interview with Klapek [36] regarding this minigame, the developer asserted that it was intended to be an accurate depiction of slave stacking (the method by which slaves were restrained during transport to fit as many as possible into the available space) and that the minigame was intended to communicate how absurd and cruel the practice had been. Although intended to be educational, the implementation of this minigame was interpreted as being insensitive and a completely inappropriate way to depict the suffering of slaves by trivializing the terrible conditions they were kept in, and consequently the developers were pressured into removing it from the game entirely. Even though built with good intentions, the poor reception of this one gameplay element was ultimately detrimental to the entire game and its reputation as a learning resource, and for that reason considering gameplay elements for a Holocaust visualization is ill-advised. A gameplay feature, even if it may be considered an engaging way for learners to interact with a topic and further their understanding of it, can be utterly detrimental to the finished product because there is no appropriate way of creating gameplay mechanics for the Holocaust without trivializing the horrific conditions and suffering that were endured by the victims. For this reason, engagements should be limited to interactions without gameplay features, because as Pepper [20] claimed, the Holocaust is not and should never be a playground for learners.

17.6 Visualization Methods for Digital Reconstruction

Technological development has evolved exponentially in a manner of decades and culminating in technologies such as augmented reality along with the means of digitally reconstructing parts of the real world in virtual space. Creators of learning experiences for augmented reality now have multiple options for asset creation that is no longer limited to manually creating a 3D model; however, there are challenges with some of these new methods that include logistical issues of deploying them, financial costs of acquiring the necessary hardware or even the skill level required to effectively utilize them. For this purpose, is it necessary to identify which methods are suitable for visualizing artifacts and scenes from the Holocaust for their usage within augmented reality?

Visual rendering in augmented reality is unfortunately more limited than with other technologies due to the operational hardware employed by mobile devices or bespoke peripherals. Although hardware is constantly improving with time, the levels of detail that can be smoothly rendered on a mobile device is still far from the standard of visuals that can be attained on a computer or a console, consequently presenting the challenge of how should artifacts and people be presented in augmented reality space. Historically, when the computer games industry was building toward the same limited standards of a device such as the Microsoft HoloLens, there were two methods of visual presentation: as realistic as possible despite the limitations of the hardware or with a stylized aesthetic. The realism approach involved attempting to make all assets look as closely comparable to their real-world counterparts as possible while adhering to the limited rendering capacity of the console or computer running the game. At the time, this was considered an acceptable compromise as players had no alternative frame of reference to compare to, but in the modern day where exceptionally accurate visualizations have become the standard including a full range of advanced rendering technologies, it is possible that audiences may reject visualizations that are built using older techniques as they look dated. Tamburro [37] details this in his article on the aesthetics of retro computer games and why they have aged poorly, citing older expectations of visual standard and comparisons against modern visualizations as the main reasons.

Regardless, the alternative to using the realistic method for visualizations would be to instead opt in favor of a stylized aesthetic: the act of depicting the rendered world in a way that does not conform to visual conventions imposed by the real world. This results in aesthetics such as cartoon visuals with overly vibrant or desaturated color palettes or characters that are presented with impossible anatomical proportions and exaggerated strength and any other deviation from the real world as desired by the artist defining the style. The primary advantage to this method is the subversion of expectations of the audience; the viewer does not expect the visuals to adhere to the standards of the real world, and because of this, the limited rendering capacity of the hardware can be adapted to present a visual style that will be interpreted as unique rather than a dated attempt at realism. Context is an important factor when determining visual style however, as these methods worked for computer games but may not be appropriate for environments such as a Holocaust museum in which accuracy is an important factor. When comparing against postmemory [33], the concept of presenting Holocaust survivors and their experiences in a stylized fashion was a method for the children of those survivors to interpret the stories told by their parents, but as aforementioned, attempts by others to do the same could potentially be interpreted as offensive, which also creates a difficult middle ground to navigate as stylized presentations can be a more suitable approach to educating children. When building an augmented reality learning environment for other applications, a stylized approach could be a very sensible direction for compensating for the limited rendering capacity and providing visuals appropriate for younger audiences, but for teaching the Holocaust, the realistic approach will maintain a higher degree of accuracy and lessen the risk of accidentally trivializing the subject matter with cartoon effects.

Creating assets for the realistic visualization workflow has become much more accessible in recent years due to the prevalence of technologies that allow for digitalization of real-world objects instead of dependence upon an artist to accurately create them within a polygon modeling software package. This includes the usage of technologies such as photogrammetry and laser scanning, which can generate 3D models and record depth information without rigorous modeling or sculpting practices required and allowing for a higher degree of accuracy when depicting the finished result. For history educational purposes, accuracy is the most important attribute in a visualization, which results in the usage of these technologies becoming increasingly desirable, but is one method more appropriate than the other?

Photogrammetry for digitizing objects or even people is not a new concept, with a comprehensive photogrammetry setup having famously been used to create a presidential bust of Barack Obama [38] as far back as in 2014. Photogrammetry works by taking a series of photographs of the subject, which can then be used to generate a 3D model inside a software package along with correctly projected textures, allowing for a mesh that accurately represents the original object. The challenge of creating an asset in photogrammetry is one for a photographer rather than a 3D artist, as the photographs need to correctly encompass the subject from all angles and be lit with neutral diffuse lighting to ensure no shadows are cast, as these would also be recorded in the texture projection. While this is very attainable for reconstructing artifacts, the challenge of digitizing a human person is substantially more difficult because of the logistical challenge of taking the required photographs without the subject moving, blinking or adjusting their facial expressions in any way between the shots being taken. Capturing humans using photogrammetry requires a much more advanced camera system, including a series of cameras instead of just one, positioned in a rig surrounding the subject from a broad range of angles, a lighting system to keep their facial details as even in diffuse as possible and an activator system to trigger all the cameras simultaneously. A setup such as this one was presented in an interview by Hideo Kojima, the creator of the game *Death Stranding* [39] about how the technology was used to digitize actors for the game [40] and how a complete 360-degree set of captures were required to create accurate character models. Such a complex setup would guarantee the most accurate results attainable, but there are two factors to consider before attempting this: the first is the setup cost, which could be substantial; the second is the level of detail captured against what can be presented in augmented reality. *Death Stranding* was produced for the PlayStation 4 console and had access to its full rendering capacities to present exceptionally accurate renders of the actors; however, mobile and bespoke devices used for augmented reality cannot attain these visuals and consequently not all of the detailing will be able to be presented. This should not be a compelling argument against using photogrammetry for digitizing people, but the decision must be made when considering digitization method regarding the cost of implementing such a complex photogrammetry rig and if it justifies the expenditure. Regardless, using photogrammetry for artifacts is still a sensible data acquisition approach to ensure all details are captured (Fig. 17.1).

An alternative data acquisition method is 3D laser scanning, which is not new but requires hardware that was not widely available until recent years. Laser scanning can capture detailed surface information of the subject and create

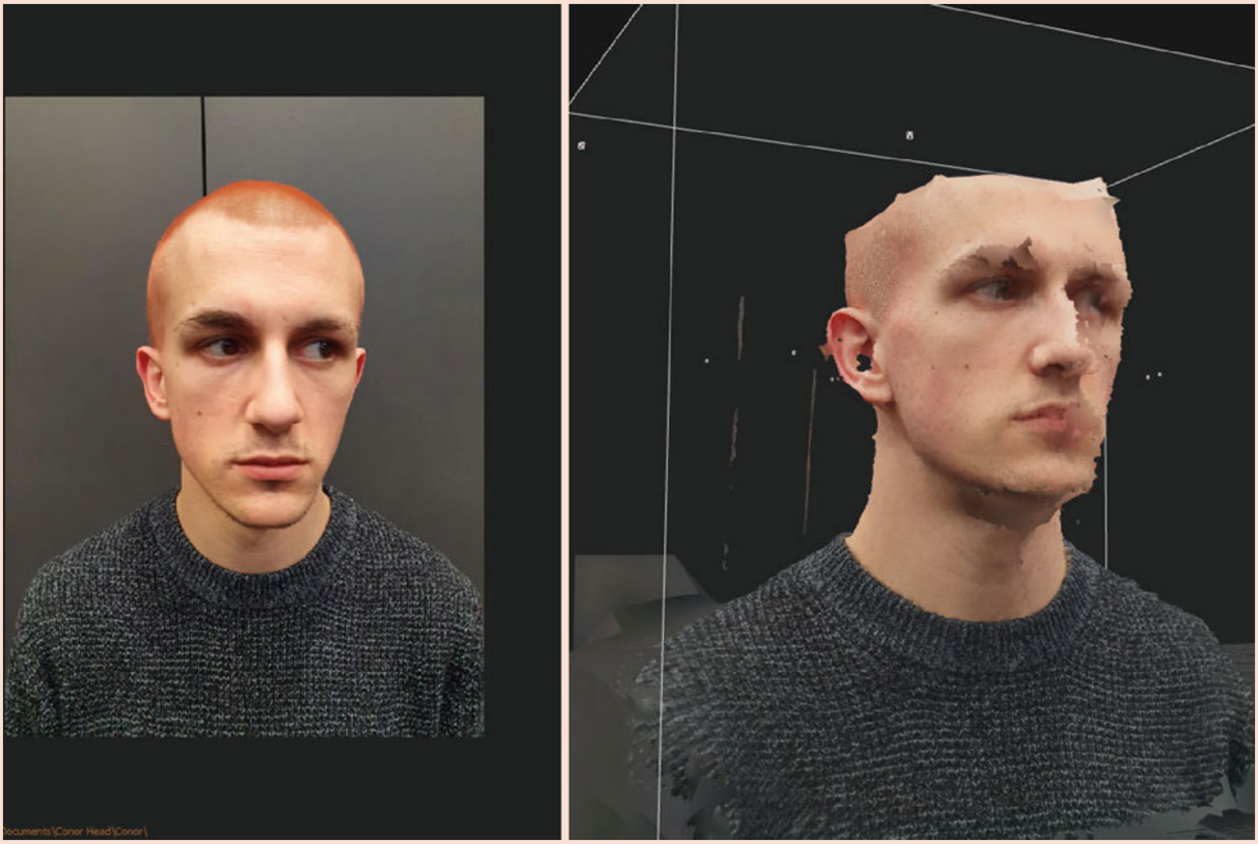


Fig. 17.1 Photogrammetry of Conor-Jack Molloy. (With permission from Conor-Jack Molloy)

a 3D model of it in real time. The process can be repeated as many times as required from different angles with the results combined by software to make a single mesh. The level of detail attainable by laser scanning is exceptional, with some scanners offering calibration down to scales equal to fractions of millimeters. However, laser scanners cannot capture texture information as photogrammetry does, and so photographs must be taken if the intention is to project texture data onto the final mesh. Capturing artifacts with this method is relatively simple as no lighting conditions are required for the laser to function optimally. However, this method shares a problem with photogrammetry when attempting to digitize a human subject because any facial movements will distort the collected data, and human hair must also be covered as scanners cannot interpret it. Unlike photogrammetry, building a complex rig to capture from multiple directions would not be a viable option due to the astronomical costs of purchasing multiple scanners, and even if this was attempted, the reconstruction software works in real time as opposed to the batch rendering system employed by photogrammetry software, meaning that custom software would also be required to make the idea into reality. Laser scanning for human subjects is a new field as many of the available scanners were built for the automotive industry rather than digitizing people, meaning further development may yet happen to expand this field, but as of now, the niche remains vacant. Looking once again to augmented reality, laser scanning suffers the same caveat as photogrammetry regarding the level of detail captures versus what can be feasibly rendered on a mobile or bespoke device; however, through the lens of pure accuracy, the scanner will obtain drastically more surface information including skin pores and wrinkles, which translates to a much more sophisticated normal map, but will not include any albedo data for texture purposes. Comparably, the full photogrammetry setup is superior for digitizing human subjects providing that the cost of implementation can be met, but if not then the laser scanning approach would also provide an accurate digitization (Fig. 17.2).

Laser scanning and photogrammetry, while being revolutionary tools, have a high cost of entry due to the equipment or software required. For institutions or individuals without the budget for such tools, the traditional alternative of manually creating the models using a software package

Fig. 17.2 Laser scan of Conor-Jack Molloy. (With permission from Conor-Jack Molloy)

such as Blender, Modo, or Autodesk Maya is still an option that is still being utilized by augmented reality projects, such as one performed by Gimeno et al. [41], which created 3D assets using 3DS Max. This study used 3DS Max to create assets for a house, furniture, and eight character models for a museum, all of which were made by hand and animated using motion-capture data. This study did not attempt to create detailed models for environmental pieces due to its usage of mobile augmented reality and instead used simplified models that would blend with their real-world counterparts as the furniture in the museum would not be moved. Creation of 3D assets in this manner can still produce exceptionally detailed and high-quality models depending on the skill of the artist creating them; however, there are some potential complications that must be addressed. The first potential complication is scale; laser scanning produces models that are identical to their real-world scale, while both laser scanning and photogrammetry capture results that are proportionally accurate. Hand-modeling assets require either precise measurements to be taken prior to the modeling process or a remarkably good eye for estimation, which sounds novel but is often inaccurate. Another complication is the level of detail; photogrammetry and laser scans will capture all shapes, forms and details of their target object, whether that be in terms of texture information or high-frequency surface information on the created mesh. When creating by hand, the artist is responsible for ensuring that these are recreated to the best of their ability and as accurately as possible, which may involve using multiple software packages if necessary. The other major complication is the time investment required; manually creating a 3D asset can be time-consuming depending upon the level of complexity of the asset and how experienced the artist is; the other methods are substantially faster by comparison. These factors should be considered when deciding upon a virtualization method.

17.7 Potential Applications of Augmented Reality for Holocaust Education

Development on augmented reality experiences for its usage within Holocaust education has already demonstrated interest by institutions to utilize the technology to expand pedagogical methods and visualize the topic in ways that were not previously possible in a traditional museum or classroom environment; however, this does raise the question of what possibilities exist that are only possible due to the availability of this technological innovation. Previous research has shown interest in creating narrative pieces or visualiz ing demolished architecture, but what other opportunities are available, and what can be visualized that complies with ethical requirements?

Visualizing artifacts is not a concept that is unique to Holocaust education nor even the field of augmented reality yet is still a distinct possibility for the field. Historical artifacts are often scarce or fragile and consequently are kept away from the hands of the general public to avoid damaging them, instead leaving them to the hands of trained professionals who can appropriately handle them without fear of causing destruction, accidentally or otherwise. Instead artifacts are mostly kept safely behind the confines of glass cases or similar display pieces, which keeps them safe but also prevents learning opportunities for learners to interact with them or examine them in closer detail. Augmented reality offers new possibilities in this regard and allows for opportunities that would not otherwise be available, something which was explored by Alelis et al. [42] during research into emotional responses of age demographics when interacting with artifacts in augmented reality. By utilizing augmented reality technology, the authors digitally recreated a series of artifacts using photogrammetry and made them available for viewing using either a tablet device using printed fiduciary markers that participants could freely move around, therefore providing a means of interaction with the artifacts that otherwise not have been permitted. This design is functional with usage of physical fiduciary markers, which allows for basic interaction with the virtual manipulatives but only within a limited capacity as the user cannot perform any interactions other than movement or rotation. Furthermore, the usage of markers is a somewhat dated method of implementing augmented reality considering that markerless solutions are becoming increasingly viable due to the prevalence of hardware with the capacity to track spatial data. Regardless of method however, by implementing solutions such as this one, artifacts need no longer be confined to cabinets or cases, and audiences become free to examine them or sate their curiosities without the concern of causing damage to an object that may be priceless. Within the context of Holocaust education, there is a large quantity of potential artifacts that could be presented in this manner ranging from personal effects to articles of clothing, old newspapers to propaganda leaflets, and many more, all of which contain educational value. Providing students with these opportunities would not create any ethical dilemmas as there is nothing inherently disrespectful about examining artifacts; however, in the case of personal effects, it may be advisable to get permission from the owners' families (where possible) before attempting to digitize them.

Interactions with artifacts are only the beginning for augmented reality-enhanced education; basic movements and rotations allow for a closer examination of historical objects, but this alone is insufficient for education as there is no context provided for the learner. Augmented reality is not limited to visualization alone; there are many potential uses for the technology including the capability to provide additional information to the learner in formats such as written text, audio and video files, or even gameplay mechanics depending upon the context. Ridel et al. [43] examined this in their research into using this technology for exploring cultural heritage artifacts, within which they took a slightly different approach to using augmented reality by enhancing a museum exhibit using a projector. Instead of utilizing a device through which to render virtual content, the authors made use of an audience-controlled projector to augment an exhibit by overlaying it with additional information. By digitally reconstructing the exhibit artifacts using laser scanning and photogrammetry, the authors were able to create a visualization of the surface details of the artifacts that was then projected onto the real objects in a small circle, similar to that of a flashlight. Learners could then control this *flashlight* with use of one of two control mechanisms: either an electromagnetically tracked physical prop or with their finger using a LeapMotion controller. While certainly a different approach to the concept of augmented reality, this study demonstrates that there are multiple ways of conveying information to learners and the limits of what can be achieved either with visualization or interaction methods have still yet to be reached. For Holocaust education, surface information may not be the most useful information to convey, but there is certainly a requirement to provide context for artifacts or props in order to assist with learning. Simply examining an artifact such as a star of David armband would serve only to show it was an article of clothing; it would not inform the learner of the deeply anti-Semitic political ideologies and social sentiments that lead to the Jewish populace being legally forced to wear them, but this could be educated via the use of augmented content. For ethical reasons, it could even be argued that augmented exhibits are obliged to provide as much accurate information as possible for learners to fully comprehend how overtly malicious the Holocaust truly was.

Understanding the context behind artifacts is essential to understanding their place in history, but simply providing facts is a waste of augmented reality as a medium because the audience could learn this information from reading a textbook or a web page. Considerations must be made into the full usage of the technology to truly reap its benefits along with justifying its implementation over alternative mediums, including the use of narratives to drive the learning experience. Examples of narrative-based implementations of augmented reality have already been mentioned, and it is evident that both heritage sites and museums are interested in them, but as learning experiences, are they pedagogically sound? A study by Yearwood and Stranieri [44] claims that they are, as narrative-based learning experiences can offer alternative perspectives on complex reasoning and are valuable because they are both believable and memorable. The study argues that narrative-based learning environments are beneficial because of their effect on reasoning within the cognitive domain along with both learning and remembering, providing learners with knowledge in a format that they can both understand and retain. Understandably, teaching of any historical period is unto itself an exercise in narrative educational structure with topics being taught as a story, with historic figures presented as characters within it as a means of condensing complicated subject matters into a format that learners can be formally educated in within the short period available for the lesson. Augmented reality learning experiences can benefit from this too with the added benefits of immersion within the physical scene and elements of interactivity; however, as aforementioned, the audience should not be encouraged to participate in empathetic substitution during the narrative, and care must be taken to ensure that viewers are grounded within reality as a spectator and not a participant regardless of whatever interactions they may engage with. Because the Holocaust impacted upon the lives of millions, there are countless opportunities for visualizations for learning experiences, and any attempts to preserve the stories of survivors should be pursued for the benefit of future generations.

Preservation is a key concept for historians across the globe to maintain a record of the past, a task that can be bolstered using visualisation technologies. This concept is logical for preserving artifacts from exhibits or collections but becomes potentially exponentially more difficult when reconstructing objects or locations that have perished or been destroyed over the course of time. While some historic locations or artifacts have been fully preserved, some have only sections remaining, and others are completely destroyed with no lingering trace of the original objects or structures. The challenge of digitally reconstructing these pieces of history will vary considerably depending on the state of the subject along with the availability of documentation or imagery of its original state, and presentation may be limited depending upon how much information can be found to validate the authenticity of the finished result. For the Holocaust, many concentration camps were repurposed or destroyed in the aftermath of the war with no trace remaining of the original structures and only a memorial to demonstrate the historical significance of the location. Augmented reality is the ideal medium for preserving these locations and visualizing them in a manner that allows audiences to be physically present at the original location and gain an understanding of both how the site once looked and the events that transpired within it. Both Kerti et al. [25] and Pacheco et al. [26] have already demonstrated the logistics of this concept and proved that it is possible to create a reconstruction with a limited dataset for educational purposes, and doing so is pivotal for the future of heritage virtualization, especially for locations such as Bergen-Belsen or Sobibor camps where so many died but no longer exist.

The features and development of these varying projects is compared in the table below to identify their differences.

Judging from prior applications, the usage of marker-based augmented reality seems to be minimal in favor of using spatial solutions. This may simply be due to a technological hardware trend rather than methodology but does indicate that marker-based solutions are indeed becoming less preferable when compared to spatial alternatives. Augmented reality solutions requiring markers provide a constant problem for the user; their entire experience is contingent upon markers being placed in accessible locations with adequate lighting for the hardware to recognize them. In a museum environment, this can be problematic as exhibits are often let to reflect the mood of their content, which could cause difficulties for a user attempting to utilize an augmented reality system. The other complication is available space; the amount of space for a user or multiple users to try and engage with an augmented reality system is potentially limited, and this problem is exacerbated when groups are all attempting to keep a device pointed at the same marker. These problems are not specific to a Holocaust museum; however, the trend in other research suggests that other researchers or participants have taken preference to markerless solutions. In a study by Brito and Stoyanova [45], which compared the two methods, the authors created two augmented reality applications for customizing a pair of Converse brand footwear: one with fiduciary markers and the other using a markerless spatial system. Participants were placed into groups, and each tested either the markerless or the marker-based solution, in which they would view a sample pair of Converse footwear that they could rotate and customize using virtual buttons. Afterwards, they were tested on emotional response, usability and intention to recommend the Converse brand. The study found that participants "did not report statistically significant differences" between each solution for emotional response, which would indicate that the response to augmented reality is a response to the technology rather than the methodology by which it was implemented. The study also reported that the markerless solution offered more stimulation to consumers to recommend the Converse brand. These are interesting responses to analyze from the perspective of Holocaust education as emotional response and user engagement are two critical factors of determining the success of an augmented reality educational environment. Unfortunately, it is difficult to use this study as a basis for advisement on Holocaust education because the potential emotional response from learning about genocide should vary rather drastically from the response one may have when shopping for footwear. The stimulation results are somewhat more universal because it suggests that users were better able to engage with a markerless solution,

Application name	Authors	Hardware	Software	Features	Year
Unnamed	Alelis et al.	7-inch Samsung Galaxy Tablet, Apple MacBook Pro (13.3 inch screen)	Unity, 3DS Max, Mudbox, 1234D Catch	Target-based augmented reality	2014
The Revealing Flashlight	Ridel et al.	Razer Hydra, LeapMotion, Qumi Q5 Projector, Playstation Eye camera	ARToolKit, Meshlab	Spatial augmented reality, gesture-based controls, meta-information visualization	2014
Make Your Own Case Study	Yearwood and Stranieri	Personal Computer	Web browser, PHP programming language	Web-based project, no augmented reality	2007
Unnamed (Referred to as Guidance System)	Pacheco et al.	iPad	3DS Max, Maya, Autocad, Unity	GPS based augmented reality (Markerless), touch-screen controls	2015
Interact	Ma et al.	Microphone	Unspecified	Voice control, adaptive responses	2015
MemoryScape	Stapleton and Davies	Unspecified	Unspecified	Spatial augmented reality, marker-based augmented reality, multiple time periods	2011
The Journey	Yunshui Jin	Microsoft HoloLens	Unity	Spatial augmented reality, interactive narrative	2019

which generically applies to all augmented reality solutions. For the purposes of education, a further study comparing the two implementations of augmented reality may be needed, with the focus on a more emotional topic than shopping.

17.8 Conclusion

This chapter set out to determine the measures that should be taken to ensure that augmented reality learning experiences have educational value while not depicting the Holocaust in a manner that would be disrespectful to the victims, and while many studies and viewpoints have been examined, there is no singular definitive answer to this question. However, based upon the viewpoints of the subject matter experts, there are several overlaps in opinions that allow for a broad set of ethical guidelines to be produced that would be applicable to any Holocaust-themed visualization for educational purposes to inform their creation and assist with development.

Defining this set of ethical guidelines is not a simple endeavor as considerations must be made regarding not only pedagogical methodology but also to the emotional connotations regarding this period of history along with how cultural differences influence how the topic is educated. Such variable factors make it impossible to create a single set of guidelines that would be universally applicable; however, there have been recurring themes identified throughout investigation into this matter that can be used to create an appropriate baseline. These are as follows:

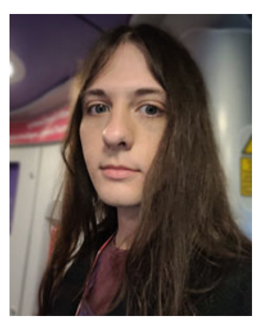
1. Strive for complete accuracy at every opportunity [12, 14, 16, 18]. Details may be omitted when educating younger audiences to avoid exposing them to traumatic themes that they are not mature enough to comprehend, but this is no excuse for revising elements of history to make them more approachable. Show everything as it was, acknowledge when potential inaccuracies arise and do not try to revise history.
2. Keep content appropriate for the experience [18, 29]. The intent of the lesson should always be to educate, not to traumatize the learner.
3. Avoid any possibilities that allow for empathetic substitution [31–33]. Immersion within the learning experience should not correlate to the learner believing they understand how it would have felt to be a victim of the Holocaust as to do so would be grossly disrespectful to those who suffered during that time.
4. Do not turn interactive elements into gameplay features [18, 20, 30, 36]. Interacting with artifacts and virtual characters allows for new possibilities, but creating gameplay features would be highly inappropriate given the nature of the subject.
5. Create realistic depictions of history [33]. Virtual characters and artifacts should be modelled, proportioned, and styled as accurately as possible to the real world to create an authentic experience, even when the technology does not allow for accurate photorealism.

These ethical guidelines, if adhered to, should ensure that future visualizations for Holocaust education are constructed in a way that is appropriate for learners and objectively presents the time period in a respectful fashion. When developing an interactive experience, it is remarkably easy to become fixated on the possible opportunities for creativity, but any ideas formed that are not relevant or appropriate for learning should be discarded. The intention is not to build a computer game or a form of entertainment; the learners have a right to be educated, and the designer has a responsibility to fulfill that right. Developers for Holocaust-themed augmented reality learning experiences must remember that the ultimate goal of creating such an experience is to facilitate learning and enhance understanding beyond what can be provided with traditional mediums, but with that objective also comes the burden of responsibility to create an experience that has been developed to depict this period appropriately.

References

1. Ambrosewicz-Jacobs, J., Eckmann, M., Stevick, D., International Holocaust Remembrance Alliance: Research in Teaching and Learning about the Holocaust: A Dialogue beyond Borders. Metropol Verlag, Berlin (2017)
2. Ellison, J.A.: How they teach the Holocaust in Jewish day schools. Cogent Educ. **4** (2017). https://doi.org/10.1080/2331186X.2017.1310010
3. Suissa, J.: Testimony, Holocaust education and making the unthinkable thinkable. J. Philos. Educ. **50**, 285–299 (2016). https://doi.org/10.1111/1467-9752.12207
4. Charny, I.W.: Encyclopedia of Genocide. ABC-CLIO (1999)
5. Langer, L.L.: Preempting the Holocaust. Yale University Press, New Haven/London (1998)
6. Blum, L.: The Holocaust in American Life by Peter Novick, As a Moral Text. Moral Philos Holocaust, ed Eve Garrard Geoffrey Scarre (1998)
7. Pettigrew, A., Foster, S.J., Howson, J., Jonathan, P.M., Salmons, P., Lenga, R.-A., Andrews, K. (Holocaust educator): University of London, Institute of Education, Department of Arts and Humanities. Teaching About the Holocaust in English Secondary Schools: An Empirical Study of National Trends, Perspectives and Practice (2009)
8. Brunetti, N.: "Holocaust fatigue" a risk at British schools because pupils are not taught anything else about anti-Semitism, says Sir Simon Schama. Telegraph (2019). https://www.telegraph.co.uk/news/2019/03/31/holocaust-fatigue-risk-british-schools-pupils-not-taught-anything/ . Accessed 10 Dec 2019
9. Short, G.: Teaching the Holocaust: the relevance of children's perceptions of Jewish culture and identity. Br. Educ. Res. J. **20**, 393–405 (1994). https://doi.org/10.1080/0141192940200403

10. Short, G.: Teaching about the Holocaust: a consideration of some ethical and pedagogic issues. Educ. Stud. **20**, 53–67 (1994). https://doi.org/10.1080/0305569940200105
11. International Holocaust Remembrance Alliance How to Teach About the Holocaust in Schools|IHRA. https://www.holocaustremembrance.com/educational-materials/how-teach-about-holocaust-in-schools. Accessed 5 Nov 2019
12. United States Holocause Memorial Museum Guidelines for Teaching About the Holocaust – United States Holocaust Memorial Museum. https://www.ushmm.org/educators/teaching-about-the-holocaust/general-teaching-guidelines. Accessed 3 May 2019
13. Hammer, R., Kellner, D.: Multimedia pedagogy for the new millennium – ProQuest. J. Adolesc. Adult. Lit. **42**, 522–526 (1999)
14. Herman, M.: The Boy in the Striped Pajamas. Walt Disney Studios Motion Pictures (2008)
15. Boyne, J.: The Boy in the Striped Pajamas: A Fable. David Fickling Books (2006)
16. Gray, M.: *The boy in the Striped Pyjamas*: a blessing or curse for Holocaust education? Holocaust Stud. **20**, 109–136 (2014). https://doi.org/10.1080/17504902.2014.11435377
17. Spielberg, S.: Schindler's List. United International Pictures (UIP) (1993)
18. Classen, C.: Balanced truth: Steven spielberg's *Schindler's list* among history, memory, and popular culture. Hist. Theory. **48**, 77–102 (2009). https://doi.org/10.1111/j.1468-2303.2009.00499.x
19. Polanski, R.: The Pianist. Pathé Distribution (2002)
20. Spielberg, S., Hanks, T.: Band of Brothers. HBO (2001)
21. Sledgehammer Games: Call of Duty WWII. Activision Publishing, inc. (2017)
22. Rosenberg, A.: Call of Duty: WWII' takes one step forward, many steps back in its handling of the Holocaust. Mashable (2017). https://mashable.com/2017/11/08/call-of-duty-wwii-review-holocaust-critique/?europe=true. Accessed 29 Nov 2019
23. Alien Games Cost of Freedom (2018)
24. Pepper, J.: The Holocaust is no subject for a video game|Jack Pepper|The Blogs. Times Isr. (2018). https://blogs.timesofisrael.com/the-holocaust-is-no-subject-for-a-video-game/. Accessed 20 Dec 2019
25. Machine Games: Wolfenstein: The New Order Bethesda. Softworks (2014)
26. Ma, M., Howard, S., Walker, C.: Question-answering virtual humans based on pre-recorded testimonies for Holocaust education. In: Minhua, M., Oikonomou, A. (eds.) Serious Games and Edutainment Applications, pp. 391–410. Springer (2017)
27. Stapleton, C., Davies, J.: Imagination: the third reality to the virtuality continuum. In: 2011 IEEE International Symposium on Mixed and Augmented Reality – Arts, Media, and Humanities, ISMAR-AMH 2011, pp. 53–60 (2011)
28. Jin, Y.: AR Journey testing in National Holocaust Centre and Museum – YouTube (2019). https://www.youtube.com/watch?v=nXnFgRKQYRY&t=37s. Accessed 10 Dec 2019
29. Kerti, J.: Explore Lager Sylt (2017). https://lager-sylt.website/reconstruction.html. Accessed 27 May 2019
30. Pacheco, D., Wierenga, S., Omedas, P., Oliva, L.S., Wilbricht, S., Billib, S., Knoch, H., Verschure, P.F.M.J.: A location-based Augmented Reality system for the spatial interaction with historical datasets. In: 2015 Digital Heritage, pp. 393–396. IEEE (2015)
31. Bathrick, D., Prager, B., Richardson, M.D.: Visualizing the Holocaust: Documents, Aesthetics, Memory. Camden House, Rochester (2008)
32. Libeskind, D.: Daniel Libeskind: The Space of Encounter. Universe (2000)
33. Hirsch, M.: Surviving images: holocaust photographs and the work of postmemory. Yale J. Crit. **14**, 5–37 (2001). https://doi.org/10.1353/yale.2001.0008
34. Spiegelman, A.: Maus: A Survivor's Tale. Penguin Books, London (1980)
35. Serious Games Interactive: Playing History 2: Slave Trade. Serious Games Interactive (2013)
36. Klepek, P.: Backlash over supposedly educational 'slave tetris' video game. Kotaku (2015). https://kotaku.com/this-is-a-screenshot-from-slave-tetris-which-i-can-as-1728483373. Accessed 20 Dec 2019
37. Tamburro, P.: Here's why retro games look awful compared to how you remember them. Gamerevolution (2018). https://www.gamerevolution.com/features/378895-heres-retro-games-look-awful-compared-remember. Accessed 17 Jan 2020
38. Smithsonian Digitization Program Office, USC Institute for Creative Technologies: Scanning and Printing a 3D Portrait of President Barack Obama (2014). http://gl.ict.usc.edu/Research/PresidentialPortrait/. Accessed 17 Jan 2020
39. Kojima Productions: Death Stranding. Sony Interactive Entertainment (2019)
40. Shaver, M.: Kojima Talks Photogrammetry in Death Stranding. Hyperxgaming (2017). https://ag.hyperxgaming.com/article/2256/kojima-talks-photogrammetry-in-death-stranding. Accessed 17 Jan 2020
41. Gimeno, J.J., Portalés, C., Coma, I., Fernández, M., Martínez, B.: Combining traditional and indirect augmented reality for indoor crowded environments. A case study on the Casa Batlló museum. Comput. Graph. **69**, 92–103 (2017). https://doi.org/10.1016/j.cag.2017.09.001
42. Alelis, G., Bobrowicz, A., Ang, C.S.: Comparison of engagement and emotional responses of older and younger adults interacting with 3D cultural heritage artefacts on personal devices. Behav. Inf. Technol. **34**, 1064–1078 (2015). https://doi.org/10.1080/0144929X.2015.1056548
43. Ridel, B., Reuter, P., Laviole, J., Mellado, N., Couture, N., Granier, X.: The revealing flashlight. J. Comput. Cult. Herit. **7**, 1–18 (2014). https://doi.org/10.1145/2611376
44. Yearwood, J, Stranieri, A.: Narrative-based interactive learning environments from modelling reasoning. J. Educ. Technol. Soc. **10**, 192–208 (2007)
45. Brito, P.Q., Stoyanova, J.: Marker versus markerless augmented reality. Which has more impact on users? Int. J. Hum. Comput. Interact. **34**, 819–833 (2018). https://doi.org/10.1080/10447318.2017.1393974



Jennifer Challenor is a PhD student and Lecturer of Games Art at Staffordshire University in the UK. She is researching the usage of augmented reality for history education with a focus on how it can be used to stimulate long-term memory retention .

17

Professor Minhua Ma is the Provost of Falmouth University in the UK. She is a world-leading academic developing the emerging field of serious games. She has published widely in the fields of serious games for education, medicine and healthcare, virtual and augmented realities, and natural language processing in over 120 peer-reviewed publications.

Augmented Reality into Live Theatrical Performance

18

Dan Lisowski, Kevin Ponto, Shuxing Fan, Caleb Probst, and Bryce Sprecher

Contents

18.1 Background . . . 433

18.2 Augmented Reality in Theatrical Performance . . . 434
18.2.1 ALICE Project . . . 435
18.2.2 What the Moon Saw . . . 440

18.3 Discussion . . . 446
18.3.1 Potentials for Nonlinear Storytelling . . . 446
18.3.2 Interfacing for Immersion . . . 446
18.3.3 Accessibility . . . 447
18.3.4 Future Directions . . . 448

18.4 Conclusion . . . 449

References . . . 449

Abstract

Live theatrical performance is an ever-evolving art form in which visionary theater makers are incorporating augmented reality (AR) into performances to connect and engage modern audiences. While theater productions are generally limited by the physical environment's constraints, AR offers a means to significantly expand the types of sets, effects, and stories that can be told. Furthermore, the addition of 3D tracking and interactive projections enables a new performance methodology with exciting new options for theatrical storytelling, educational training, and interactive entertainment. In this chapter, the authors discuss recent inclusions of AR in live performance, present helpful insights for those looking to include AR into productions, and explore the future of AR live theatrical performance.

Keywords

TYA · Theater · Live performance · Motion capture · Safety · Game engine · Kinect

D. Lisowski (✉) · S. Fan
Department of Theater and Drama, University of Wisconsin, Madison, WI, USA
e-mail: dan.lisowski@wisc.edu; sfan22@wisc.edu

K. Ponto
Wisconsin Institute for Discovery/Design Studies, University of Wisconsin, Madison, WI, USA
e-mail: kevin.ponto@wisc.edu; kbponto@wisc.edu

C. Probst
Curriculum and Instruction/Theater and Drama, University of Wisconsin, Madison, WI, USA
e-mail: cwprobst@wisc.edu

B. Sprecher
Wisconsin Institute for Discovery/Virtual Environments Group, University of Wisconsin, Madison, WI, USA
e-mail: bjsprecher@wisc.edu

18.1 Background

The theater director Peter Brook has argued that all that is needed to make theater is for a person to walk across any empty space while someone else watches [1]. Of course, most theatrical productions involve far more than this, but Brook suggests that these are the minimum requirements. Three key elements within this definition will be important to consider as we discuss how augmented reality (AR) can be used in theatrical performance. Brook's definition is framed in the present tense, suggesting an immediacy that distinguishes theater from film—its closely related storytelling cousin. When someone watches a film, they are watching something that has already happened. Film is fixed and can no longer respond to outside input.

On the other hand, when someone watches theater, they are watching something that is happening now. However subtly, the act of an audience member bearing witness changes the performance of the actor on stage. Therefore, while virtual reality technology could enable someone to experience the opening night of *Hamilton*, the Hip Hop Broadway musical that has captured international attention, the experience would be more akin to film than to theater as it happened in the past.

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_18

This immediacy creates a sense of community between the performers and the people watching. This community is yet another feature of theater that separates it from film and is perhaps part of what has made theater an enduring art form, despite the popularity of television and movies. Although VR can provide someone with the ability to experience a performance that is happening live, it may still be difficult for the technology to create the communal nature of theater when everyone is sealed off from one another inside their headsets. AR, however, offers a broader range of possibilities, and as such may be a better method for innovating theater spaces in the twenty-first century.

Brook states that someone walking across a space can be theater. Thus, any space where someone is doing something can become a theater space. There is a long history of researchers exploring the combination of theater and other live performance media with augmented reality. Some of this research has involved augmenting the space itself, while the rest has augmented the human action or interaction. Researchers have developed virtual sets [2], platforms for dance and theater events [3, 4], and the ability to combine actors and robots [5]. Dorsey et al. were some of the first to present research in this subject area [6], designing virtual sets for opera stages with computer-generated projected environments. Sparacino et al. presented an augmented reality platform for dance and theater [4, 7]. This work used body and gesture tracking in conjunction with virtual actors and human actors in performance events. Other types of augmented reality in live performance events include dancing events [8], body painting [9], and games for children with learning difficulties [10]. Jessop described approaches for mapping user gestures for performances [11, 12], providing a framework for developing performance expression recognition systems in live performance, interactive installations, and video games. Jessop's work outlines the tools, technologies, challenges, and opportunities for utilizing gestures in interactive mediums. Benford et al. provided a survey of many augmented performance events in their work, *Performance-Led Research in the Wild* [3]. The work describes how practice, studies, and theory are all interleaved into interactive public exhibitions and how challenges, such as balancing artistic and research interests, can push up against institutional norms.

Much of this work has explored two different aspects of augmenting live performance events. One aspect of augmentation that has been explored is the use of virtual actors. For example, Mavridis developed an augmented reality technology that enabled combinations of pseudo-3D projections and humanoid robots to create a mixed reality theater piece [5]. This work is similar to the presented work in that it aimed to empower the actor but focused mainly on the human and virtual character interaction as opposed to the user interface. Another aspect for augmentation of live performance is the projection of virtual backdrops and sets. For example, Jacquemin et al. described an implementation of interactive projective sets [2]. More recently, Marner et al. presented *Half Real*, which demonstrated a theater production that featured projected environments, actor tracking, and audience interactivity [13]. Lee et al. have also demonstrated projection-based augmented reality for dynamic objects in live performance events [14]. The presented work's motivation is similar to these works in the augmentation of stage performance, building upon them to enable new modes of stage-based interaction.

The onstage hologram concerts by Hatsune Miku provide a unique experience for fans of the virtual vocaloid singer. Developed in 2007 by Crypton Future Media in Japan, Hatsune Miku is part digital 3D avatar and part music production software. Miku is capable of singing any lyrics a user provides. Despite being an entirely digital persona, both in vocal talent and appearance, through the use of onstage holographic projections, she can perform live at concerts [15]. This blending of fictional personas and real-life spaces has been seen in other musicians' live performances such as the Gorillaz. They frequently use projection mapping and onstage holograms to allow the fictional band members to appear before their fans [16]. These touring acts demonstrate the viability of fictitious personas appearing live before audiences using augmented reality technologies, which pushes against the boundaries of what constitutes live theater.

The IETM, the International Network for Contemporary Performing Arts, comprises more than 500 performing arts organizations from around the world and represents individuals who create theater, dance, circus, interdisciplinary live art forms, and new media. At their 2016 Spring Plenary Meeting in Amsterdam, keynote speaker Joris Weijdom presented on mixed reality and the future of theater [17]. At the close of his remarks, he quoted Shakespeare and reminded his audience that "all the world's a stage, and all the men and women merely players." In doing so, Weijdom challenged theater makers worldwide to think not only of ways that technology could augment and enhance a performance on a traditional stage but also how technology might enable theater to be made by anyone, anywhere.

18.2 Augmented Reality in Theatrical Performance

The following two examples of uses of augmented reality (AR) in live theatrical performance demonstrate the two versions of Weijdom's challenge for the future of theater making. In the ALICE Project, AR technology allowed all of the production's technical elements to be controlled by the performer. The result of which is perfect synchronization of technology and human movement. Next, *What the Moon Saw* utilized AR to facilitate a theatrical production in a nontraditional performance space and create opportunities for audience interaction and agency. These two projects, viewed together, show the breadth of technical possibilities of integrating AR into theater. At one end of the spectrum, *What the Moon Saw* presented an augmented performance environment that was so intuitive that young audience members rarely needed any help navigating the world of the play. At the other end, the ALICE Project relied on a highly trained and well-rehearsed performer to safely and effectively execute the visual spectacle of Lewis Carroll's Wonderland in front of an audience.

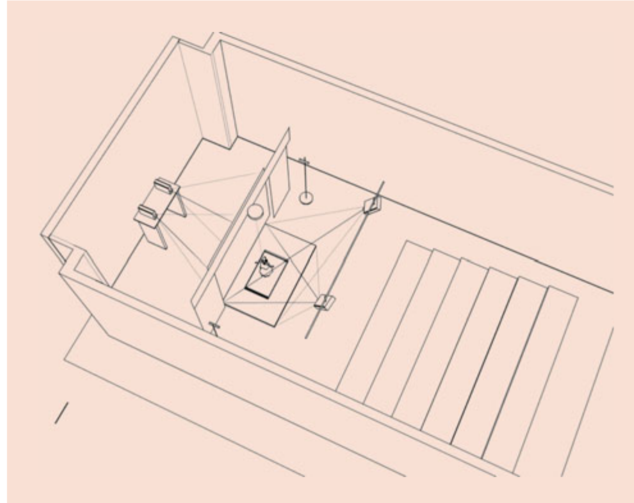


Fig. 18.1 Projector placement

18.2.1 ALICE Project

The technology behind most live theatrical performance events has been standardized into multiple entertainment control systems. Traditional performance practice dictates that each of these systems has an operator to run it, and this work is in conjunction with, though independent of, the onstage performer. The performers' movement either influences the operation of these technology systems or is driven by the output of the technology of these systems. For example, a spotlight operator follows the movement of a dancer, while an actor adjusts their speaking tempo to sync with a recorded video. The philosophy of our Augmented Live Interactively Controlled Environment (ALICE) Project is to place the control of these systems with the performer. This dynamic shift was developed and tested through multiple performances in the spring of 2014. Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation (WARF).

The ALICE Project is a multidisciplinary interactive production methodology that melds traditional theatrical production disciplines with emerging technologies. The ALICE Project enables the performer (i.e., actor, dancer, musician, etc.) to simultaneously interact with and control multiple aspects of a dynamic stage environment. By integrating video projection, motion control, motion capture, a video game engine, and virtual reality technologies together, the project enables new possibilities in live performance that enhance the experiences of both the performer and the audience.

Technical Components

The ALICE system was comprised of several interconnected, automated control systems that interacted with the actor and responded to the motion of their body. The system consisted of three main elements: motion sensors, a video game engine, and a stage motion control system. The actor's joint locations were continuously monitored on the stage using multiple Microsoft Kinect v2 sensors that transmitted these joint locations as data to both the video game engine and the stage motion control system. The research team elected Unity as the video game engine to provide the interactive content environment because of its physics modeling and object interaction capabilities. The computer running the custom simulated Unity world powered a multi-projector system. Two projectors provided a rear-projected display behind the performers, thus preventing the performers from occluding the projection. Two ceiling-mounted stage projectors displayed images on the stage floor around the performer, allowing the virtual world to spill out onto the stage (Fig. 18.1).

The video display responded to how the actor moved within different operational zones of the stage motion control system. One such zone was the treadmill. As the actor locomoted on the treadmill, the Kinect sensors provided the positional data necessary to maintain their position in the center of the zone, regardless of pace, while the Unity world responded accordingly. Thus, when the actor walked, the environment changed slowly; when the actor ran, the environment changed quickly. The augmentation of the stage space in this way therefore created the sense of moving through vast environments.

Similarly, the flying zone made it possible for the actor to create the sense of both flying and falling. The actor's hand positions controlled both the height at which the performer flying system raised them above the stage, as well as the speed at which the video game world changed. For example, hands straight up created rapid falling, but hands out to the sides would slow the apparent falling. All these motion functions were safeguarded against failure by SIL 3-rated functional safety features (Safe Limited Speed, Safely

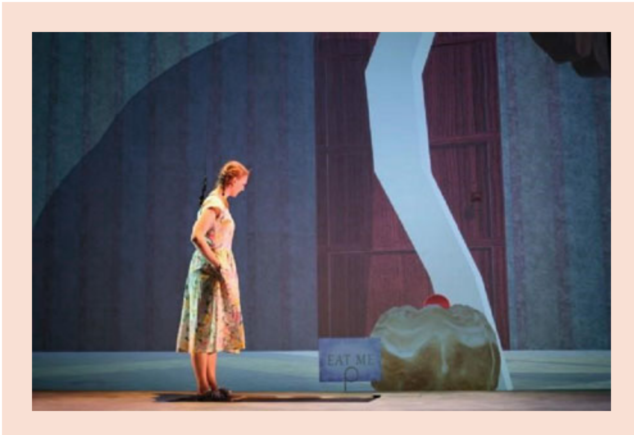


Fig. 18.2 Actor using flying performer system

Limited Acceleration, etc.). This research project utilized IEC 61800-5-2 (adjustable speed electrical power drive systems - Part 5-2: Safety requirements - Functional) functional safety features to enhance performer safety in this unique performance environment (Fig. 18.2).

Technological Affordances in Performance

The research team chose to adapt Lewis Carroll's *Alice's Adventures in Wonderland* into the script for this stage performance. Following the start of the show, no external input, except for one hold-to-run safety button, was required for the actor to create an Alice character who truly lived in the imagined worlds created by Lewis Carroll. The actor's movements, captured via multiple Microsoft Kinect v2 sensors, directly controlled all the show systems. Their position on the treadmill was maintained at the center regardless of the speed at which they walked or jogged. The projected world behind Alice scrolled along to keep pace via a communication link between the automation system and the Unity video game engine, which managed the digital world. This allowed Alice to interact with projected images and traverse the world via the treadmill (Fig. 18.3). For example, when an animated White Rabbit appeared, Alice could follow it. Unlike previous theater methods, however, the actor was not beholden to the predetermined and fixed speed at which the White Rabbit would run away. In the ALICE Project, the actor playing Alice could walk curiously, quicken slightly, and then break into a full run. The dynamic system responded to ensure the White Rabbit would always remain one step ahead.

When Alice followed the White Rabbit down the rabbit hole, the actor raised their hands and the performer flying system raised them above the stage. The Unity engine responded accordingly, and Alice fell down "a very deep well" [18] into Wonderland, and the higher the actor raised their hands, the faster they appeared to fall (Fig. 18.4). Position

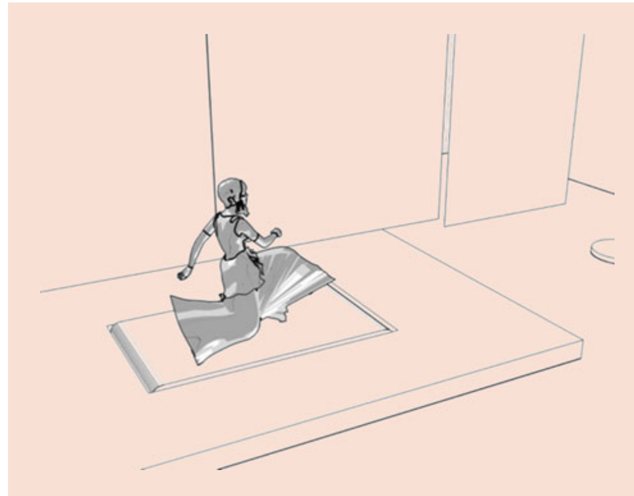


Fig. 18.3 Actor on treadmill

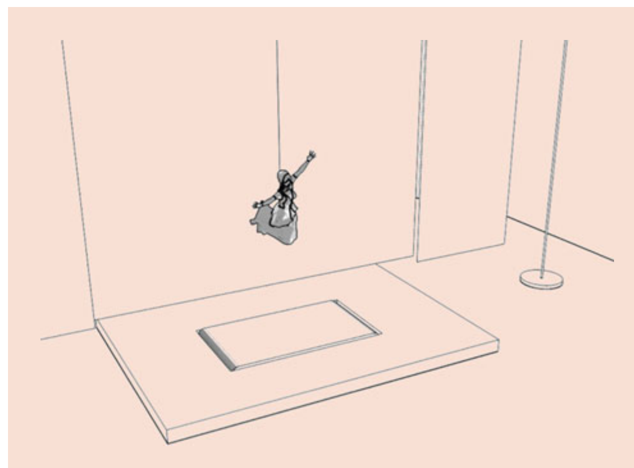


Fig. 18.4 Actor appearing to fall

in the stage environment was detected by the Kinect sensors and dynamically controlled by the automation system.

The inclusion of automation into the ALICE Project is what makes this performance methodology genuinely unique. Traditional performance practice dictates that each specialization (e.g., performer flying system, video projections, etc.) has an operator who runs each system. Within the ALICE Project, however, the actor is in control of the stage environment. Through the use of multiple technologies, the research team has developed a stage environment in which the actor dynamically controls lighting, sound, projection, and automation systems (Fig. 18.5). This liberates the actor to live more fully within an imagined environment.

Safety Considerations

Safety was a primary design consideration throughout the development of this entertainment technology. The aforementioned hold-to-run enable button empowers an external safety observer to halt the automation system if needed to maintain safety. This safety function was one of several utilized in the system. Formal Risk Analysis and Risk Reduction (RA/RR) processes conducted throughout the development determined that multiple levels of safety were required to protect the Alice performer throughout the performance. Highlights of the safety system include dual-zone sensing to activate the treadmill and lift zones; advanced variable frequency drive technology using Safe Limited Speed (SLS), Safe Limited Acceleration (SLA), and Safe Limited Position (SLP) [19]; and use of an E1.43-compliant performer flying hoist system [20]. The performer flying system allowed for performer-controlled flying within the normal operational zone while preventing overtravel and excessive speed or acceleration (Fig. 18.6). The integration of these safety features with purpose-designed interactive automation control software allowed actor-driven automation to be performed without incident throughout the performances.

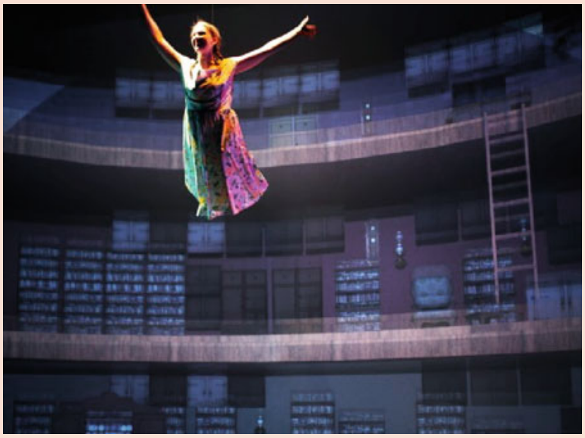


Fig. 18.5 Flying controlled by actor

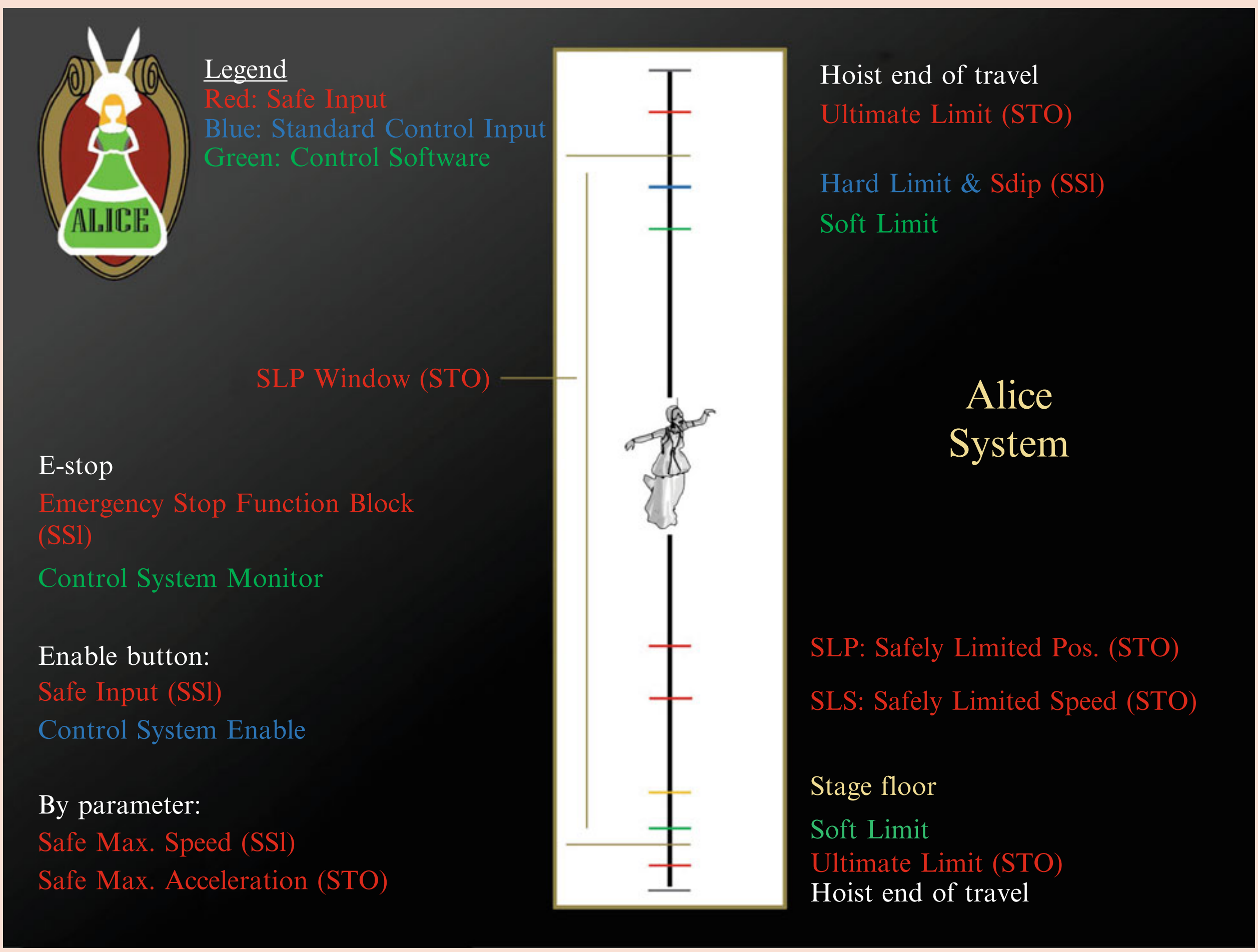


Fig. 18.6 ALICE Project performer flying safety diagram

Sensor Considerations

With the performer's safety as a primary goal, the consideration of the limitations and reliability of the sensing systems used is an important element in such interactive design. These limitations can be intrinsic to the capture technology used or extrinsic environmental factors that influence the sensing of the performer. Intrinsic limitations include the resolution, range, capture rate, depth-sensing technologies, and spectrum visible to the camera systems. Extrinsic factors include the size of the tracked space, the lighting and visibility of the tracking targets, and the number of targets to be tracked. Two broad categories of sensing involve augmenting the performer or augmenting the space. Augmenting the performer might include passive motion capture reflective targets, active trackers such as the HTC Vive trackers, or accelerometer and positional data from activity trackers and smart devices. While some of these involve adding tracking cameras or base stations to the environment, they all also involve premeditated preparation of the performance participants. Another approach is to augment the space such that it is capable of directly sensing the desired data. This was the route taken in the ALICE Project, using two Microsoft Kinect v2 cameras.

The Microsoft Kinect family of sensors use infrared light to generate their depth data. However, the technique varies based on the model. The first-generation Kinect is a structured light sensor, which projects a known pattern of infrared light onto the scene and calculates depth information based on the resulting changes to that pattern. This can result in errors when the two sensor's structured light patterns interfere with each other. Subsequent Kinect models use a time-of-flight camera system that also suffers from competing infrared signals in different use cases. For example, stage lighting, or anything very hot, serves as a potent infrared source that can disturb the sensing capabilities. Despite limitations, infrared depth-sensing confers a powerful advantage over visible light depth systems. That is, they can sense depth without visible light. This is of particular interest in theatrical productions where one faces a high dynamic range in the intensity, configuration, color, and coverage of lighting. Using infrared-based systems facilitates the tracking of performers in complete darkness, allowing performers to be tracked offstage or, in this project, while in midair. Visible light-based stereo systems are not without their advantages. Other depth cameras that utilize disparity provide a significantly longer-depth tracking range in well-illuminated venues such as outside. Choosing the sensor best suited for project needs involves considering the environment of the performance and the goals of the data captured.

Data Considerations

While the Kinect sensor primarily captures depth, the system can try to infer human body positions and movements. As these systems provide the system's best guess, body parts that are occupied or are in unusual positions can be incorrectly identified. Fortunately, the Kinect v2s used in this project can indicate the confidence levels of joint positions reported. The system may report points with poor confidence for a number of different reasons. For example, when a performer's body is completely out of the range of the sensor's view, the system will report no confidence. Low confidence points include those that are at the edges of sensing thresholds or those that are occluded. Occluded points will have their joint locations predicted by the Kinect's machine learning kinematic models. Medium and high confidence points have their predicted joint location correspond with actual points visible to the sensor (Fig. 18.7).

The distinction between predicted and sensed points is particularly important for the automatic control systems used in augmenting the performance in the ALICE Project. As stated above, the Kinect generates joint positions by running machine learning algorithms trained to detect human body poses from depth information. This is highly beneficial for robustness and usability in general application development. It allows for a consistent representation of the body pose and position to be maintained during occlusion, interference, or range issues. However, there is no guarantee on the accuracy or precision of the predicted models. While an incorrect guess is not catastrophic when displaying a digital avatar, these inferred points potentially become dangerous when those data are responsible for driving a treadmill or lift system used by a human performer. These safety considerations led the ALICE Project research team to use multiple sensors. Multiple sensors ensured ample high confidence points, overcame occlusion with extra viewing angles, and provided redundancy for safety. Poor confidence points could be rejected from the control systems to ensure the performer's safety and the reliability of the control system. In practice, the developed system provided a stable representation of the actor, enabling the seamless transitions between the various operational zones around the stage.

Considerations for Working with Performers

One surprise the research team discovered from working with actors and performers was how different this group of people was from human subjects in other research studies. Most actors' training prepares them to quickly adapt their performance to align with the technology's capabilities. This, however, created an unexpected tension between the researchers and the actors. The researchers sought to design technology that adapted an actor's physicality, but this was a completely foreign experience for the actors who were unaccustomed to having their individual preferences attended to so closely. This was particularly true as the team fine-tuned the flying and treadmill zones for a particular actor during the design process. Because the team build the flying

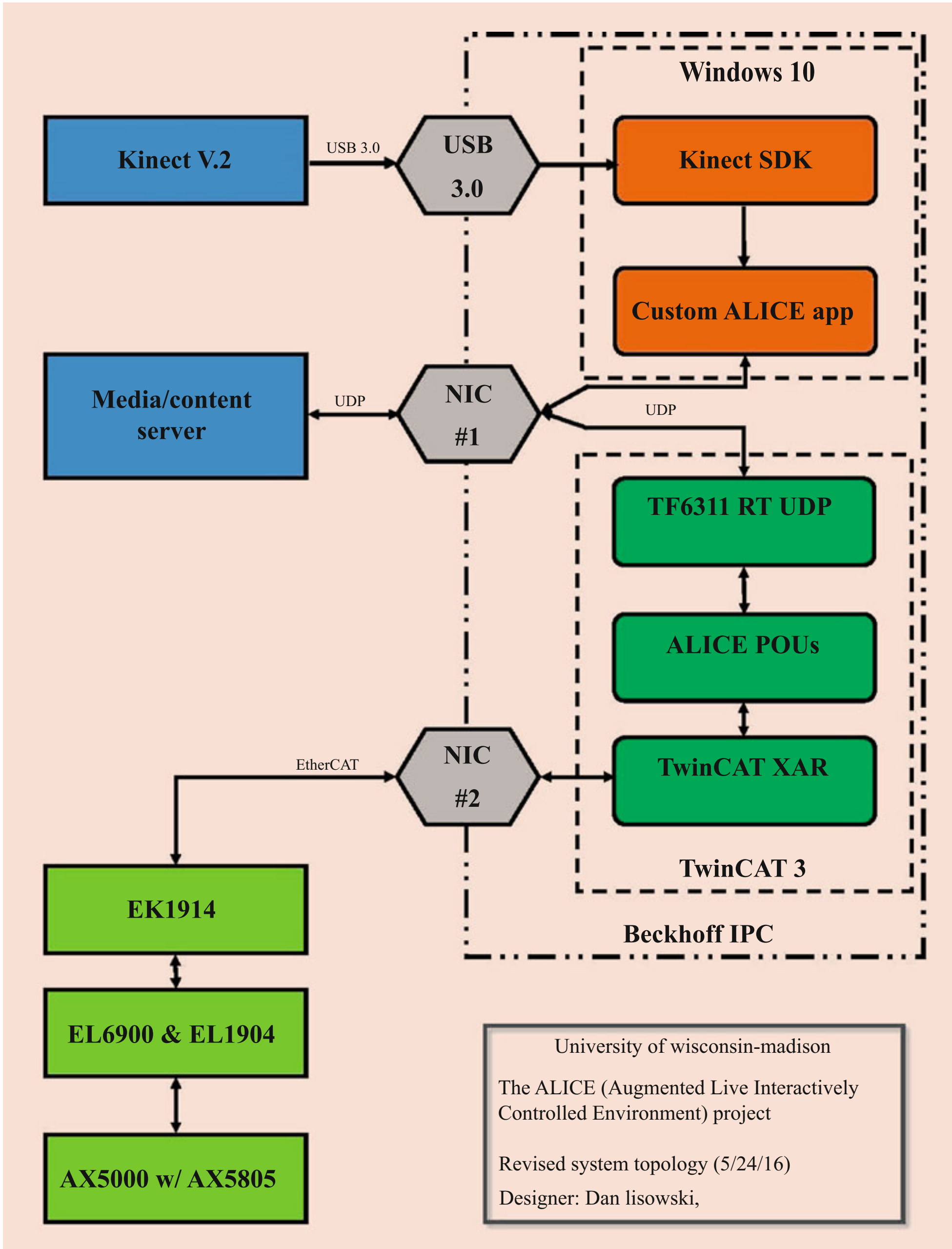


Fig. 18.7 ALICE Project system topology

18

system specifically for the actor, they could not directly test much of the user interface. Instead, they had to rely on the actor's responses to changes in various parameters. As a result, the only way to test the responsiveness between the actor's actions and lift, for example, was through trial and error (Fig. 18.8). This led to many conversations in which the research team would create a series of mock-ups and ask the actor which one felt most natural, only for the actor to respond that their preference was irrelevant and that they would do whatever seemed to fit best the overall aesthetic of the production.

Another challenge for the actors came from interacting with 3D spatial triggers for which the actors could not see. These triggers were set up such that from the audience's point of view, the actor's actions would align with virtual objects (Fig. 18.9). For instance, the actor could grab a virtual bottle by passing their hand through the space the bottle would physically occupy. Unfortunately, as the projected perspective was given for the audience and not the actor, the actor was without visual cues as to where these physical/virtual collisions should occur. While actors are trained in hitting marks, these generally occur in 2D space, for instance,

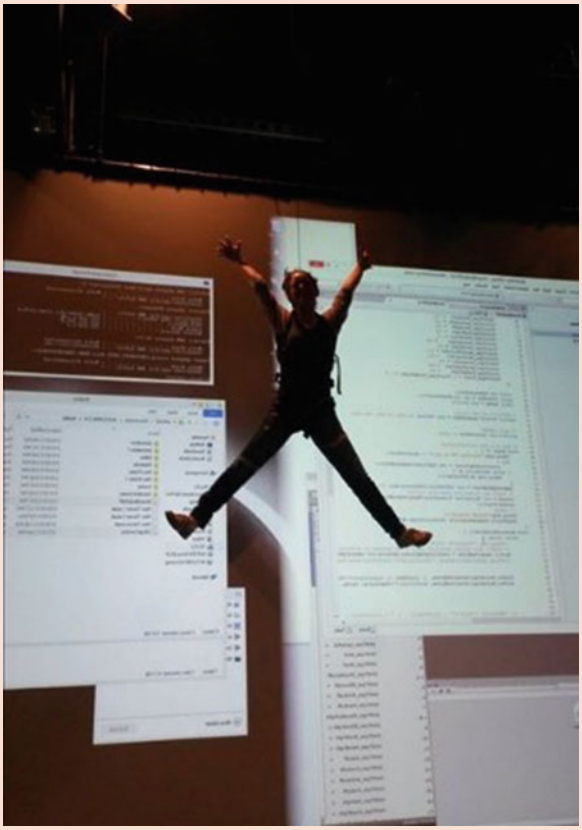


Fig. 18.8 ALICE Project actor testing the flying system

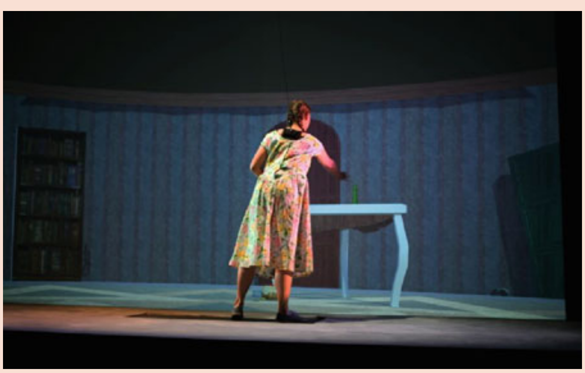


Fig. 18.9 ALICE Project actor reaching for a virtual object

making sure one is standing in the correct location on the stage. The adjustment toward enabling an action to move through a volume of space with limited cues sometimes proved to be a challenge to the performer.

Finally, having actors be part of the technological development process proved to be a unique experience all around. As both groups were unsure of what was possible, either physically or technologically, both groups needed to learn how each other operated. Much like other interdisciplinary research forms, while integration challenges existed between the groups, the payoff is more than made up for it.

18.2.2 What the Moon Saw

Lewis Carroll's *Alice's Adventures in Wonderland* contains fantastic environments that were previously difficult to produce for audiences given their strong preconceived visual notions created by major motion pictures. However, the technology of the ALICE Project afforded the creative team the ability to create an augmented live performance that satisfied these cultural expectations for the audience. The significance of this technology for the practice of theater making is the shift in the actor's ability to interact with the created narrative world, but this remained a passive experience for the audience. Building upon these developments of using video game and interactive technologies in live theatrical performance, the research team worked to develop an environment that would facilitate interactive possibilities for young audiences. To do this, the team elected to adapt Hans Christian Andersen's *What the Moon Saw* [21] into a dynamic and interactive play for children. These efforts yielded a new storytelling methodology that allowed for nonlinear storytelling and audience interaction within augmented reality performance.

To fully develop an interactive production, the project team developed a new script based on Hans Christian Andersen's fairy tale *What the Moon Saw* that sought to incorporate the new performance methodology. The original tale is a loosely connected set of 32 vignettes framed by a story of the Moon paying nightly visits to a lonely child, who is Andersen's nameless protagonist. As the Moon tells stories of what he has seen in his travels around the world each night, the child fills a sketchbook with drawings inspired by the tales. Because each vignette within the original Andersen text essentially functions as a unique and self-contained story unit, the order in which an audience encounters them does not matter. This made *What the Moon Saw* an excellent source of material for the creative team to adapt into a nonlinear story. For the purposes of providing a coherent narrative frame for the adaption, the team's playwright crafted a story that takes place in the course of a single evening. In the adapted script, the Moon visits a child, whom the playwright named Erika, and shows her various episodes inspired by Andersen's original vignettes. The playwright designed the script in such a way to provide significant audience agency over the narrative, which will be discussed later in the chapter. The resulting play combined augmented reality (AR) technologies with more traditional theatrical conventions.

The Unity game engine was the technical medium through which the team created the audience's interactive theatrical experience. *What the Moon Saw* featured a variety of both interactive and noninteractive scenes rendered by the team's designers in Unity. Some of the noninteractive elements included static backdrops akin to traditional matte paintings, prerendered cartoon-style computer-generated movies that played as a dynamic backdrop, and dynamic cameras that rendered the digital scenes using Unity's real-time graphics engine. The show's interactive elements included body motion-tracked mini-games for audience members, performer body tracking-driven scene transitions, showrunner-cued transitions, and digital puppeteering for audience members. The whole project ran off of a single VR-capable desktop with a dedicated GPU. The use of body tracking for the interactive elements of the performance allowed for a seamless onboarding process of audience participants which gave them a sense of stepping into the simulated world on the stage.

The research team achieved body tracking through the use of one Kinect v2 camera system; the Kinect is a visible and depth camera with machine learning models that enable both human body and pose tracking. Data generated by the Kinect streamed to Unity by the Body2Basics Microsoft app, which provides 25 body joint positions for up to six humans. The researchers mounted the Kinect on a frame above and downstage of the performers, allowing full tracking of space in front of the projected screens (Fig. 18.10). An active USB 3.0 extension cable was necessary to connect the stage-mounted Kinect to the backstage rendering and projection systems. This data was fed into the Unity rendering system to enable the input to the simulated reality, which was then output to the users enabling the interactive experiences.

Simulation output was through two theater-grade Epson 1080p projectors which the team vertically oriented to rear project the AR environments. Manual keystoning and lens offsets allowed the team to tune the displays specifically to the projected surface. The performance space for *What the Moon Saw* contained a wall of frosted glass doors that served as the projection surface and the actors' entrances and exits. The mobility of the projection surface allowed for immersive transitions by the performers. Through opening doors in the projected surface, performers were able to step out of the virtual world of the digital game and into the stage's physical world.

Fig. 18.10 Actors standing in front of static projected image

Running off a single computer and using a single Kinect camera lent a great deal of portability and ease of setup and takedown to the system. As part of the performance space's limitations, the team had to remove the entire technical infrastructure after each rehearsal and then reassemble it before the next rehearsal session. The system was versatile enough to enable rapid deployment, as well as to augment the performance space.

18

Agentic Affordances of AR in Theater for Young Audiences

The research team behind *What the Moon Saw* utilized live motion capture technology, facilitated through a Microsoft Kinect v2 and the Unity video game engine, developed in the ALICE Project to augment the performance environment. In creating a new form of performance methodology, it was important for the research team to first consider the unique affordances that AR could contribute to live theater. Perhaps most notably, AR provided young audience members a chance to join actors in the performance space and express agency over the narrative outcome. Young people could enact their agentic capacity in two ways. First, young audience members could interact with elements within the video game world. For example, someone could push a rock within the video game world over a cliff. Second, young audience members could embody figures or characters and control their movements within the environment. This phenomenon could be imagined as a sort of digital puppetry.

Chicken Game

In considering ways to utilize the first interactive affordance, the playwright adapted two vignettes from Hans Christian Andersen's original text into scenes requiring the successful completion of a task-oriented game. During one night in the Andersen text, the Moon shows the protagonist a chicken farm in which dozens of chickens have escaped the coop. A young girl is distraught over the situation because she is responsible for the uncaged chickens and is worried that her father, the farmer, will be angry. This premise provided the foundation for a scene with a game-driven narrative, which is to say successful completion of the game is a vital element for successful storytelling.

At the start of the scene, 30 chickens bounce and squawk about the screen. The actors portraying the Moon and Erika

Fig. 18.11 Joint tracking of young audience members playing the Chicken Game

Fig. 18.12 Young audience members playing the Chicken Game

explained to the young audience that the girl would need help getting the chickens back into the coop and then selected multiple volunteers to come to the stage to herd the chickens. Facilitated through the joint tracking of the Kinect sensors' joint tracking (Fig. 18.11), the young audience members then used their hands and feet to move the bouncing chickens into a targeted chicken coop (Fig. 18.12). By successfully rounding up the chickens, the volunteers solved the girl's problem of the escaped chickens and brought the scene's narrative to its conclusion.

Penguin Game

Similarly, the playwright adapted a second vignette set in the Arctic in which walruses hunt seals into task-oriented game that involved coaxing penguins back into their exhibit at the zoo. In this instance, the actors portraying the Moon and Erika elicited help from young audience members to recapture the penguins before they waddled too close to the walrus exhibit. Again, facilitated through the Kinect sensors, the young audience members then used their bodies to deflect sliding penguins into the targeted penguin exhibit (Fig. 18.13). During the performance, two audience members would play the game, and once they collected all the penguins, the scene would advance after the audience members returned to their seats. This game could support up to three players. The number of escaped penguins could be adjusted for increased difficulty. Seven penguins provided a sufficient challenge during the show without taking up too much time.

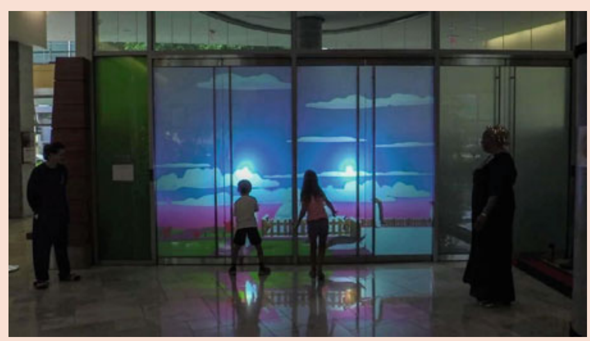


Fig. 18.13 Young audience members playing the Penguin Game

The tasks of the Chicken Game and the Penguin Game were virtually identical. The actors provided the audiences with the ability to choose between the two scenes. Thus, only one of the games was played during an individual performance. Choosing the way in which the larger narrative unfolded on stage was yet another way in which the young audience members exerted agency over the story. This will be discussed later in the chapter.

Swan Game

The final scene of *What the Moon Saw* utilized the second affordance: embodiment. This interaction involved coordinated effort by audience members to help an injured swan fly back to its flock. This included the most specific body tracking of all of the interactions. Using the position and rotation of their shoulders and arms, audience volunteers were able to puppeteer the wings of the swan. Though capable of working with one user, the experience was designed to be a cooperative effort from two players.

The actors brought two volunteers to the stage and asked them to help the swan fly again. Thus, the young audience members were presented with an ill-defined problem that they had to work together to solve. This provided young people an opportunity to cooperate within their agentic control over the narrative. Standing facing the projected screen, the left audience member's left arm controlled the swan's left wing and similarly for the audience member standing on the right. In order to successfully make the swan fly, both players had to flap their wings in unison (Fig. 18.14). If one player flapped too fast or too frequently, the extra force would tip the swan over, and the duo would have to try again. Successfully helping the swan fly off from the lake brought about the narrative conclusion to the scene. Once the swan reached a certain height, a prerendered video of the swan flying away to its flock played and triggered the scene's transition.

Storytelling Affordances of AR

In addition to affording young audience members the opportunity to help complete the story of each scene, the AR created by the video game world of Moon also allowed the audience to choose the direction of the larger narrative as well. Unlike Andersen's original literature version which takes place of 32 nights, this theatrical version takes place in just one night. Like Andersen's protagonist, Erika is lonely and isolated. In this version, these feelings are because her family has recently moved to a new town. The Moon magically comes to the window and offers to show Erika many wonderful aspects about her new home. By design, however, the Moon does not have time to show Erika everything, and this allows the actors to ask the audience to vote for the next scene at the conclusion of each preceding scene.

To make the voting process simple, the playwright wrote two options for the audience to choose between each time. The designers created text to appear on clouds with single words describing the settings of the scene options. In the previously mentioned scenes with uncaged chicken and escaped penguins, the words *farm* and *zoo* appeared. The actors asked the audience to vote and then, utilizing the technology's interactive affordance, slid the appropriate text cloud to the side to trigger the next scene to start.

Central to the story of *What the Moon Saw* is Erika's sketchbook. The graphics designer created the sketchbook that was larger than life and often filled the entire screen. The interactive features of the technology allowed the actors to turn the giant pages of the AR sketchbook and trigger the start of animated drawing that gave Erika's illusion, drawing the elaborate scenes the Moon described.

Fig. 18.14 Young audience members playing the Swan Game

Considerations for Public Interaction

The creative team staged *What the Moon Saw* in a large, open space within the Wisconsin Institute for Discovery (WID) on the University of Wisconsin–Madison. This building was open to the public and is not a traditional theater space. The location was chosen, in part, because of the presence of large, frosted glass walls that would allow for rear projection on one side and audience seating on the other. In using the frosted glass as a projection surface, the team solved one of the primary concerns of attempting this project with young audience members that is potential damage to a projection screen during performances. The team predicted that the ubiquity of touch-screen technology would make playing a large video game fairly intuitive for young audiences. While interacting with the AR environment did not require touching the screen, projecting onto sturdy glass walls mitigated the risk of damage.

The motion capture sensors facilitated the human interaction within the AR environment, which meant that the team needed to consider how and where to place the sensors. Microsoft recommends that Kinect sensors be set between 2 feet and 6 feet from the ground [22]. This, however, is not ideal in a theatrical setting because the sensors would then obstruct the audience's view of the stage area. The team was concerned that placing the sensors on the floor would be a tripping hazard that could cause injury to young audience members and damage to equipment. The team found that hanging the sensors from scaffolding, such that they were 8 feet above the ground (Figs. 18.15 and 18.16), alleviated the tripping hazard without sacrificing sensor efficacy.

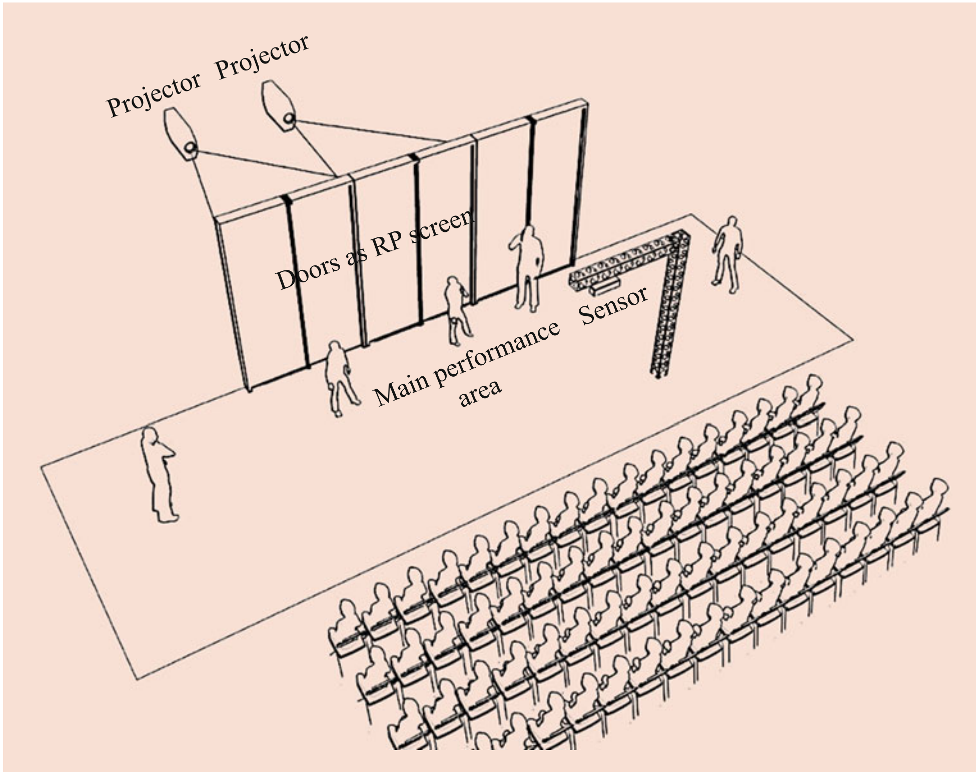


Fig. 18.15 Placement of Kinect and projectors

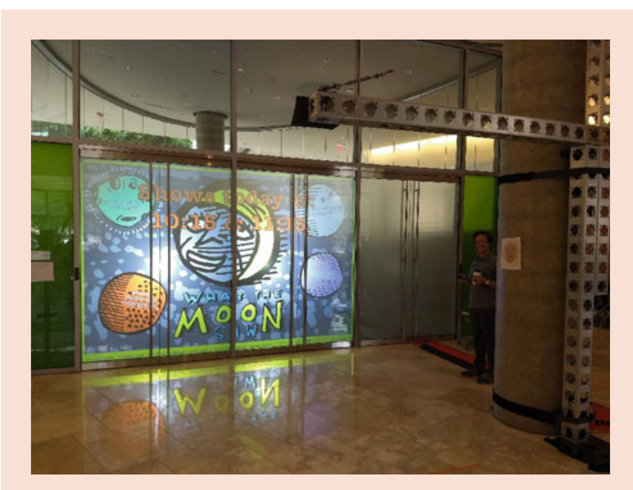


Fig. 18.16 Eye-level view of placement of Kinect with a member of the research team, for scale

Sensor Considerations

Whereas in the ALICE Project performer safety was a primary goal, in *What the Moon Saw,* the sensing goal was to provide a robust and flexible system. In contrast to the single performer tracking in the ALICE Project, *What the Moon Saw* involved tracking a wide range of body sizes and shapes. The audience members ranged from kindergarten to middle school students. Also, unlike the single performer in the ALICE Project, there were anywhere between two and six bodies to be tracked at one time. The research team decided to use a single Kinect v2 as it provided the ability to track up to six targets simultaneously. Needs for redundant data or occlusion resolution were minimized in this application because the data was driving interactive games instead of physical systems with safety concerns.

One of the problems that eluded resolution for quite some time was a phenomenon of ghost bodies. Occasionally, the Kinect would begin confidently tracking floating bodies that did not actually exist. Once the ghost bodies were tracked, the true performers were no longer reported by the Kinect. Fully blocking the sensor's cameras with a piece of cardboard for a few seconds cleared these errors. After further observation, the research team noticed a sliver of light was leaking through the gaps in the projection surface. The heat from the light overwhelmed the camera. The team confirmed this looking through the Kinect's IR diagnostic view and observing the very bright patches coming from the hot, intense IR source of the projector's lamp. This demonstrates the importance of evaluating how tracking systems and equipment might interfere with one another. Another example of these conflicts is the IR laser tracking systems used in Valve's lighthouse system; if placed close to a Kinect, the IR laser beams projected from the Kinect can overpower the IR lasers from the base stations and cause catastrophic tracking failures.

Data Considerations

The Chicken and Penguin Games were achieved by using depth masking information alone. Collisions with the space occupied by participants was sufficient to interact with the Unity physics system that drove the animals' movements. This method has the advantage of quicker implementation and a single object against which to check collisions. It would also allow for many users to be in the space because the sensor tracked occupied space and not individual bodies and joints. However, the Swan Game required the ability to recognize the position of two users and track specific arm movements. This cannot be easily done directly from the depth data but is a task better suited for body and joint tracking. This requirement led the team to forgo using a simple depth masking technique throughout the project instead of using the full-body tracking systems in the Kinect API, application programming interface.

A further consideration that applied to both interaction approaches resulted from relocating the physical mounting location of the Kinect. In the ALICE Project, the Kinect tracked the performer from the front, but in *What the Moon Saw*, the team needed to mount the Kinect above the stage. The research team made this change primarily to prevent young audience members from tripping over the sensor or its cables. Still, this nonstandard positioning required a transformation, measurement, and calibration step in Unity for each performance location. Using the live preview mode in Unity, the performer would walk the space's boundaries. At the same time, an operator adjusted the digital space to correspond with the sensor's reported body positions before each performance.

Fig. 18.17 Humanoid collider shown behind ALICE Project actor

Multiple Forms of Public Engagement

The Wisconsin Institute of Discovery hosts a monthly event called Saturday Science, which is open to the public. Learners of all ages have an opportunity to explore a variety of scientific disciplines through hands-on engagement. In July 2019, the team displayed the motion capture technology and the interactive games at the monthly Saturday Science. Following minor tweaking based on this user testing, the team then mounted the full production of *What the Moon Saw* at the August Science Saturday. In October 2019, the team remounted the production for the WID's annual Science Fest, attended by more than 2000 elementary and middle school students from across the state of Wisconsin. In total, the team produced 10 performances of *What the Moon Saw* that nearly 500 audience members enjoyed.

In between performances, the research team ran the audience participation games as an open arcade experience. Following the conclusion of the show, they invited audience members to play any of the three games from the play. During the performance, two or three audience members would play the game, and once they completed the game, the scene would advance after they returned to their seats. However, this was not a technical limitation, and the game could support up to six players. Six players frequently played the Chicken Game successfully during postshow free play. The number of chickens to capture could be adjusted based on numbers of players and desired game length. Thirty chickens provided a sufficient in-show challenge without taking up too much time. Postshow free play variants included hundreds of chickens for larger groups, creating a veritable sea of chickens that moved in waves like a fluid. The Chicken Game was a clear crowd favorite as it both amazed toddlers and excited teenagers.

While there was an opportunity to play the Penguin Game during the open arcade, the game did not scale up as well as the Chicken Game. When there were too many penguins or too many players, the penguins stacked up and blocked the entrance to the target enclosure or players accidentally blocked the access to the enclosure. If the team made the humanoid colliders visible to the players (Fig. 18.17), then the young people had an easier time understanding that they were blocking the entrance. Though many audience members would play and win the penguin game once, it was the chicken game that people wanted to play multiple times.

The popularity of the Swan Game seemed to fall somewhere in between the other two. For the open play, the designers disabled the video projection trigger from the performance, enabling players to fly the swan as high as possible. Though not prompted by any team member, players would often compete against one another to see who could fly the highest without tipping over. Future considerations will include adding a maximum height score or additional content to encourage this emergent play behavior.

Actors, Audience, and AR

Although the interactive game element of the production proved intuitive to most audience members, the team nonetheless spent multiple rehearsal sessions helping the actors gain expertise in using the Kinect motion sensors and manipulating the motion-based triggers AR platform. While this aspect of the rehearsal process nearly doubled the amount of time needed to prepare the cast of actors for a 25-minute performance, the team needed to have ample time to teach the cast how to navigate the potential need to troubleshoot during a performance successfully. In any theatrical setting, the actors are best positioned to solve problems that occur on the stage. Knowing this, the creative team spent roughly 9 h working with the actors to acclimate them to the technology's various elements.

The AR experience generalized across multiple performers. Over the lifetime of the production, two different sets of performers filled the cast. With minimal adjustments, the team could adapt the AR production to the users' new sizes and profiles for motion capture. This generality also continued in the diversity of the audience's size and body type during their interactive scenes, including audience members who used wheelchairs.

The technical iteration process of getting feedback from the performers during tech rehearsal was particularly vital in the development stage of the project, highlighting the importance of user testing and feedback when designing interfaces and interactions, especially in performance and dynamic technology.

18.3 Discussion

Interactive performance research is a new and growing field of inquiry—this research project investigated both the technology behind the interaction and the human reaction to the performance. The aforementioned ALICE Project was solely focused on developing the initial technology behind the methodology. *What the Moon Saw* extended this inquiry into how to incorporate interactive technology into production, demonstrating that the video game engine environment can be a flexible and powerful tool for use in augmented reality performance. Using motion capture technology, both performers and audience members can dynamically interact with the stage environment. The ALICE Project shows what fantastic spectacles can be achieved with trained performers, while *What the Moon Saw* shows what is possible by simply walking onto the stage. AR and live motion capture allow for a relatively seamless onboarding process, making it possible for anyone to interact with the characters and the story. While examples of mixed reality, augmented reality, and interactive technology are not new to artistic installations, this production methodology's scale and inclusive nature is unique.

18.3.1 Potentials for Nonlinear Storytelling

One of the potential contributions of this research to theatrical performance is the possibility of creating dynamic storytelling environments. In producing *What the Moon Saw*, the research team tested the use of AR technology to enable the audience to choose the direction of the next scene, similar to a choose-your-own-adventure storybook. This endeavor's success demonstrated the viability of dynamically programming story/action options into a video game engine for theatrical performance purposes.

This technical advancement opens new possibilities to playwrights, theater directors, and designers for creating nonlinear storytelling experiences. Nonlinear storytelling is not a new concept to theater makers. The New York production of *Sleep No More* provides a recent example of nonlinear storytelling in theater. In this production, the creative team transformed an entire building into a nontraditional theater space. Actors performed their portions of the story simultaneously in different rooms of the building, and the individuals in the audience were permitted to travel throughout the theatrical environment. As a result, each member of the audience experienced the play in a unique way.

Sleep No More created an analog theatrical storytelling experience similar to how narratives unfold within the created worlds of video games. However, the performance methodology of *What the Moon Saw* created this same sense of agentic control over the story's direction, but it does so in a communal way for the audience. In addition to creating a dynamic and communal experience for the audience, this application of AR into theatrical settings also enables theater makers to utilize traditional theater spaces. Therefore, the building's winding corridors used in *Sleep No More* could be recreated using the technology applied in *What the Moon Saw*, allowing a similar story to be told in a much smaller space.

18.3.2 Interfacing for Immersion

Designing an interface for immersion requires as much a consideration of how to acquire the data as how to use it. Differences in the goals, environment, and interactions between these two projects led the team to take different sensing and data collection approaches. The predictable, controlled, and high-stakes aspects of the ALICE Project led the team to focus on high-quality data, redundant systems, and performer training. *What the Moon Saw*, however, was unpredictable, dynamic, and playful, and this led the team to focus on scalability, simplicity, and emergent spontaneous interaction design. Establishing working prototypes and periodically reevaluating the goals of the interface help lead to a functional result that minimizes overcomplicating either system.

Despite similarities to virtual reality or traditional video games, additional considerations must be addressed when designing for augmented reality in live theatrical performance. As all these systems can be designed in the Unity Game engine, one might assume an entire AR story could also be experienced in an immersive VR perspective in a head-mounted display, HMD. The technical change in Unity from a projected AR display to an HMD unit is relatively straightforward, but the experience would not be so easily translated. Certain goals and assumptions will conflict. For example, when designing for AR, the virtual environments can be designed to facilitate linear movement (e.g., scrolling backdrops). This could be done to avoid the camera accidentally clipping through virtual objects. However, when given the unconstrained perspective of VR, these environments designed to be viewed from a forced perspective look empty. For a fully immersive first-person point of view, it is essential to create scenes that enable arbitrary viewing directions.

A secondary challenge of this approach is in the full integration of physical and virtual worlds. As is exploited in redirected walking research, humans often utilize visual cues to walk in straight lines. When using augmented reality technologies, the user can still maintain awareness of their environment and receive visual feedback. However, in a fully virtual experience, any misalignment in a treadmill system's limited width can cause the user to walk off the treadmill and risk injury. Furthermore, many HMD systems currently require physical connections to computation nodes. These cords and cables can pose serious safety concerns, especially for multiuser spaces. And while the use of HMD is incongruous with the definition of theater used by this team, these considerations nonetheless demonstrate that when designing an experience relying on a strong merging of physical and virtual interactions, augmented reality provides many suitable safety, design, and execution options.

18.3.3 Accessibility

In thinking about the potential for AR technology to create immersive theatrical environments for audience members, it is also important to consider issues relating to audience access to theater in the first place. The research teams of the ALICE Project and of *What the Moon Saw* made deliberate choices to make theatrical experiences that did not rely on the use of secondary devices to mediate the AR experiences for the audience. While there are certainly creative possibilities that can be achieved through audience members using smartphones or tablets to facilitate an experience, some of which will be discussed later, theater makers should balance these design choices against the degree to which they might exclude some audience members from full participation.

The use of mobile devices by people in nearly every demographic has increased significantly in recent years, and an overwhelming majority of school-aged young people indeed have access to a smartphone or tablet [23]. However, it would be misguided to assume that the overwhelming majority of audience members would therefore each be able to have a personal device available to mediate an augmented theatrical performance. Some students who report having access to Web-enabled devices for homework also report that they share the device with other members of their family. Socioeconomic status can already present a barrier to experiencing certain theatrical performances because high ticket prices can make attending events like Broadway plays cost-prohibitive. While supplying devices to every audience member would be a possible solution, many theater companies would likely find such a financial investment equally cost-prohibitive.

In addition to considering ways to make an augmented theatrical performance accessible to people across the socioeconomic spectrum, it is also important to keep in mind audience onboarding. The ALICE Project featured no audience onboarding as audience members needed only to observe the spectacle in the same way they would a traditional theatrical performance. *What the Moon Saw* provided a nearly seamless onboarding because audience members appeared to find the interactive technology intuitive, and the performers modeled its use throughout. If theater makers and designers expect audience members to utilize new technologies in order to facilitate an AR experience, then they should make efforts to ensure that all audience members are comfortable doing so.

Beyond simply taking steps to ensure that audience members understand how to interact with the necessary technology, designers should also take steps to make experiences that are accessible to people with different abilities. For example, holding a mobile device or wearing a headset may not be possible for some patrons. Although it may not be possible to accommodate every audience member's unique situation, theater makers considering AR technology in the design should weigh the options as they would in choosing whether or not to use strobe lighting or the sounds of gunfire. The team wanted to make audience participation in *What the Moon Saw* accessible to anyone who wanted to take part. To do so, the team tested the Kinect sensor's ability to track someone in a wheelchair accurately. They also looked to see that the sensor could track the movements of someone with a missing limb and people of varying height. The Kinect worked well in all of these situations, but this was not necessarily a given.

Finally, creating a live theatrical performance using AR for young audiences meant making the technology elements accessible to children. Some of the safety considerations for protecting both people and equipment have been discussed previously, but this research project provides a deeper lesson about design that can generalize for a wide array of AR applications in theater. Good design should always have the user experience in mind, and in this case, the user is the audience. Some children are short, and so the team made sure that their motion could be tracked. Children are all at different stages in their motor development, and so the team made sure to design the games accordingly. And in addition to countless design decisions made ahead of time, the onstage performers and the showrunner were prepared to step in and assist young audience members navigate the augmented theatrical world successfully.

18.3.4 Future Directions

This project paved the way for different types of future endeavors beyond augmented user interfaces. Consumer commodity computers, smart devices, and wearables are continuing to integrate more and more of the technology required to augment reality. High-resolution screens, positioning and telemetry sensors, and even depth cameras are becoming commonplace in sophisticated smartphones. As our evermore technological society continues, it is reasonable to assume that this trend will enable augmented reality to become ubiquitous. How long that may be is speculative, but we look here to consider what this enables in live theatrical performance. In addition to live motion capture, we also experimented with voice recognition. While only tested as an early-stage prototype, the use of voice recognition presents interesting possibilities for theater and other live performance events. As scripts are well-defined (i.e., predetermined), voice recognition algorithms can be tailored toward prespecified input phrases (i.e., lines in the script). These phrases can be used to advance the narrative structure of the projected display environment.

Further applications of voice recognition technology could serve to make a theater experience more inclusive and accessible for audiences. It is common practice in opera productions to project surtitles above the stage to allow audiences to read translations of the lyrics being sung. For example, an Italian opera performed in the United States might have English language surtitles. Similarly, larger theaters sometimes use a similar practice to make a production accessible to audience members who are deaf or hard of hearing.

The current practice of surtitle projection is limited in several ways. First, they are typically prerendered, which means that if performers were to veer from the script, the surtitles would not reflect this. Second, subtitles generally only feature one language. Third, a theater venue must have architecture to support the projection of surtitles without obscuring the audience's view or creating obstacles for performers. Voice recognition captured live and translated into any one of a variety of languages could then generate subtitles sent to an audience member's mobile device, addressing all of these limitations. By creating what amounts to closed captioning for live performances, this technology would give greater accessibility by expanding linguistic options to patrons. It would also allow for performance art forms like improvised comedy, where translation of this kind has been impossible, to be accessed by more people. Finally, by sending captions to a handheld device, performance spaces that are not conducive to projecting surtitles become more accessible.

Once generated, these captions would not need to be limited to words on a handheld device. Instead, they could be delivered to an audience member's device as audio narration. Audience members with impaired vision could use a text-to-speech delivery method, allowing them to listen to descriptions of the action of the performance. Additionally, action annotations could be automatically generated. Using motion capture and artificial intelligence technologies to recognize performer actions would allow for audio descriptions of the performers' visual actions. This can be leveraged further against the content of a script allowing for meaningful descriptions of the actions within the context of the performance, again making the art form of theater more accessible.

Furthermore, we see this framework as providing a potential opportunity to create training environments for professions that require extensive human interaction. A common challenge in preparing for professions such as teaching, counseling, or courtroom litigation is learning how to respond to the unexpected. While the professions generally rely on reflection to prepare for the future, it is rarely possible to implement a new tactic in a timely manner. Using AR, however, a user could download a scenario, set which characters will be human and which will be virtual, and game through the scene with projected agents and environments. Given the advances of voice recognition since the team's initial experimentation in 2014, and the motion capture technology developed through this research, users could interact with virtual agents capable of responding in real-time to their speech and gestures. For example, a preservice teacher could then have multiple opportunities to practice managing a classroom crisis and explore a variety of possible interventions.

An additional opportunity of the ALICE Project comes from improving the methods for flying systems. As an example, it would be possible to create a system that will augment the actor's jumping abilities. As traditional systems use external operators, these personnel must decide when to pull the actor into the air in the course of a jump. This only allows the actor to react to these pulling forces and can present potentially dangerous situations if the timing is off. While current high-end approaches to these effects utilize a push-button system in the actor's hand to start the flying behavior, thus enabling the actor to be in control, actors can still miss the timing of these button pushes, causing jerking behavior. We see potential in the usage of the tracking system to aid in these types of efforts. The system could determine jumping behavior and the optimum point at which to start the flying motion. This would not only create an effect that would look more natural to audience members, but it would also provide greater safety for the actor.

A final potential application of this approach is to simulate other types of dangerous training situations. For example, this system could be augmented to be used as a jet-pack simulator in which the user would be able to feel the forces that they would expect to feel in flight. While the current rendering techniques utilize a fixed perspective for the audience, having this perspective change based on the user's head-tracked position has been shown to be effective for other CAVE-like systems using Kinect devices [24]. This approach of using motion platforms has been shown to be effective for driving simulators motivating these types of endeavors [25].

18.4 Conclusion

Designing augmented reality into theater and live performance shares many common principles with conventional design. It requires the designer to understand the relationship between the virtual three-dimensional designed objects and the real physical environment. Artistically speaking, implementing AR technology into live performance adds an additional dramatic dimension in the storytelling environment to provide the audience an immersive theatrical experience. Technically speaking, developing an AR environment requires the creative team to work collaboratively to solve the problem of setting the virtual space in real space, setting the virtual interactivity in the physical performance progression, and finding the AR environment control and manipulation solution. Due to AR's digital nature, the implementation still needs to be based on heavy digital content creation, through conventional 3D motion graphic applications and video game engines, similar to VR production.

Other examples of AR experiments can be found from Gorillaz's concerts, Diesel's *Liquid Space* holographic fashion show, and Royal Shakespeare Theater's *The Tempest*. The common artistic and technological solution applied in the above productions are the digital projections based on either the holographic projection technology or the digital projection mapping technology on moving screens to create the virtual objects added into the real theatrical environment for the dramatic illusion. Selecting the projection as the medium for AR in theatrical performance has been the solution, while the AR viewing device, such as the Microsoft HoloLens, is not technically reliable and practically applicable at the present time; however, the AR application in meeting the requirements of the large audience and space might not be limited in those wearable viewing devices as long as the augmentation and the immersive environment are created.

In conclusion, by combining the virtual and physical worlds, the described approach enables play writing to be more innovative and imaginative as many of the limitations of creating physical sets and props can be overcome. This approach enables a new performance methodology with exciting new options for theatrical storytelling, educational training, and interactive entertainment. The work described here demonstrates the possibilities for industry and performance advancements by setting aside prevailing notions of what can and cannot be done.

18

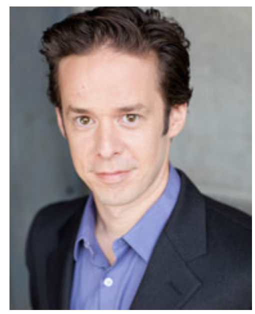
References

1. Brook, P.: The Empty Space. Touchstone, New York (1968)
2. Jacquemin, C., Gagner, G.: Revisiting the layer/mask paradigm for augmented scenery. Int. J. Perform. Arts Digit. Media. **2**(3), 237–257 (2007)
3. Benford, S., Greenhalgh, C., Crabtree, A., Flintham, M., Walker, B., Marshall, J., Koleva, B., Rennick Egglestone, S., Giannachi, G., Adams, M.: Performance-led research in the wild. ACM Trans. Comput.-Hum. Interact. (TOCHI). **20**(3), 14 (2013)
4. Sparacino, F., Davenport, G., Pentland, A.: Media in performance: Interactive spaces for dance, theater, circus, and museum exhibits. IBM Syst. J. **39**(3–4), 479–510 (2000)
5. Mavridis, N., Hanson, D.: The IbnSina center: An augmented reality theater with intelligent robotic and virtual characters. In: The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009. RO-MAN 2009, pp. 681–686. IEEE, Toyama, Japan (2009)
6. Dorsey, J., Sillion, F., Greenberg, D.: Design and simulation of opera lighting and projection effects. In: ACM SIGGRAPH Computer Graphics, Vol 25(4), pp. 41–50. ACM, New York, NY, United States (1991)
7. Sparacino, F., Wren, C., Davenport, G., Pentland, A.: Augmented performance in dance and theater. Int. Dance Technol. **99**, 25–28 (1999)
8. Meador, W.S., Rogers, T.J., O'Neal, K., Kurt, E., Cunningham, C.: Mixing dance realities: Collaborative development of live-motion capture in a performing arts environment. Comput. in Entertain. (CIE). **2**(2), 12–22 (2004)
9. Ip, H., Hay, Y., Tang, A.: Body-Brush: A body-driven interface for visual aesthetics. In: Proceedings of the Tenth ACM International Conference on Multimedia, pp. 664–665. ACM, New York, NY, United States (2002)
10. Ip, H., Kwong, B.: Smart ambience games for children with learning difficulties. In: Technologies for E-Learning and Digital Entertainment, pp. 484–493 (2006)
11. Jessop, E.: Capturing the body live: A framework for technological recognition and extension of physical expression in performance. Leonardo (2015)
12. Jessop, E.N.: A gestural media framework: Tools for expressive gesture recognition and mapping in rehearsal and performance. PhD Thesis (Citeseer 2010)

13. Marner, M.R., Smith, R.T., Walsh, J., Thomas, B.H.: Spatial user interfaces for large-scale projector-based augmented reality. IEEE Comput. Graph. Appl. **34**(6), 74–82 (2014)
14. Lee, J., Kim, Y., Heo, M.-H., Kim, D., Shin, B.-S.: Real-time projection-based augmented reality system for dynamic objects in the performing arts. Symmetry. **7**(1), 182–192 (2015)
15. Jenkins, M.: This singer is part hologram, part avatar, and might be the pop star of the future. Washington Post (5 Jul 2018)
16. Milton, J.: From holograms to headliners – How Gorillaz's inventive live shows have transformed, https://www.nme.com/blogs/nme-blogs/holograms-headliners-how-gorillaz-live-shows-changed-2003855 (March 6, 2017)
17. Weijdom, J.: Mixed reality and the theatre of the future. In: IETM Spring Plenary Meeting (Jul 2016)
18. Carroll, L.: Alice's Adventures in Wonderland. VolumeOne Publishing, Chicago (1998)
19. IEC: Adjustable speed electrical power drive systems – Part 5-2: Safety requirements – Functional, IEC 61800-5-2:2007
20. ESTA TSP: Entertainment Technology – Performer Flying Systems, ANSI E1.43 - 2016
21. Andersen, H.C.: What the Moon Saw: And Other Tales. Routledge, London (2018)
22. Microsoft.: https://beta.support.xbox.com/help/hardware-network/kinect/kinect-sensor-setup-tips (2018)
23. Anderson, M., Jiang, J.: Teens, Social Media & Technology 2018. Pew Research Center (2018). https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/
24. Ponto, K., Kohlmann, J., Tredinnick, R.: DSCVR: designing a commodity hybrid virtual reality system. Virtual Reality. **19**(1), 57–70 (2014)
25. Freeman, J.S., Watson, G., Papelis, Y.E., Lin, T.C., Tayyab, A., Ramano, R.A., Kuhl, J.G.: The Iowa driving simulator: An implementation and application overview, SAE Technical Paper (1995)

Dan Lisowski (MFA, Yale School of Drama; BS, University of Wisconsin–Madison) has been a faculty member at UW–Madison since 2009. As a technical designer, he has worked with Cirque du Soleil, Disney Theme Parks, Universal Studios, and NASA in advancing technology in entertainment. His research work focuses on developing technology to enable new or reenvisioned modes of storytelling.

Kevin Ponto is an Associate Professor at the University of Wisconsin–Madison. Dr. Ponto has utilized VR and AR technologies to better understand health in the home, generate new methodologies for crime scene investigation, create new systems to support informal learning for scientific research projects, and develop new interaction techniques to support next-generation consumer marketplaces.



Caleb Probst received a BFA from Syracuse University, Department of Drama, in 2004. He worked as a theater artist and educator in Chicago for 12 years, and he is now pursuing a PhD in Curriculum and Instruction from the University of Wisconsin–Madison.

Bryce Sprecher is a researcher in the Virtual Environments Group at the Wisconsin Institute for Discovery. They received their BSc in Computer Science and Neurobiology from the University of Wisconsin–Madison. Their research includes acquiring, processing, interpreting, and displaying spatial data for insight, education, interaction, and experience. They enjoy photography, sour apples, comfy over-ear headphones, and fountain pens.

19 How do Tourists Evaluate Augmented Reality Services? Segmentation, Awareness, Devices and Marketing Use Cases

Katrin E. Schein, Marc Herz, and Philipp A. Rauschnabel

Contents

19.1	**Introduction**	451
19.2	**Holistic Augmented Reality Marketing in Tourism**	452
19.2.1	Augmented Reality Marketing	452
19.2.2	Augmented Reality Marketing in Tourism: Use Cases	453
19.2.3	Augmented Reality Marketing in Tourism: Prior Research	456
19.3	**The Role of Augmented Reality in Tourism: Research Questions**	458
19.4	**Study**	459
19.4.1	Research Design and Methodology	459
19.5	**Results**	460
19.6	**Discussion**	464
19.6.1	General Conclusion	465
References		467

Abstract

The increased proliferation of augmented reality (AR) technology has especially impacted the tourism industry, and many players in this sector have made initial attempts to integrate AR into their marketing strategies. Although both the tourism sector and the academic tourism community have made significant contributions to the AR discipline, some gaps remain. While many studies have assessed the drivers behind the acceptance of specific use cases, a comparison of different use cases remains scarce. Likewise, little is known about how evaluations of AR differ in different tourist segments. Therefore, this research aims to examine how tourists evaluate touristic AR use cases and how these evaluations differ between tourist segments. Based on an empirical study among 553 German city tourists, the findings reveal that virtual guides and history AR apps are among the most popular touristic AR use cases and that different tourist segments (explorers, hedonists, and culture gourmets) value the usefulness of tourist AR use cases differently, whereas almost no significant differences can be found when it comes to demographic data such as age and gender. It is thus important to adapt touristic AR use cases to the motivations and needs of tourists – and not so much to the demographic data, as many organizations still do.

Keywords

Augmented reality · AR services · Tourist segmentation · AR marketing use cases

19.1 Introduction

Augmented reality (AR) is an innovative media format that realistically integrates computer-generated information into a user's perception of the real world [1]. This superimposition of virtual content onto the user's real field of view is different from virtual reality (VR), which immerses the user in a completely artificial and digital environment [2]. A popular example of AR for the consumer market is the Pokémon Go mobile app, which allows users to catch virtual creatures encountered in the real world [3].

AR can be applied in many existing mobile technologies, such as smartphones and tablets, as well as in innovative AR-specific wearables, particularly smart glasses [4]. While such advanced wearable devices are expected to seamlessly integrate digital content into the real world, existing AR smart glasses (e.g., Google Glass, Vizux Blade, Nreal, Microsoft HoloLens) must still overcome few technological challenges to deliver full potential and thus make AR a staple

K. E. Schein · P. A. Rauschnabel (✉)
Digital Marketing & Media Innovation, College of Business, Universität der Bundeswehr München, Neubiberg, Germany
e-mail: katrin.e.schein@gmail.com; philipp.rauschnabel@gmail.com

M. Herz
kleinundplaecking, Berlin, Germany
e-mail: mhe@kleinundplaecking.de

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_19

of consumer life [5]. Until then, the smartphone remains the most common device for experiencing AR due to its wide proliferation and accessibility [6]. On a smartphone or tablet, AR can be accessed either through an app or through an internet browser, which is known as web AR.

AR has gained popularity in various disciplines, partly due to its ability to create an enhanced user experience, and multiple forecasts predict increased growth of AR (e.g., [7–9]). For instance, a recent research report by Deloitte commissioned by Snap Inc. [8] expects nearly 75% of the global population to frequently use AR by 2025. Another report by the Boston Consulting Group [7] concludes that "[m]oving forward, we expect the AR ecosystem will continue to develop quickly ... [and] players such as ad agencies, app and software developers, and ad networks are staking out their own roles in the value chain." (p. 1). The broad discipline of marketing provides multiple opportunities for companies to use AR to interact and communicate with a variety of stakeholders.

The increased proliferation of AR technology has especially impacted the tourism industry, and many players in this sector have made initial attempts to integrate AR into their marketing strategies [10–12]. As will be discussed below in more detail, marketing is a broad field that includes a variety of tools subsumed under communicative activities (promotion), decisions about the goods or services of a company (product), pricing decisions (price), and distribution activities (place).

Many tourism companies have made initial attempts to integrate AR into one or several of these sets of tools. For instance, Florence offers visitors information in AR about famous buildings, Vienna offers a scavenger hunt as AR city tour and the cities of Braunschweig and Hamburg developed an AR app that allows visitors on-site to discover how the city looked like in former times. AR is also used for branding purposes (e.g., Stockholm with its AR app "Stockholm is Your Canvas") or for attracting customers (e.g., Buga 2019 with its AR app "impact.Karl"). Many other use cases of touristic AR services have been discussed in academic [10, 13, 14] and managerial literature [15–18].

Although both the tourism sector and the academic tourism community have made significant contributions to the AR discipline, some gaps remain. For instance, while many studies have assessed the drivers behind the acceptance of specific use cases [13, 19, 20], a comparison of different use cases remains scarce. Likewise, little is known about how evaluations of AR differ in different tourist segments – a topic that is highly relevant from a managerial perspective.

Based on an empirical study among 553 German city tourists, our study provides answers to the following research questions: How do tourists evaluate touristic AR use cases, and how do these evaluations differ between different tourist segments?

This chapter will first discuss AR marketing for tourism organizations and multiple use cases. Then we will formally propose three research questions. These research questions will then be tested in an empirical study and will be subsequently discussed.

The results of this study will contribute to a better theoretical understanding of AR in a tourism context. Furthermore, tourism managers can use the findings to develop AR use cases for their target segments. More specifically, they can base their decisions on the findings reported in this chapter rather than relying purely on assumptions.

19.2 Holistic Augmented Reality Marketing in Tourism

One of the many disciplines where AR offers a particularly wide range of opportunities is marketing. In this section, we will briefly introduce the topic of AR marketing as a discipline as well as discuss its application to tourism.

19.2.1 Augmented Reality Marketing

Many people associate marketing with advertising and other communicative activities. In recent decades, however, marketing has developed into a broad concept that ranges from highly strategic long-term decisions to extremely tactical activities, often in real time [21]. Since the impact of marketing extends to the society as a whole, marketers must always consider the ethical, environmental, legal, and social context of their activities [22]. Besides highly visible communication activities (promotion) such as advertising, public relations, and social media, marketing includes decisions about the goods and services a firm offers (product), how they are distributed (place), and how they are priced (price). These four elements (product, place, price, and promotion; the 4 Ps) are subsumed under the term "marketing mix" [23]. Marketing is also not limited to consumer brands. NGOs, destinations, cities, and countries, as well as people and any other organization can apply marketing techniques to communicate to customers and other internal and external stakeholders [21, 24–26].

In recent years, the range of opportunities for marketers to interact with consumers is drastically increasing and changing with the expansion of new technologies [21]. AR, for example, represents a new marketing resource that will dominate and revolutionize the future of marketing [27]. AR marketing is poised to become a mainstream that is easily accessible to consumers, significantly changing the way individuals buy and consume products and services [5, 28]. According to Rauschnabel, Felix, and Hinsch [1], AR will be indispensable for both consumption and marketing. Rauschnabel et al. [29] conceptualizes AR marketing as a strategic concept that uses AR to achieve organizational goals (e.g., increase sales, inspire users, improve branding) by providing, communicating and/or delivering value to stakeholders. Value can thereby be delivered not only in the form of financial value [30] but also by other types of value, such as functional, hedonic, symbolic, or social values, depending on the specific use case. Commercial companies, organizations (e.g., NGOs, NPOs, destinations, cities), and brands can all use AR marketing to address their stakeholders. Thus, next to (potential) customers, stakeholders may also be the public, employees, applicants, and others [29].

Not only academics but also marketing managers are seeing great potential in AR. The Boston Consulting Group [7] concluded that "[m]arketers can expect to have access to a wide array of AR-marketing options in the future" (p. 1). According to a recent report by Deloitte [8], AR is the new consumer experience which cannot be ignored in marketing. Multiple brands (ranging from typical commercial brands to NGOs and destinations) have already recognized that AR marketing offers new opportunities for interaction and communication with consumers. Retailers such as Ikea, Adidas, Lacoste, L'Oréal, and Sephora have already integrated AR technology into their marketing mix [31]. IKEA's AR app for example enables customers to easily arrange, customize, and transform virtual products out of the IKEA catalogue to suit various physical environments, and Adidas recently launched a retail AR experience to reveal the environmental impact of plastic usage. By simply holding a smartphone in front of a screen in the store, customers find themselves in an AR ocean world and are asked to collect the ocean garbage and experience how it turns into small plastic particles, spins into a thread, and transforms into the latest shoe in the collection (Fig. 19.1).

Also destinations are using AR, for example, to engage the attention of potential tourists, to increase sales by promoting their services and products, as well as to inform or entertain visitors on-site. Exemplary use cases are virtual city tours or brochures with additional information and images in AR.

Next to AR use cases in the B2C sector, AR can also be applied in the B2B sector. A typical example of B2B in tourism would be an AR room planner that is used by a hotel building company to discuss and present ideas to the client about how the hotel and its rooms may look. Our study, however, will focus on AR use cases in the B2C sector.

Academics have started to study many AR marketing use cases in more detail and found that user evaluation and perception of AR depend on multiple factors. Factors of selected prior research are depicted in Table 19.1.

Various studies have shown how these factors influence consumer behavior and attitudes. AR can, just to mention a few, increase purchase intentions [5, 28, 34, 40, 42, 47–50], lead to high levels of satisfaction [42] and engagement [51], as well as improve and facilitate decision making [37, 52, 53]. Likewise, Rauschnabel, Felix, and Hinsch [1] show in a pre-post attitude comparison that AR content can improve brand attitudes, especially once users are inspired.

Finally, AR can also generate new forms of data about users and their context (context data). This innovative type of data is futuristic but also highly controversial because of privacy and data security issues. Context data can, however, provide multiple advantages for tourism managers. For example, AR apps can track what objects visitors look at; how much time they spend at certain sights, attractions, or in museums; how crowded places are at certain times; and how polluted environments are [54].

19.2.2 Augmented Reality Marketing in Tourism: Use Cases

The expansion of AR marketing is transforming the tourism industry and changing how destinations are perceived and consumed [27]. According to Wang et al. [12], AR technology has had a significant impact on the tourism sector by changing travel behavior, decision-making, and information searching. The implementation and proliferation of AR has made information and interaction within the tourism and event environment more accessible and engaging, and tourist organizations must find innovative ways to remain competitive in this dynamic industry [47, 55, 56]. Neuhofer [57] summarizes the situation by concluding that "[o]perating competitively in a fast-paced tourism industry means recognizing cutting-edge technological developments and being at the forefront of using them as means for innovation and strategic competitive advantage" (p. 29). Buhalis et al. [58] predict that tourists will demand more AR experiences and will co-create their reality via sensory-technology-enabled tourist attractions.

Tourism managers tend to see AR as a technology for meeting new challenges in the market, and they implement AR in various ways [59]. Huge potential for AR is seen at the pre-booking and information-gathering stage as well as in the enhancement of on-site experience.

AR in the Pre-Booking and Information-Gathering Stage

Integrating AR into the booking and information process is reported to increase upsell in accommodation, travel, and tourism attractions [60]. Olsson et al. [61] explain this effect by underlining the emotional interaction created by AR due to the immediate connection between the company and tourists. This is different from traditional media such as brochures and videos. As tourists experience the destination more intensively, they build stronger emotional connections and memories [62]. Thomson Cruises, for example,

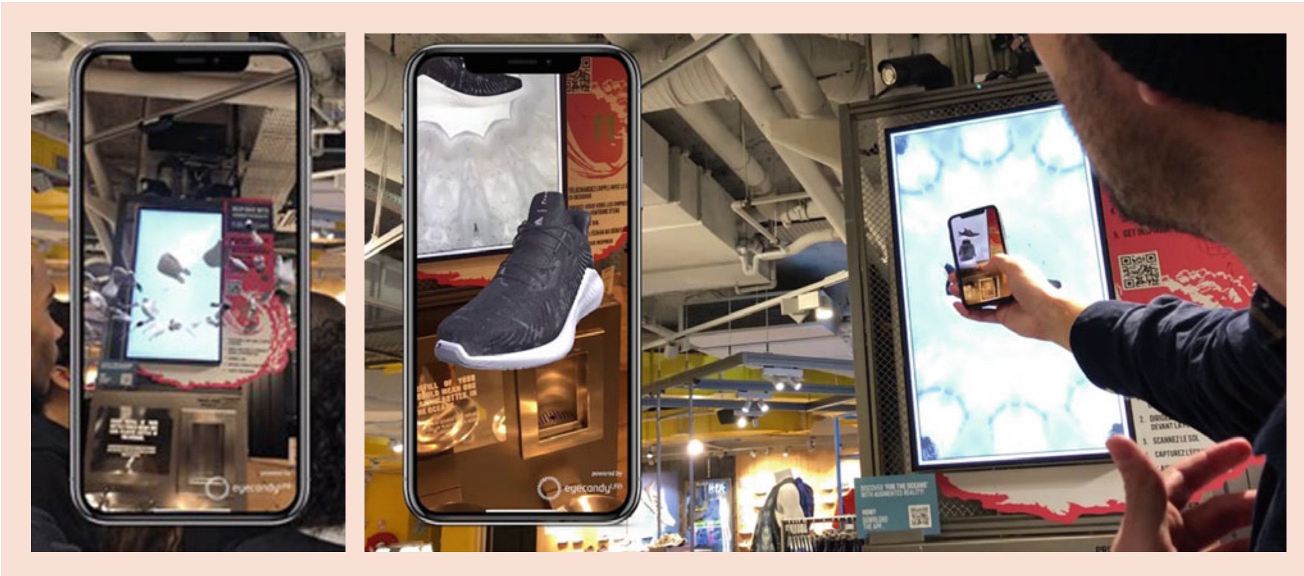


Fig. 19.1 AR marketing offers new opportunities for interaction and communication with consumers (Source: Eyecandylab, www.eyecandylab.com)

Table 19.1 Overview of relevant factors in the user evaluation of AR apps

Benefits	
	Hedonic value Aesthetic, experiential, and enjoyment-related factors (e.g., [1, 32, 33])
	Utilitarian value Functional, instrumental, and practical benefits (e.g., [1, 32–35])
	Inspiration The degree to which users perceive the content as inspirational (e.g., [1, 36])
	Creativity Creative thinking is facilitated and promoted by the use of AR (AR-enabled customer creativity, e.g., [37, 38])
Experiential factors	**Perceived augmentation quality** The extent to which a user perceives virtual content as realistic/well integrated into the real world (e.g., [1, 36, 39–42])
	Flow The holistic experience that people feel when they act with total involvement (e.g., [3, 40, 43])
Absence of risks	**Physical risks** The potential threat to an individual's safety, physical health, and well-being by fully immersing users in AR while being in the normal environment (e.g., [3, 32, 44–46])
	Privacy risks The threat to individual privacy but also other people's privacy since the whole environment is screened and processed (e.g., [3, 32, 44–46])

introduced an AR brochure as a new way of presenting their cruise ships to potential tourists [63], and Lufthansa is making it possible for users from anywhere to step onto observation decks in New York and Hong Kong to experience the countries from a unique perspective in AR (Fig. 19.2).

Stockholm developed the AR app "Stockholm Is Your Canvas" to increase awareness for the Stockholm Art Week and make art accessible also outside the gallery. Stockholm Art Week symbols were displayed on walls everywhere in and around Stockholm City. Once tourists pointed their smartphone's camera at the symbol, an artwork appeared on their display (Fig. 19.3).

On-Site AR Experiences

Next to the use of AR in the pre-booking and information-gathering stage, additional AR services are used on-site to inform, guide, and entertain visitors. AR is already used to help users retrieve tailored and interactive information about monuments and historical buildings in a city while exploring a destination [15, 16] and to provide tourists with accurate information about accommodation, restaurants, and tourist attractions so they can make the most of their experience [16–18].

Marriott, for example, collaborated with LIFEWTR to provide in-room AR entertainment for guests, restaurants

Fig. 19.2 AR in the pre-booking stage (Source: headraft, www.headraft.com)

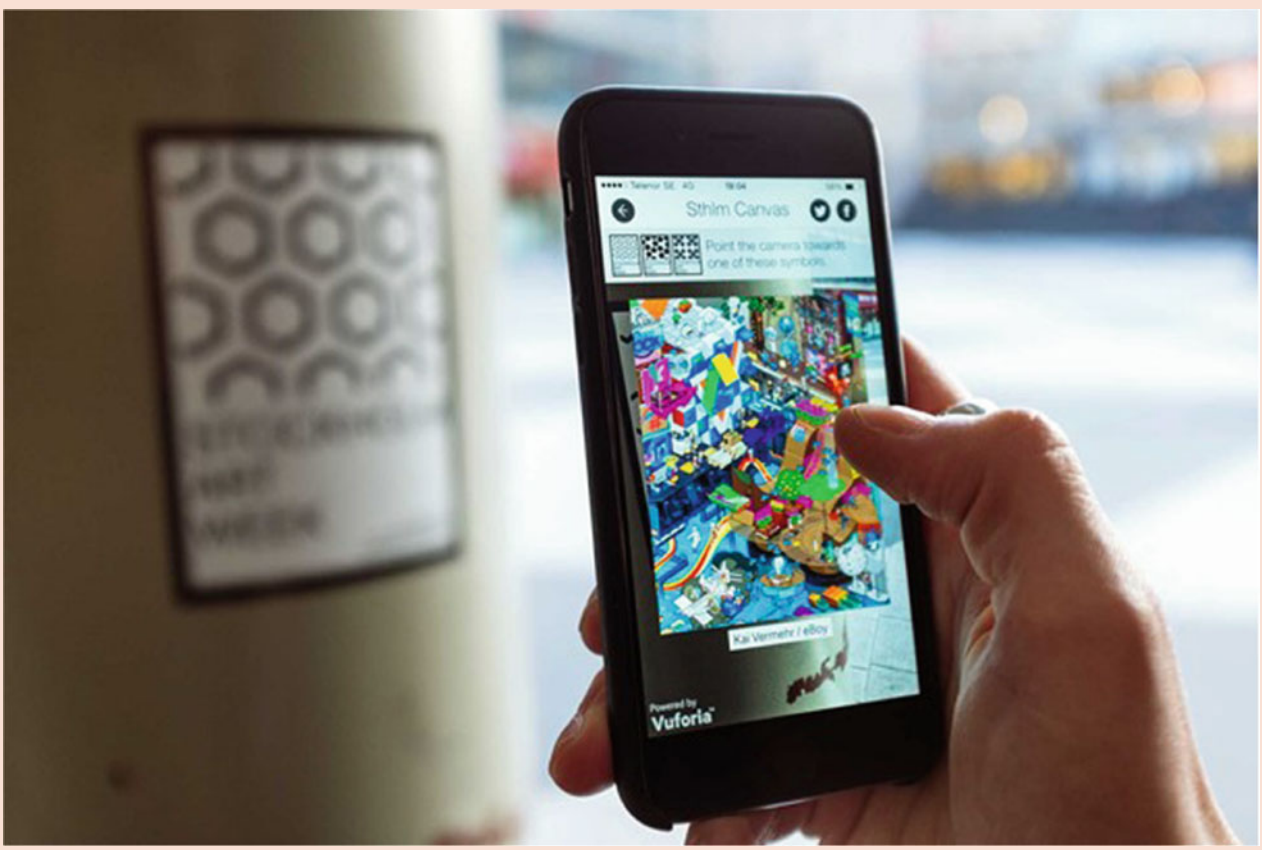


Fig. 19.3 AR to increase awareness for touristic attractions (Source: Stockholm Art Week/M&C Saatchi Stockholm)

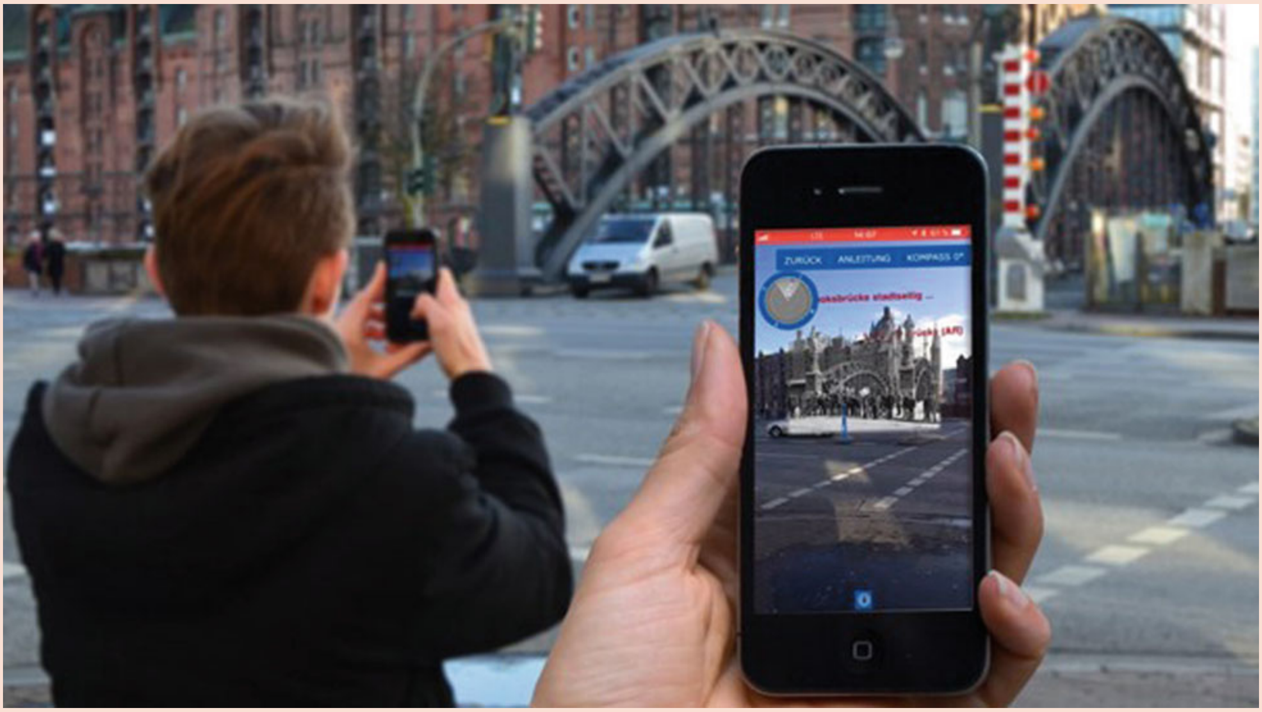


Fig. 19.4 History AR app (© Picture: Dataport; Bildmontage: Fraunhofer FOKUS)

are offering an advanced version of conventional dishes to customers by presenting menus in AR as 3D models, and cities and historic attractions are providing visitors with AR apps so they can witness the evolution of landmarks over time [59]. The city of Hamburg, for example, launched the AR app "Speicherstadt digital" that allows its users to dive into the city's past. On-site, tourists can use the AR app on their own smartphone and look at their surroundings through the screen, where historical images are overlaid and present how Hamburg used to look like (Fig. 19.4).

Sherpa Tours offers tourists to take virtual walking city tours using AR on their smartphone . An avatar guides tourists through the most popular cities, e.g., Paris, Rome, and New York, and informs them about sights (Fig. 19.5).

Another AR service on-site was developed for the Buga (Bundesgartenschau) in Heilbronn with the aim to entertain and thus attract more younger visitors. The AR app "impact.Karl" enabled visitors to explore the Buga playfully by offering several mini-games (Fig. 19.6).

In the following, we introduce nine specific AR marketing use cases in tourism, which will be discussed and evaluated in detail in this study.

As already depicted above, tourism-related entities in the market have launched similar AR activities, however with varying levels of success. Many tourism managers are likely not yet aware of AR's potential and lack a clear understanding of what (potential) tourists expect [59]. As Hassan and Rahimi [64] state, however, the adoption of AR and the exploration of its potential are crucial in the tourism sector to ensure business profitability as well as better products and services. In view of the fact that AR necessitates large investments, it is essential to explore its full potential prior to any implementation decisions [65].

19.2.3 Augmented Reality Marketing in Tourism: Prior Research

An increasing number of studies on new technology in the tourism context have been conducted in recent years in order to explore new frontiers and ways to improve the tourist experience (e.g., [2, 13, 30, 65–70]). Prior research has outlined the potential of AR applications and agrees that AR is a promising tool for enhancing the tourist experience by creating richer, more immersive experiences [19], improving entertainment and engagement [71], increasing interactivity [20], and creating more personal products [72]. It has led in part to increased numbers of visitors [67]. Jung and tom Dieck [65], for example, found that integrating AR technology enhances the intention of consumers to visit

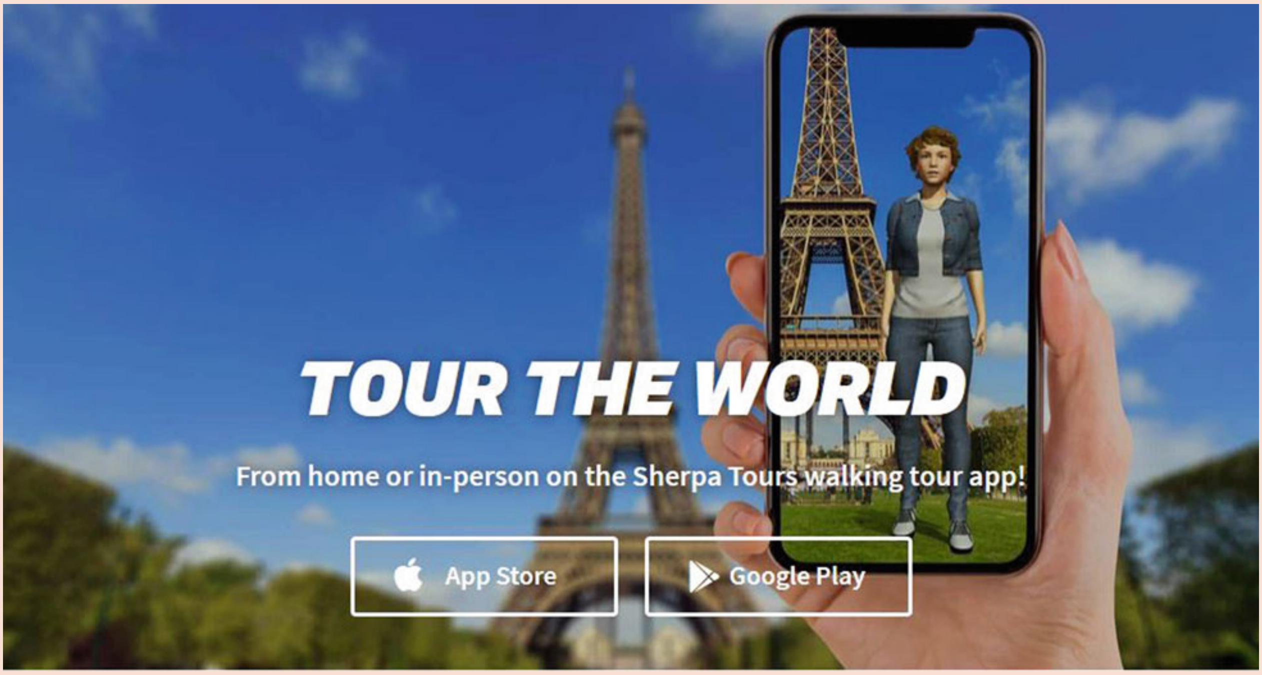


Fig. 19.5 Virtual walking city tours using AR (Source: Sherpa Tours, www.sherpatours.com)

Fig. 19.6 On-site AR to entertain visitors (Source: SevenD, www.sevend.de)

(pre-visit), provides richer information and interpretation, improves learning and enjoyment (on-site), and increases spending and the intention to revisit (post-visit). Overall, it enhances the tourist experience. In acknowledgment of this positive impact, there have been an increasing number of studies that explore the value of AR in the tourism sector. In a recent study among tourism managers, Cranmer, tom Dieck and Fountoulaki [30] explored tourism-specific AR value dimensions and found that AR marketing and sales activities are the most prominent benefits of implementing AR in the tourism sector. Many interviewees identified the potential of AR to better promote facilities, amenities, tours, and destinations and agreed that AR can be used "as a marketing tool to improve [the customers'] understanding of what the destination is really like" ([30], p.5).

In addition to the manager perspective, tourism research has also focused on tourist acceptance (e.g., [14, 19, 65, 73–75]). Tom Dieck and Jung [65], for example, proposed an AR acceptance model in the context of urban heritage tourism and identified seven dimensions that should be incorporated into AR acceptance, i.e., information quality, system quality, cost of use, recommendations, innovativeness, risk, and facilitating conditions. These influence the adoption of AR technology by young tourists. Focusing on the same context, Chung, Han, and Joun [71] defined three factors, i.e., technology readiness, visual appeal, and facilitating conditions, which are crucial to AR perception and influence the AR usage intention and destination visit intention of tourists. Jung, Chung, and Leue [71] used a quality model to investigate tourist satisfaction and the intention to recommend tourist AR services, and they identified content, personalized service, and system quality as important factors. Furthermore, Jung et al. [76] explored the effects of cross-cultural differences on mobile AR adoption in the cultural heritage tourism context, followed by a study by Jung et al. [77], which explored "how long- and short-term orientation moderates the relationship between experience economy provided by AR applications and users' perceived value" ([77], p. 2) and found that South Korean tourists (representing long-term orientation culture) regarded the educational factors of AR applications as highly valuable, whereas Irish tourists (representing short-term orientation culture) highly valued escapist experiences. In general, these studies show that AR in tourism can create enhanced experiences and improve the attitudes and behavioral intentions of tourists [47]. There is limited discussion, however, of how AR applications must be designed in order to add real value to the tourist experience, and studies exploring different needs and evaluations of AR applications among different tourist segments remain scarce.

19.3 The Role of Augmented Reality in Tourism: Research Questions

Given the exploratory nature of this study, we have developed three research questions (RQs). We will address these RQs in the following and answer them by means of an empirical study.

RQ1 addresses the awareness of AR among tourists. According to Roger's [78] Innovation Diffusion Theory, the adoption and use of a new technology typically starts with gaining basic knowledge about the innovation and being exposed to it. The second stage, persuasion, refers to the formation of positive attitudes and beliefs about the innovation in response to the experiences in the first stage. Decision, the third stage, describes behavioral intentions to implement (or reject) the innovation, followed by implementation, i.e., obvious behavior. The fifth stage, confirmation, refers to the decision to continue using or to reject an innovation. As AR is new, most tourists are in the first phase of innovation diffusion. In any case, the adoption of an innovation – in our case AR – usually starts with knowledge. Therefore, we must ask the question:

RQ1: How aware are tourists of AR?

Multiple studies have assessed the drivers behind the acceptance, evaluation, and use of AR in tourism [14, 19, 65, 71–75] and related fields [3, 32, 33, 35]. There is thus a solid body of research to explain why people tend to react positively or negatively to certain AR content. However, there is very little research to explain how consumers rate different use cases in comparison. More specifically, while the existing body of research has mostly focused on drivers (and less on the magnitude of overall acceptance), our study complements prior research by looking at overall evaluations of different use cases (see Table 19.2). Knowing from prior research that different situations lead to different preferences in the AR device (i.e., mobile devices or wearable devices such as AR smart glasses, e.g., [4]), we also study device preferences in the tourism context.

These findings can be of help for AR managers in tourism organizations. The results can also complement prior research by identifying relevant use cases. Thus, we must ask the question:

RQ2: How do tourists evaluate certain touristic AR use cases?

Previous studies have investigated AR acceptance in different contexts and focused on different dimensions of AR acceptance (e.g., [14, 19, 71]). Some studies have provided detailed insights into specific target groups, e.g., young female tourists in tom Dieck and Jung [14]. While these studies

Table 19.2 AR use cases in this study

Virtual guides	
1	Virtual guides in an AR app lead tourists to famous sights in the city. Tourists are guided from one sight to another via AR navigation, similar to Google Maps
2	Virtual guides also lead tourists to sights outside the city and accompany them, for example, on hiking or biking tours. With AR, tourists are navigated to a certain spot and receive information on that site outside the city
3	An AR app presents a "real" tourist guide who is superimposed on the tourist's field of vision in real time. The user can interact with the AR tour guide and ask questions
AR discovery apps	
4	AR apps for historical sites overlay the real world with historical images and thus allow tourists to see what a certain building or district looked like in the past
5	City guides provide tourists with up-to-date information on specific sights. Once tourists arrive at a site, the camera recognizes the environment, and additional information about the site is displayed in AR
6	Based on the principle of Pokémon Go, AR games allow tourists to discover a city or environment in a fun way and motivate users to continue the discovery tour
Additional information in AR	
7	AR apps display additional information to printed brochures. For example, they can show videos in the user's field of view which bring to life a printed text or picture
8	AR review apps, for example, "write" virtual reviews on restaurant walls so that users can see reviews immediately when looking through the screen and passing by a restaurant
9	AR discount apps display vouchers for certain shops and attractions into the tourist's view when tourists approach such places

provide rich and broad overviews of tourists – or detailed insights into specific target groups – there is less research that starts with the heterogeneity of tourists. This is surprising, especially since different tourists may react differently to AR.

Understanding this heterogeneity is highly relevant for managers. Many companies apply an STP approach (segmentation, targeting, and positioning). This starts with the identification of consumer segments, i.e., a broad market normally consisting of existing and potential tourists is divided into subgroups based on shared characteristics. Many firms use demographics (e.g., four groups: male/female vs. young/old). If available, behavioral variables (e.g., analyzing log files) can also serve as segmentation criteria. Other approaches use psychographic variables such as tourism needs or expectations for segmentation. Especially if there are more than just two variables, managers need to apply multivariate techniques – such as cluster analysis – to identify these segments. Once segments are identified, managers rate each segment and chose one or more promising segments. This is called targeting. This allows them to position their businesses and/or activities.

However, little is known about how different tourism segments respond to AR. This study therefore starts with a basic approach and identifies segments based on tourism motivation (i.e., what tourists expect during travel) using cluster analysis. We then show how the evaluation of use cases diverges between these – yet unknown – segments. We therefore must ask the question:

RQ3: How does the evaluation of touristic AR use cases differ between tourist segments?

19.4 Study

19.4.1 Research Design and Methodology

This study surveyed 553 city tourists in Germany with the help of a professional market research firm. The sample included a broad age distribution and covered city tourists from all over Germany. To identify our target group, we specifically asked respondents about their attitude towards choosing a city as their tourism destination. Anyone who did not meet this qualification criterion was automatically excluded. Respondents received financial compensation for their participation. Table 19.3 summarizes the sample in more detail.

Respondents rated various tourism needs with regard to their personal importance on a scale of 1 (absolutely unimportant) to 5 (very important). We presented them a list consisting of 14 criteria including wellness and relaxation; luxury and pampering; culinary experiences; rest and relaxation; sports and exercise, hiking, nature, and outdoor activities; cycling; experience and adventure; family friendliness; events and activities; shopping; culture, theatre, and music; historic sites; regional tradition and original landscapes. These scores were averaged and aggregated to three scores (pleasure-

Table 19.3 Demographics of the sample

Age in years		**Gender**	
Min	18	Male	52%
Max	83	Female	48%
Mean	44		
Profession			
Employed (full- or part-time)		59%	
Self-employed		7%	
Students/pupils/in training		8%	
Housewives/house Husbands/parental leave		6%	
Retired		15%	
Jobseekers		5%	
Education			
No degree (yet)		1%	
Certificate lower than high school		44%	
High school diploma		24%	
University degree		31%	
Income (per month)			
Less than 1,000 EUR		12%	
1,000–1,999		21%	
2,000–2,999		25%	
3,000–3,999		19%	
4,000–4,500		6%	
More than 4,500		9%	
Population of the respondents' city			
Fewer than 5,000		17%	
5,001–50,000		35%	
50,001–500,000		26%	
More than 500,000		22%	

seeking (alpha = .685), activity-seeking (alpha = .730), culture-seeking (alpha = .715)), which served as cluster variables for further analyses.

Respondents then received a definition of AR including pictures to ensure they all understand what AR is. Respondents were asked to indicate their level of AR experience by choosing the answer that best fit them: *I consider myself as pretty knowledgeable and I have actual usage experience*; *I consider myself as pretty knowledgeable, but I do not have actual usage experience so far*; *I have only heard about AR, but don't really know what it is*; *I don't know what AR is*.

We then measured the evaluation of various AR use cases (as discussed in the theory section and in Table 19.2) in tourism on five-point Likert scales [How useful do you find the following augmented reality (AR) apps in the tourism sector?]. Higher values indicate a more positive evaluation. Finally, since AR is dependent on devices (i.e., it is used on a smartphone, tablet, or glasses), we also asked how respondents would prefer to use AR (see Fig. 19.10).

19.5 Results

RQ1 asks how aware tourists are of AR. To answer this question, we asked respondents to choose the response that best fits. Only one tenth of all surveyed tourists had knowledge and actual AR user experience. One third of respondents said they had knowledge of AR but no actual experience with it. In sum, almost 44% had good knowledge of AR. On the other hand, a third of those surveyed had heard about AR but did not really know what it was, and almost one fourth did not know at all what AR was. This indicates that a broad mass adoption of AR among tourists has not been achieved yet. Figure 19.7 summarizes the results. In addition, readers may find it interesting to explore AR knowledge between different demographic target groups. Table 19.4 therefore shows the means and standard deviations of the construct for the whole sample and selected sub-samples. In general, these analyses show that males and younger tourists tend to be more familiar with AR. However, tourists from more rural areas tend to be more familiar with AR than urban respondents.

In order to answer RQ2, we presented the nine touristic use cases from Table 19.2 to respondents and measured their acceptance. Higher values indicate a better evaluation. The results are presented in Fig. 19.8. We are also interested how and with what device tourists preferably use AR.

It is important to note that – across the sample – the averages are around the midpoint scale (3). This indicates that acceptance is still fairly low. However, we found that more than two out of three respondents (67.5%) considered at least one of the proposed use cases to be (very) useful (scale values of 4 or 5). On average, respondents actually rated three to four of the nine use cases with 4 or higher ($M = 3.51$; $SD = 3.12$). The top 2 values, i.e., the percentage of respondents who rated a use case with four or higher (shown in Fig. 19.8), indicate which of the use cases are considered especially useful. Almost half of the respondents indicate that virtual guides that lead tourists to sights in the city (49.6%) or sights outside the city (46.8%), history apps (48.7%), and virtual city guides (50.6%) are (very) useful.

Moreover, we found that the more experienced tourists are with AR, the more positive they rate the proposed use cases across all cases (see Fig. 19.9). Thus, as familiarity with AR increases, so do evaluations of AR use cases.

When it comes to the AR device, the majority of respondents prefer an app on their own smartphone or tablet, followed by almost one fourth preferring to use AR through an internet browser on their mobile/smartphone. The use of AR apps on rented mobile devices or on borrowed smart glasses (data glasses) is less popular among German tourists. For detailed results, see Fig. 19.10.

RQ3 asks how the evaluation of touristic AR use cases differs between different tourist segments. We used a two-step

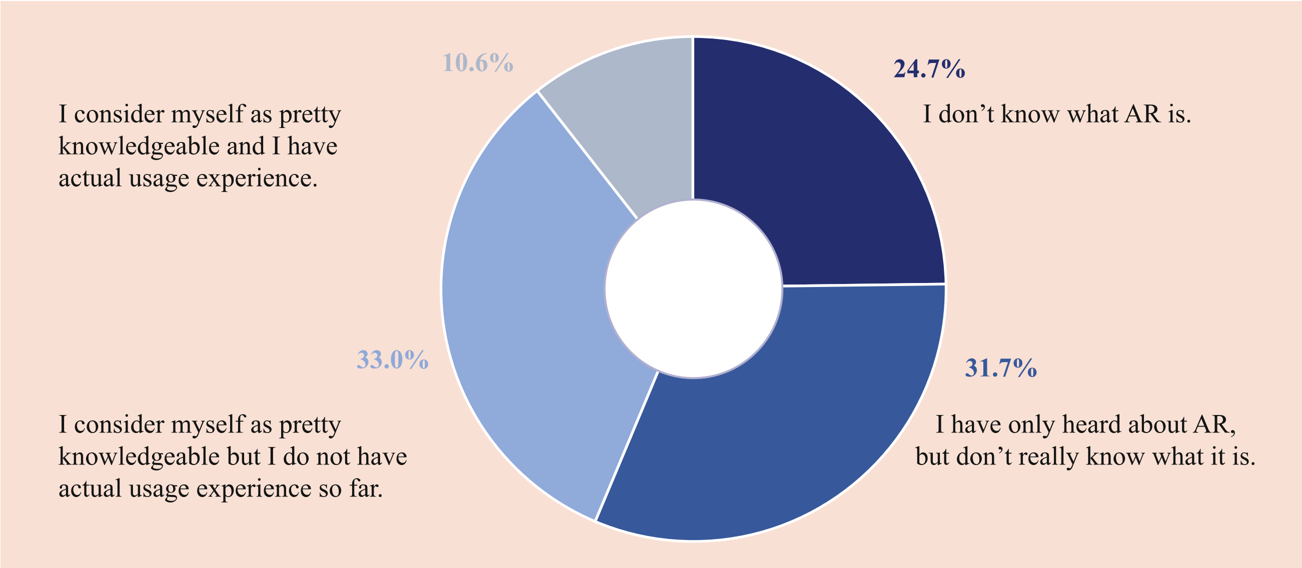


Fig. 19.7 AR experience among tourists (Germany). Note: Question Text: "Are you familiar with Augmented Reality (AR)? Please select the answer that fits best." For a reference in the US, see eyecandylab [79]

Table 19.4 Additional statistics on differences in AR knowledge regarding demographics in percentage

	Overall sample	Group comparisons								
		Gender[1]		Age[2]			Population[3]			
		M	*F*	<30	31–45	>45	<5,000	5,000–50,000	50,000–500,000	>500,000
I consider myself as pretty knowledgeable, and I have actual usage experience	10.6	12.8	8.2	16.9	13.4	5.4	14.6	9.7	12.0	7.3
I consider myself as pretty knowledgeable, but I do not have actual usage experience so far	33.0	39.4	26.2	41.0	37.5	25.6	22.5	29.7	36.1	42.3
I have only heard about AR but don't really know what it is	31.7	29.6	34.0	26.5	25.9	39.5	31.5	29.7	36.8	29.3
I don't know what AR is	24.7	18.2	31.6	15.7	23.2	29.6	31.5	30.8	15.0	21.1

Note: *F* female, *M* male, *SD* standard deviation; (1) Chi^2-Test: significant ($p = .000$); (2) significant ($p = .000$) based on an F-Test; (3) Chi^2-Test: partially significant ($p = .009$)

approach to answer this question. Inspired by prior academic and managerial research in tourism [65, 76], we first identified a set of items which reflects three groups of variables that cover tourist needs and interests: pleasure-seeking (wellness and relaxation; luxury and pampering; culinary experiences; rest and relaxation; alpha = .685), activity-seeking (sports and exercise, hiking, nature, and outdoor activities; cycling; experience and adventure; family friendliness; alpha = .730) and culture-seeking (events and activities; shopping; culture, theatre, music; historic sites; regional tradition and original landscapes; alpha = .715). To answer RQ3, we applied a hierarchical cluster analysis. We used mean scores for each construct as cluster variables. Since some respondents tend to give all questions a higher (or lower) value than others, we applied a 0 to 1 standardization by case to parcel out these variations. We estimated the distances matrix using a Squared Euclidean Distance, and applied the Ward's Method as clustering algorithm. This method aims at minimizing

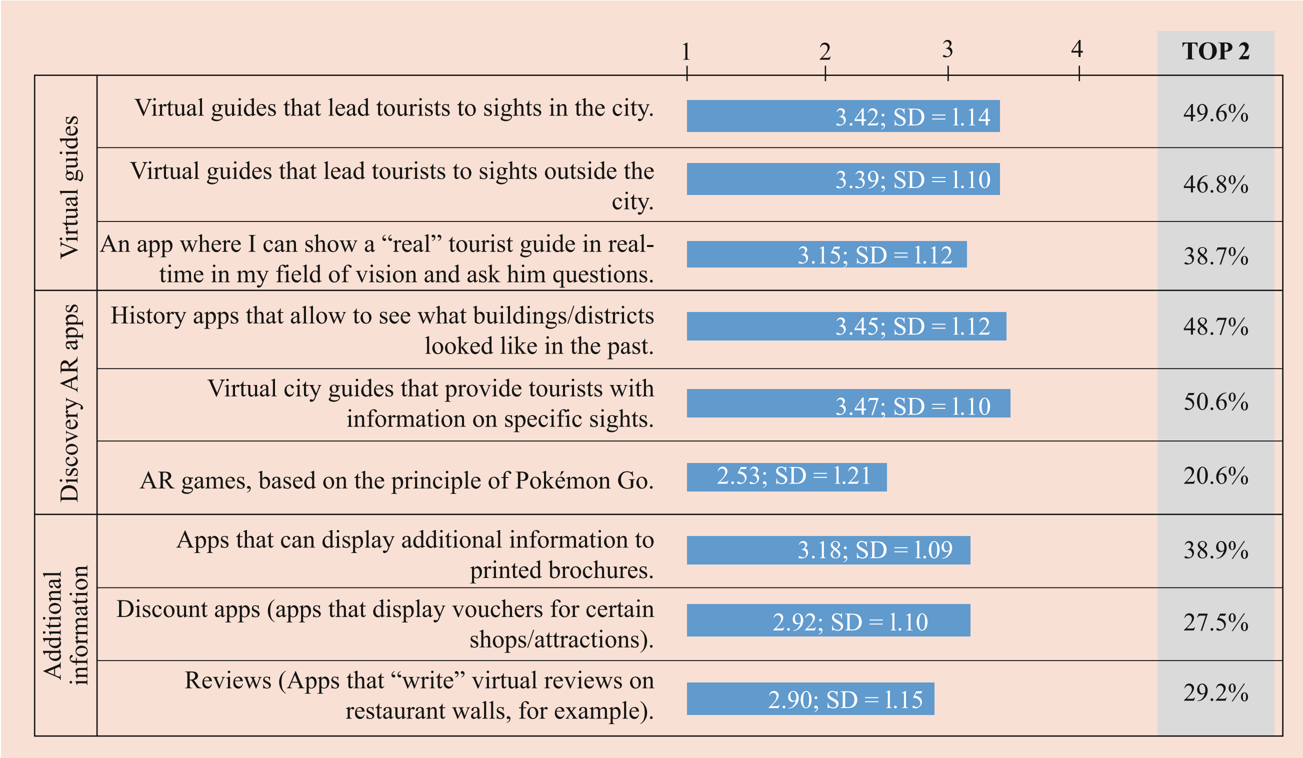


Fig. 19.8 Evaluation of tourist AR use cases among German tourists (base: all respondents). Note: Question Text: "How useful do you find the following Augmented Reality (AR) services in the tourism sector?" (Scale ranges from 1 (not useful at all) to 5 (very useful)); $n = 553$; top 2 values: Percentage of respondents who rated a use case with 4 or higher on a scale from 1 to 5

the variance within each cluster. The strength of the Ward's Method is that it produces solutions with very low intra-cluster and very high inter-cluster heterogeneity. An inspection of the dendrogram helped us to find the ideal number of segments.

Before we discuss how these clusters differ in terms of the evaluation of AR use cases, we will first present each cluster in more detail (see Table 19.11). ANOVAs clearly show that the three clusters differ significantly for each variable (all $p < .004$). Cluster 1 ($n = 154$; 29.0%) is characterized by higher levels of activity ($M = 3.55$, SD = .72) and culture ($M = 3.41$; SD = .79) but lower levels of pleasure ($M = 3.07$; SD = .81). We named this segment *Explorers*. We named the second cluster ($n = 160$; 30.2%) *Hedonists* as this segment shows the highest interest in pleasure-seeking activities ($M = 3.71$, SD = .59), with all other interests (activity-seeking ($M = 3.09$, SD = .70) and culture-seeking ($M = 2.94$, SD = .61)) being considerably lower. The last cluster ($n = 216$; 40.8%) was named *Culture gourmets* as individuals in this segment showed the highest interest in pleasure-seeking ($M = 3.37$, SD = .65) as well as culture-seeking activities ($M = 3.63$, SD = .56), with interest in activity ($M = 2.75$, SD = .73) being considerably lower. To assess whether the differences between the clusters are significant, we applied Bonferroni post-hoc tests to compare the three values between clusters and within each cluster. All differences were highly significant (all $p < .004$).

To better understand "who" these segments are and how marketers could target them, we also analyzed demographic data. As shown in Table 19.5, the three segments are surprisingly similar in terms of demographics. Only small trends are observable. For example, hedonists (cluster 2) tend to have a slightly higher proportion of men, and culture gourmets (cluster 3) tend to have a slightly higher average age.

Having identified and described the clusters, we will now focus on answering RQ3. To do this, we used ANOVAs to compare the evaluations of the nine uses across the three segments. In situations where we found significant differences, we inspected them in more detail using Bonferroni post-hoc tests (see Table 19.6). The results are shown in Fig. 19.11.

Cluster one tourists (*Explorers*) show the greatest interest in virtual guides, apps that can display additional information to printed brochures, discount apps, and AR games.

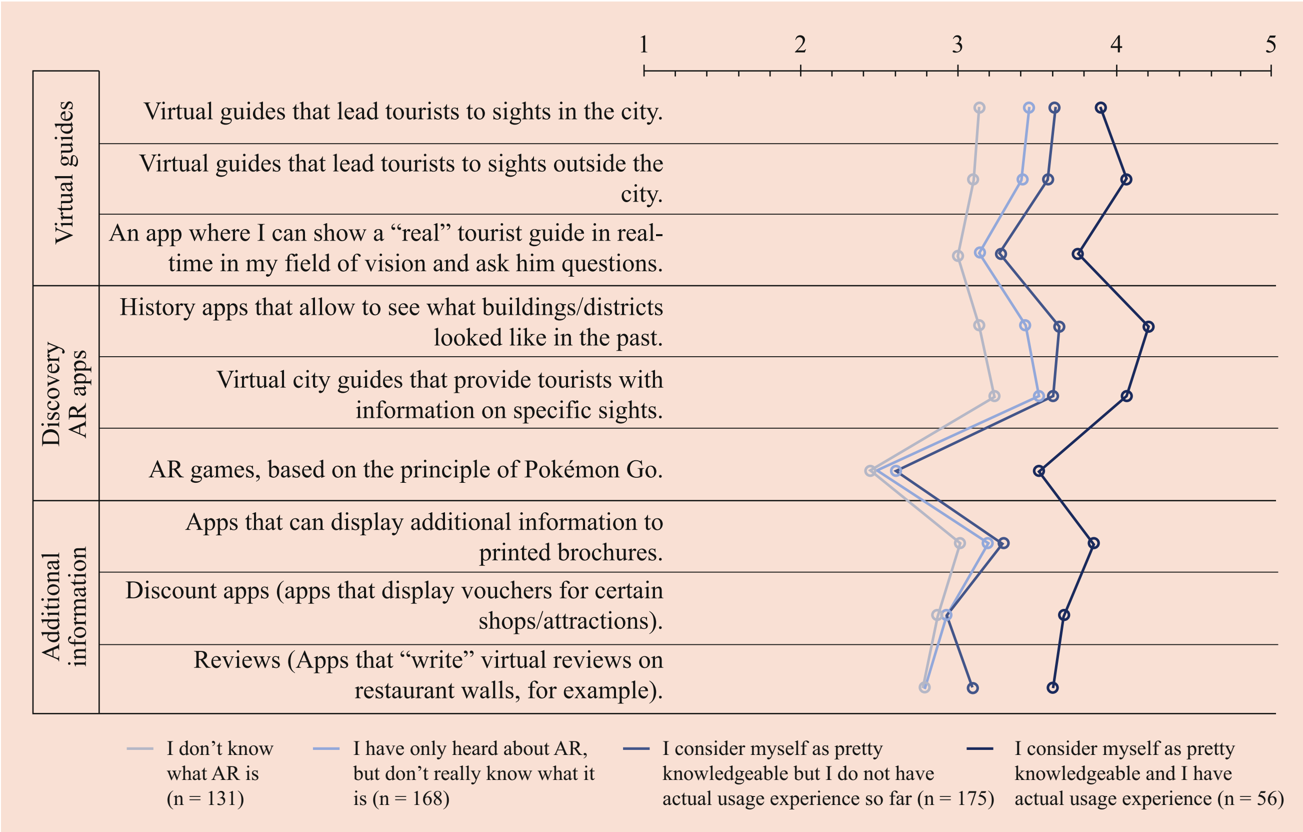


Fig. 19.9 Evaluation of use cases depending on AR experience

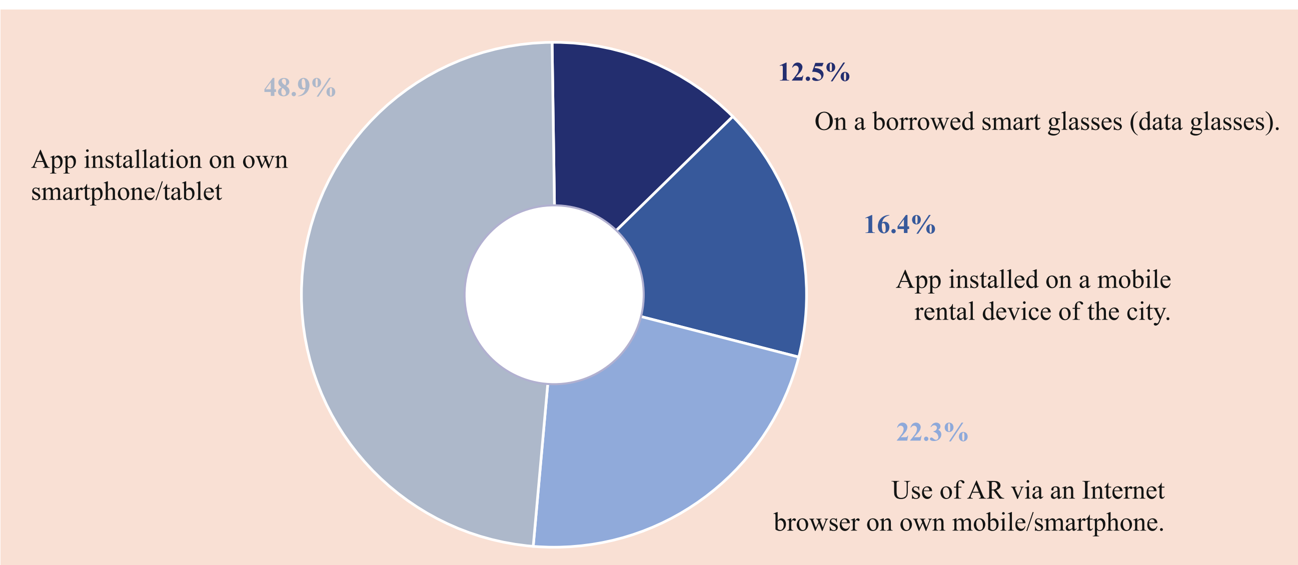


Fig. 19.10 Usage preference of AR use cases among tourists. Note: Question Text: "How would you prefer to use AR? Please select the answer that fits the most." $n = 553$

Cluster two tourists (*Hedonists*) tend to show the lowest overall interest in any AR use case compared to the other two clusters. As it is primarily focused on relaxing and seeking pleasure, this segment has the lowest interest in any AR use case. This segment has the youngest average age (42.2) and the largest proportion of men (55.5%).

Cluster three tourists (*Culture gourmets*) are interested in virtual city guides, history apps, virtual guides that lead

Table 19.5 Overview of identified clusters

		Cluster 1 *Explorers* ($n = 154$)	Cluster 2 *Hedonists* ($n = 160$)	Cluster 3 *Culture gourmets* ($n = 216$)	**Total** ($n = 530$)
Segmentation variables					
Pleasure-seeking	*M* SD	3.07 .82	3.71 .59	3.37 .65	3.39 .73
Activity-seeking	*M* SD	3.55 .73	3.09 .70	2.75 .73	3.08 .73
Culture-seeking	*M* SD	3.41 .79	2.94 .61	3.63 .56	3.36 .71
Demographics age (average)[1]					
	M	42.4	42.2	44.9	43.4
Gender[2]					
Male	%	50.6	54.4	50.5	51.7
Female	%	49.4	45.6	49.5	48.3
Population[3]					
Fewer than 5,000	%	15.6	18.8	16.2	16.8
5,001–50,000	%	39.0	33.7	32.9	34.9
50,001–500,000	%	24.6	23.1	26.8	25.1
More than 500,000	%	20.8	24.4	24.1	23.2

Note: (1) Partially significant ($p = .098$) based on an F-Test; (2) Chi^2-Test: not significant ($p = .719$); (3) Chi^2-Test: not significant ($p = .849$)

tourists in the city as well as outside the city, and apps that can display information not found in printed brochures. Overall, this cluster shows the greatest overall interest in AR use cases, which makes this segment particularly interesting to marketers.

Furthermore, we found that explorers and culture gourmets have similar needs. Their evaluation of uses only differs significantly in the evaluation of AR games ($p = .002$). The comparison between explorers and hedonists is slightly different. These groups rate the usefulness of AR games ($p = .001$), discount apps ($p = .013$), and apps that display additional information to brochures ($p = .001$) differently. The largest difference in AR use case evaluation can be found between hedonists and cultural gourmets.

We also investigated whether differences in the evaluation of use cases depend on gender and age. While there is a significant difference in the evaluation of discount apps (male: $M = 2.81$; SD $= 1.15$; female: $M = 3.04$; SD $= 1.03$; $p = .013$), there is no significant difference for gender. Similar results can be found for age. There are no significant differences in the evaluation of use cases for most of the use cases ($p > .05$). There are significant differences between different age groups only for the ratings for discount apps ($p = .000$), AR games ($p = .000$), and review apps ($p = .001$).

19.6 Discussion

As our results show, virtual guides that lead tourists to sights in the city as well as outside the city (e.g., hiking tours), history AR apps, and virtual guides that provide tourists with information on specific sights are among the most popular touristic AR use cases. We also found that different tourist segments (explorers, hedonists, and culture gourmets) value the usefulness of tourist AR use cases differently, whereas almost no significant differences can be found when it comes to demographic data such as age and gender. It is thus important to adapt touristic AR use cases to the motivations and needs of tourists (and not so much to demographic data, as many organizations still do). Explorers and culture gourmets, for example, evaluate AR use cases in a similar fashion, whereas hedonists and culture gourmets are interested in different AR use cases. Overall, we found that culture gourmets and explorers are the most positive about tourist AR use cases. Thus, tourism managers should focus on those two segments when implementing AR use cases in the tourism sector. We also found that AR is still new to the majority of tourists. With increasing knowledge and exposure to AR, tourists rate AR use cases more positively. As we discussed, the AR market is still in an early stage. We thus expect a growing interest and demand in tourist AR use cases in the coming years. According to our findings, this may translate into more positive evaluations.

Regarding the AR device, we found that the majority of respondents prefer an app on their own smartphone or tablet. The use of AR apps on borrowed AR-specific smart glasses is less popular among German tourists so far. However, once the hardware will be further developed and AR glasses will become mainstream, we expect AR glasses to replace the smartphone when experiencing touristic AR use cases.

Table 19.6 AR evaluation by cluster

		Explorers ($N = 154$)	**Hedonists** ($N = 160$)	**Culture gourmets** ($N = 216$)	**Total** ($N = 530$)			
AR knowledge[1]								
I consider myself as pretty knowledgeable, and I have actual usage experience	%	11.7	11.3	9.3	10.6			
I consider myself as pretty knowledgeable, but I do not have actual usage experience so far	%	29.2	34.4	34.7	33.0			
I have only heard about AR but don't really know what it is	%	30.5	28.7	34.7	31.7			
I don't know what AR is	%	28.6	25.6	21.3	24.7			
Use of AR[2]								
App installation on own smartphone/tablet	%	44.8	50.0	50.9	48.9			
Use of AR via an Internet browser on own mobile/smartphone	%	24.7	22.5	20.4	22.3			
App installed on a mobile rental device of the city	%	14.9	16.3	17.6	16.4			
On borrowed smart glasses (data glasses)	%	15.6	11.3	11.1	12.5			
						Bonferroni Post-hoc (p-values)[3]		
Evaluation of AR use cases						E< >H	E< >C	H< >C
Virtual guides that lead tourists to sights in the city	*M* SD	3.40 1.03	3.23 1.22	3.57 1.13	3.42 1.14	n.s.	n.s.	.011
Virtual guides that lead tourists to sights outside the city (e.g., hiking tours)	*M* SD	3.47 .98	3.22 1.15	3.47 1.12	3.39 1.10	n.s.	n.s.	.088
An app where I can show a "real" tourist guide in real time in my field of vision and ask him questions	*M* SD	3.25 1.01	3.03 1.15	3.17 1.15	3.15 1.12	n.s.	n.s.	n.s.
History apps that allow to see what buildings/districts looked like in the past	*M* SD	3.46 .98	3.21 1.18	3.63 1.14	3.45 1.12	n.s.	n.s.	.001
Virtual city guides that provide tourists with information on specific sites	*M* SD	3.46 .99	3.26 1.18	3.63 1.10	3.47 1.10	n.s.	n.s.	.004
AR games, based on the principle of Pokémon Go	*M* SD	2.85 1.17	2.37 1.22	2.42 1.19	2.53 1.21	.001	.002	n.s.
Apps that can display additional information to printed brochures	*M* SD	3.35 .97	2.91 1.12	3.26 1.12	3.18 1.10	.001	n.s.	.006
Discount apps (apps that display vouchers for certain shops/attractions)	*M* SD	3.10 .10	2.74 1.16	2.93 1.11	2.92 1.10	.013	n.s.	n.s.
Reviews (apps that "write" virtual reviews on restaurant walls, for example)	*M* SD	3.03 1.04	2.86 1.18	2.83 1.19	2.90 1.15	n.s.	n.s.	n.s.

Note: (1) Chi^2-Test: not significant ($p = .585$); (2) Chi^2-Test: not significant ($p = .730$); (3) Reading example: "Virtual guides that lead tourists to sights in the city': no significant differences in the evaluation of this use case between explorers and hedonists, no significant differences in the evaluation of this use case between explorers and culture gourmets, significant differences in the evaluation of this use case between hedonists and culture gourmets ($p = .011$); n.s.: $p > 0.10$

19.6.1 General Conclusion

AR provides numerous advantages for tourism organizations, and the interest of tourists will likely increase once AR develops into a more mainstream technology. This study sheds light on specific touristic AR use cases. A core finding is that different tourist segments prefer different use cases. Managers and scholars alike should keep this heterogeneity in mind when conducting and interpreting research.

In order to provide concrete insights on AR adoption strategies for tourism managers, we established key learnings and working questions that can help to assess the appropriateness and approach of AR adoption in their strategic setting. Based on the study results, our conclusions for tourism managers are as follows:

Be Aware that Consumers are Not Aware

As our results indicate, tourists are still in an early stage of adopting AR in tourism. When looking at Roger's [78] innovation diffusion theory, we can see that consumers are in the first phase of gaining basic knowledge about this technology. Consequently, marketers should not overstress consumers with overly complicated AR use cases but should rather focus on the initial phases and emphasize basic knowledge and

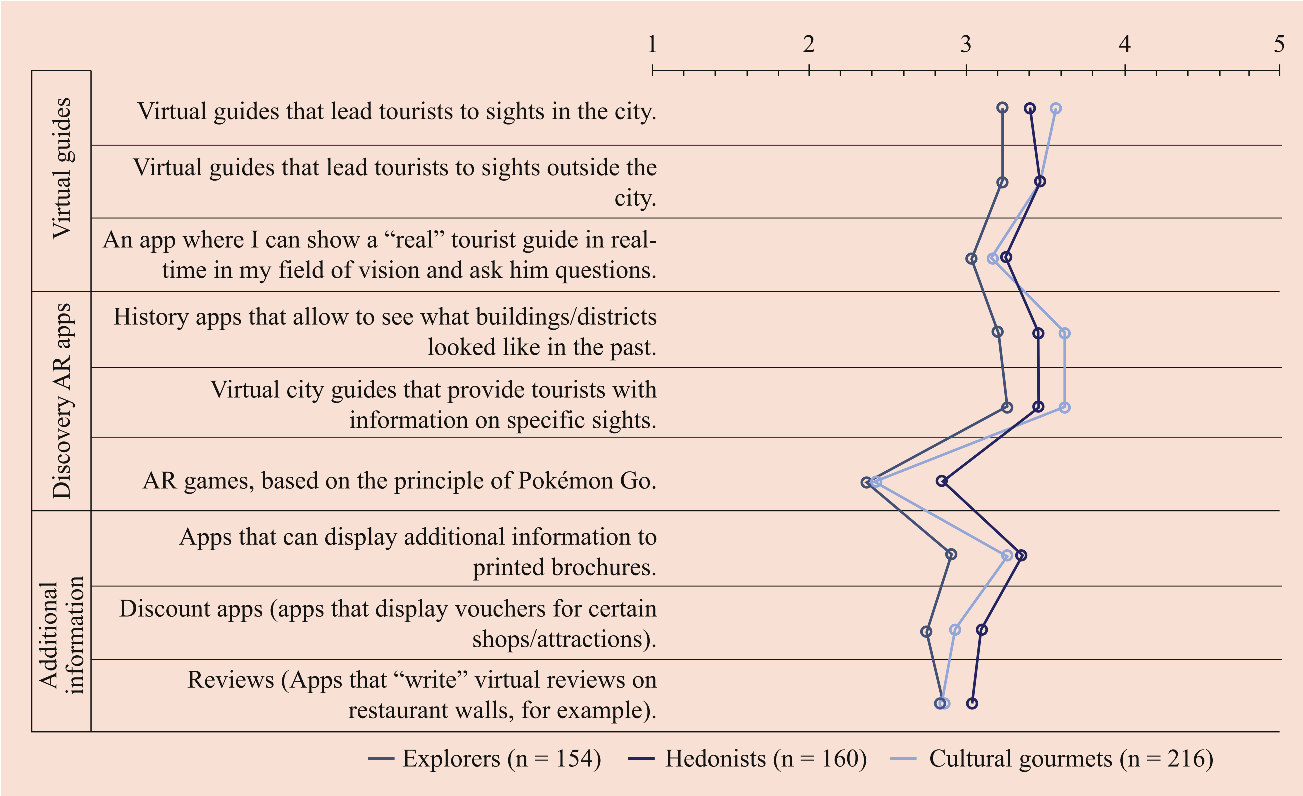


Fig. 19.11 Evaluation of touristic AR use cases between different tourist segments

persuasion. It is also important to note that all the marketing activities of tech companies and the announced devices will likely push the average familiarity with AR across segments. Thus, once familiarity with AR has reached higher levels, the reactions of consumers (and this includes tourists) will become more positive.

Work question: How can we push consumers from gaining knowledge to the next stage of persuasion and how can we optimize persuasion?

AR is Just Getting Started

Every week, there are new technological advances, startup ideas, and new possibilities in the field of AR which allow for new and promising AR use cases in tourism. Ideas that appeared impossible are now available and are downloadable with a single click. Although there are use cases, as we have found, that are already interesting to tourists, these use cases only represent the first of many to come. New AR use cases may increase convenience for tourists, may help tourists to save time and money, may help to personalize individual travel, and thus may optimize the tourism experience. AR use has just started.

Work question: First, consider a goal a tourist may have when maximizing his or her travel experience and, second, establish a theoretical use case in which AR achieves this goal.

Focus on Individual Tourist Needs

In our study, we determined which tourist segments exist when it comes to city traveling and we checked how their attitudes towards different AR use cases differ. We found that age and gender appear to play a minor role, and psychographic properties tend to play a more salient role. Our analysis revealed three relevant tourism segments: The *Explorers*, the *Hedonists*, and the *Culture gourmets*. While these segments showed few differences in terms of demographics, they showed strong differences when it came to AR use cases. Surprisingly, the hedonists, who had a slightly lower average age and a slightly higher proportion of men, showed the lowest overall interest in AR use cases. The explorer segment showed a slightly higher interest in more specific AR use cases. Culture gourmets, on the other hand, showed the highest overall level of interest in AR use cases. Our study reveals that consumers not only differ in terms of general traveling interests but that these psychographic differences also play an important role when it comes to their interest in AR use cases.

Work question: Try to get into the head of a person in each of the three segments, and brainstorm specific AR use cases for a typical member of this segment (goals, specific content, devices, requirements about usability, and so forth).

Finally, it is important to target each of the segments using communication strategies. Each segment has different preferences and expectations. Tourists may use different media and may prefer different communication styles.

Work question: How can we target each segment? What do effective communication messages look like? How do we create interest in AR apps?

References

1. Rauschnabel, P.A., Felix, R., Hinsch, C.: Augmented reality marketing: how mobile AR-apps can improve brands through inspiration. J. Retail. Consum. Serv. **49**, 43–53 (2019)
2. Rauschnabel, P.A., Felix, R., Hinsch, C., Shahab, H., Alt, F.: What is XR? Towards a framework for augmented and virtual reality. Comput. Hum. Behav. **133**, 107289 (2022)
3. Rauschnabel, P.A., Rossmann, A., tom Dieck, M.C.: An adoption framework for mobile augmented reality games: the case of Pokémon Go. Comput. Hum. Behav. **76**, 276–286 (2017)
4. Rauschnabel, P.A.: Virtually enhancing the real world with holograms: an exploration of expected gratifications of using augmented reality smart glasses. Psychol. Mark. **35**(8), 557–572 (2018)
5. Herz, M., Rahe, V.: Virtual reality becoming part of our lives—assessing consumers' perceived applicability of virtual reality supported tasks and a critical reflection on the development. In: Augmented Reality and Virtual Reality, pp. 113–122. Springer, Cham (2020)
6. Wang, D., Xiang, Z., Fesenmaier, D.R.: Smartphone use in everyday life and travel. J. Travel Res. **55**(1), 52–63 (2016)
7. BCG. Augmented reality: is the camera the next big thing in advertising? Retrieved June 3, 2020, from https://www.bcg.com/de-de/publications/2018/augmented-reality-is-camera-next-big-thing-advertising.aspx (2018)
8. Deloitte. Snap consumer AR. Global report 2021. Retrieved Mai 22, 2021, from https://downloads.ctfassets.net/inb32lme5009/hSj5P9vzaox79vlTqXZfj/d6094404212f1327e8b7d69535337941/Snap_Consumer_AR_Research_GlobalReport.pdf (2021)
9. PWC. Seeing is believing. How virtual reality and augmented reality are transforming business and the economy. Retrieved April 16, 2020, from https://image.s50.sfmc-content.com/lib/fe31117075640475701c74/m/1/9b4ff71e-c4f0-4746-bada-ccf96e2018de.pdf (2019)
10. Nayyar, A., Mahapatra, B., Le, D., Suseendran, G.: Virtual Reality (VR) & Augmented Reality (AR) technologies for tourism and hospitality industry. Int. J. Eng. Technol. **7**(2.21), 156–160 (2018)
11. Sigala, M.: New technologies in tourism: from multi-disciplinary to anti-disciplinary advances and trajectories. Tour. Manag. Perspect. **25**, 151–155 (2018)
12. Wang, X., Love, P.E., Kim, M.J., Wang, W.: Mutual awareness in collaborative design: an augmented reality integrated telepresence system. Comput. Ind. **65**(2), 314–324 (2014)
13. Han, D.I., Jung, T., Gibson, A.: Dublin AR: implementing augmented reality in tourism. In: Information and Communication Technologies in Tourism 2014, pp. 511–523. Springer, Cham (2013)
14. tom Dieck, M.C., Jung, T.H.: A theoretical model of mobile augmented reality acceptance in urban heritage tourism. Curr. Issue Tour., 1–21 (2015)
15. Bogomolov. Top 5 ideas how to use Ar in tourism. In Hospitalitynet. Retrieved June 3, 2020, from https://www.hospitalitynet.org/opinion/4092421.html (2019)
16. Gibson, A.: Augmented reality and virtual reality: new realities in digital marketing. Hotel Yearbook 2017. Foresight and Innovation in the Global Hotel Industry. Special Issue on Digital Marketing, 32–33 (2017)
17. Raghavan, L., Dev, C. S.: Digital branding for hotels: the four E's to success. Hotel Yearbook 2020. Foresight and Innovation in the Global Hotel Industry. Special Issue on Digital Marketing, 18–20. (2020)
18. Sharma, S.: How augmented reality is revolutionizing hospitality industry (From Inside Out). In AR Post. Retrieved June 3, 2020, from https://arpost.co/2019/10/11/augmented-reality-revolutionizing-hospitality-industry/ (2019)
19. Jung, T., Chung, N., Leue, M.C.: The determinants of recommendations to use augmented reality technologies: the case of a Korean theme park. Tour. Manag. **49**, 75–86 (2015)
20. tom Dieck, M.C., Jung, T., Han, D.I.: Mapping requirements for the wearable smart glasses augmented reality museum application. J. Hosp. Tour. Technol. **7**(3) (2016). https://doi.org/10.1108/JHTT-09-2015-0036
21. Kotler, P., Keller, K.L., Opresnik, M.O.: Marketing Management. Konzepte, Instrumente, Unternehmenfallstudien, 15th edn. Pearson Deutschland GmbH, Hallbergmoos (2017)
22. Sisodia, R., Wolfe, D., Sheth, J.N.: Firms of Endearment: How World-Class Companies Profit from Passion and Purpose. Pearson Prentice Hall, Upper Saddle River (2003)
23. McCarthy, E.J., Perreault, W.D.: Basic Marketing: A Global-Managerial Approach, 14th edn. McGraw-Hill/Irwin, Home-wood (2002)
24. Herz, M.F.: The country-of-origin concept reassessed–The long path from the 'made-in 'label. In: Impulse für die Markenpraxis und Markenforschung, pp. 141–161. Springer Gabler, Wiesbaden (2013)
25. Herz, M., Arnegger, J.: Country image formation, stability, and change: a longitudinal assessment of consumers' image of Azerbaijan. J. Travel Tour. Mark. **34**(9), 1169–1183 (2017)
26. Herz, M., Arnegger, J., Mayer, M.: How stable is event-related country-image change? In: Forum Markenforschung 2016, pp. 141–157. Springer Gabler, Wiesbaden (2018)
27. Kostin, K.B.: Foresight of the global digital trends. Strateg. Manag. **23**(1), 11–19 (2018)
28. Park, M., Yoo, J.: Effects of perceived interactivity of augmented reality on consumer responses: a mental imagery perspective. J. Retail. Consum. Serv. **52**, 101912 (2020)
29. Rauschnabel, P.A., Babin, B.J., tom Dieck, M.C., Krey, N., Jung, T.: What is augmented reality marketing? Its definition, complexity, and future. J. Bus. Res. **142**, 1140–1150 (2022)
30. Cranmer, E.E., tom Dieck, M.C., Fountoulaki, P.: Exploring the value of augmented reality for tourism. Tour. Manag. Perspect. **35**, 100672 (2020)
31. Shopify. How these retailers use augmented reality to enhance the customer experience. Retrieved June 16, 2020, from https://www.shopify.com/retail/how-these-retailers-are-using-augmented-reality-to-enhance-the-customer-experience (2018)
32. Rauschnabel, P.A., He, J., Ro, Y.K.: Antecedents to the adoption of augmented reality smart glasses: a closer look at privacy risks. J. Bus. Res. **92**, 374–384 (2018)
33. Spreer, P., Kallweit, K.: Augmented reality in retail: assessing the acceptance and potential for multimedia product presentation at the PoS. SOP Trans. Mark. Res. **1**(1), 20–35 (2014)

34. Yim, M.Y.C., Chu, S.C., Sauer, P.L.: Is augmented reality technology an effective tool for e-commerce? An interactivity and vividness perspective. J. Interact. Mark. **39**, 89–103 (2017)
35. Yim, M.Y.C., Park, S.Y.: "I am not satisfied with my body, so I like augmented reality (AR)": consumer responses to AR-based product presentations. J. Bus. Res. **100**, 581–589 (2019)
36. Hinsch, C., Felix, R., Rauschnabel, P.A.: Nostalgia beats the wow-effect: inspiration, awe and meaningful associations in augmented reality marketing. J. Retail. Consum. Serv. **53**, 101987 (2020)
37. Jessen, A., Hilken, T., Chylinski, M., Mahr, D., Heller, J., Keeling, D.I., de Ruyter, K.: The playground effect: how augmented reality drives creative customer engagement. J. Bus. Res. **116**, 85–98 (2020)
38. Scholz, J., Duffy, K.: We ARe at home: how augmented reality reshapes mobile marketing and consumer-brand relationships. J. Retail. Consum. Serv. **44**, 11–23 (2018)
39. Daassi, M., Debbabi, S.: Intention to reuse AR-based apps: the combined role of the sense of immersion, product presence and perceived realism. Inf. Manag. **58**(4), 103453 (2021)
40. Javornik, A.: 'It's an illusion, but it looks real!' consumer affective, cognitive and behavioural responses to augmented reality applications. J. Mark. Manag. **32**(9–10), 987–1011 (2016)
41. Poushneh, A.: Augmented reality in retail: a trade-off between user's control of access to personal information and augmentation quality. J. Retail. Consum. Serv. **41**, 169–176 (2018)
42. Poushneh, A., Vazquez-Parraga, A.Z.: Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. J. Retail. Consum. Serv. **34**, 229–234 (2017)
43. Smink, A.R., Van Reijmersdal, E.A., Van Noort, G., Neijens, P.C.: Shopping in augmented reality: the effects of spatial presence, personalization and intrusiveness on app and brand responses. J. Bus. Res. **118**, 474–485 (2020)
44. Herz, M., Rauschnabel, P.A.: Understanding the diffusion of virtual reality glasses: the role of media, fashion and technology. Technol. Forecast. Soc. Chang. **138**, 228–242 (2019)
45. Yavuz, M., Çorbacıoğlu, E., Başoğlu, A.N., Daim, T.U., Shaygan, A.: Augmented reality technology adoption: case of a mobile application in Turkey. Technol. Soc. **66**, 101598 (2021)
46. Schein, K.E., Rauschnabel, P.A.: Augmented reality in manufacturing: exploring workers' perceptions of barriers. IEEE Transactions on Engineering Management (2021)
47. He, Z., Wu, L., Li, X.R.: When art meets tech: the role of augmented reality in enhancing museum experiences and purchase intentions. Tour. Manag. **68**, 127–139 (2018)
48. Heller, J., Chylinski, M., de Ruyter, K., Keeling, D.I., Hilken, T., Mahr, D.: Tangible service automation: decomposing the technology-enabled engagement process (TEEP) for augmented reality. J. Serv. Res. **1094670520933692** (2020)
49. Heller, J., Chylinski, M., de Ruyter, K., Mahr, D., Keeling, D.I.: Touching the untouchable: exploring multi-sensory augmented reality in the context of online retailing. J. Retail. **95**(4), 219–234 (2019)
50. Javornik, A.: Augmented reality: research agenda for studying the impact of its media characteristics on consumer behaviour. J. Retail. Consum. Serv. **30**, 252–261 (2016)
51. Olya, H., Jung, T.H., Tom Dieck, M.C., Ryu, K.: Engaging visitors of science festivals using augmented reality: asymmetrical modelling. Int. J. Contemp. Hosp. Manag. **32**(2), 769–796 (2020)
52. Alimamy, S., Deans, K., Gnoth, J.: An empirical investigation of augmented reality to reduce customer perceived risk. In: Proceedings of the Academy of Marketing Science and Marketing Transformation: Marketing Practice in an Ever Changing World AMSWMC 2017 Developments in Marketing Science, pp. 127–135 (2017)
53. Dacko, S.G.: Enabling smart retail settings via mobile augmented reality shopping apps. Technol. Forecast. Soc. Chang. **124**, 243–256 (2017)
54. Dwivedi, Y.K., Ismagilova, E., Hughes, D.L., Carlson, J., Filieri, R., Jacobson, J., et al.: Setting the future of digital and social media marketing research: perspectives and research propositions. Int. J. Inf. Manag. **102168** (2020)
55. Neuhofer, B., Buhalis, D., Ladkin, A.: A Typology of technology-enhanced tourism experiences. Int. J. Tour. Res. **16**(4), 340–350 (2014)
56. Ströbel, T., Germelmann, C.C.: Exploring new routes within brand research in sport management: directions and methodological approaches. Eur. Sport Manag. Q. **20**(1), 1–9 (2020)
57. Neuhofer, B.: Innovation through co-creation: Towards an understanding of technology-facilitated co-creation processes in tourism. In: Open Tourism, pp. 17–33. Springer, Berlin/Heidelberg (2016)
58. Buhalis, D., Harwood, T., Bogicevic, V., Viglia, G., Beldona, S., Hofacker, C.: Technological disruptions in services: lessons from tourism and hospitality. J. Serv. Manag. **30**(4), 484–506 (2019)
59. PRNewswire. AR in travel & tourism, worldwide, 2019: market size & forecast, technologies, trends, competitive analysis, value chain analysis. Retrieved June 16, 2020, from https://www.prnewswire.com/news-releases/ar-in-travel%2D%2Dtourism-worldwide-2019-market-size%2D%2Dforecast-technologies-trends-competitive-analysis-value-chain-analysis-300992745.html (2020)
60. Gerrity, C.: How augmented reality travel will re-shape the tourism industry. Retrieved June 3, 2020, from https://www.agencymabu.com/augmented-reality-traveltourism/ (2018)
61. Olsson, T., Lagerstam, E., Kärkkäinen, T., Väänänen-Vainio-Mattila, K.: Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. Pers. Ubiquit. Comput. **17**(2), 287–304 (2013)
62. Herz, M., Brunk, K.H.: Conceptual advances in consumers' semantic and episodic brand memories: a mixed methods exploration. Psychol. Mark. **34**(1), 70–91 (2017)
63. Hassan, A., Ekiz, E., Dadwal, S.S., Lancaster, G.: Augmented reality adoption by tourism product and service consumers: some empirical findings. In: Augmented Reality and Virtual Reality, pp. 47–64. Springer, Cham (2018)
64. Hassan, A., Rahimi, R.: Consuming "Innovation" in tourism: augmented reality as an innovation tool in digital tourism marketing. In: Global Dynamics in Travel, Tourism, and Hospitality, pp. 130–147. IGI Global (2016)
65. tom Dieck, M.C., Jung, T.H.: Value of augmented reality at cultural heritage sites: a stakeholder approach. J. Destin. Mark. Manag. **6**(2), 110–117 (2017)
66. Buhalis, D., Law, R.: Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tour. Manag. **29**(4), 609–623 (2008)
67. Cranmer, E.E., tom Dieck, M.C., Jung, T.: How can tourist attractions profit from augmented reality? In: Jung, T., tom Dieck, M.C. (eds.) Augmented Reality and Virtual Reality, pp. 21–32. Springer, Cham (2018)
68. Guttentag, D.A.: Virtual reality: applications and implications for tourism. Tour. Manag. **31**(5), 637–651 (2010)
69. Huang, Y.C., Backman, K.F., Backman, S.J., Chang, L.L.: Exploring the implications of virtual reality technology in tourism marketing: an integrated research framework. Int. J. Tour. Res. **18**(2), 116–128 (2016)
70. Jung, T.H., Han, D.I.: Augmented Reality (AR) in urban heritage tourism. e-Rev. Tour. Res. **5** (2014)
71. Xu, F., Buhalis, D., Weber, J.: Serious games and the gamification of tourism. Tour. Manag. **60**, 244–256 (2017)
72. Kounavis, C.D., Kasimati, A.E., Zamani, E.D.: Enhancing the tourism experience through mobile augmented reality: challenges and prospects. Int. J. Eng. Bus. Manag. **4**, 10 (2012)
73. Chung, N., Han, H., Joun, Y.: Tourists' intention to visit a destination: the role of augmented reality (AR) application for a heritage site. Comput. Hum. Behav. **50**, 588–599 (2015)

74. Han, D.I., tom Dieck, M.C., Jung, T.: User experience model for augmented reality applications in urban heritage tourism. J. Herit. Tour. **13**(1), 46–61 (2018)
75. Jung, T., tom Dieck, M.C., Lee, H., Chung, N.: Effects of virtual reality and augmented reality on visitor experience on museum. In: Inversini, A., Schegg, R. (eds.) Information and Communication Technologies in Tourism 2015, pp. 621–636. Springer International Publishing, Heidelberg (2016)
76. Jung, T., Lee, H., Chung, N., Tom Dieck, M.C.: Cross-cultural differences in adopting mobile augmented reality at cultural heritage tourism sites. Int. J. Contemp. Hosp. Manag. **30**(8) (2018)
77. Jung, T., tom Dieck, M.C., Lee, H., Chung, N.: Moderating role of long-term orientation on augmented reality adoption. Int. J. Hum. Comput. Interact. **36**(3), 239–250 (2020)
78. Rogers, E.M.: Diffusion of Innovations, 4th edn. Free Press, New York (1995)
79. Eyecandylab. The Power of Video-triggered AR. US Market Study 2019. Retrieved July 10, 2020, from https://cdn2.hubspot.net/hubfs/3934347/studies/eyecandylab_ThePowerofVideoTriggeredAR_September2019.pdf (2019)

Marc Herz is partner at the Berlin-based brand consultancy Kleinundpläcking GmbH. He is an expert in consumer research and customer insights. Prior to that he was analyst and consultant at a UK-based market research and consultancy company. Marc holds a PhD in Management from the University of Vienna and is a marketing and research lecturer in Germany, Austria, France, Denmark, Bosnia and Azerbaijan. In his research, Marc focuses on branding, tourism management, market research and consumer behavior. His work has thereby won several awards and was published in international top journals.

Katrin E. Schein is a research associate and PhD student at the Department of Digital Marketing and Media Innovation at the Bundeswehr University Munich, Germany, and holds a Master's degree in Marketing Management from the University of Groningen, Netherlands. Katrin is involved in multiple research and teaching projects related to XR with a strong focus on augmented reality (AR) and its influence on consumer behavior and user acceptance as well as psychometric measurement. Prior to that she worked as a product manager in the marketing department of international companies in the cosmetics industry.

Prof. Dr. Philipp A. Rauschnabel is a Professor of Digital Marketing and Media Innovation at the Bundeswehr University Munich, Germany. Prior to that he was a faculty member at the University of Michigan-Dearborn, USA. His research focuses on contemporary issues in the field of XR, especially augmented reality (AR), mixed reality (MR) and virtual reality (VR). He has published numerous journal articles on these topics. Rauschnabel holds a PhD in Marketing (University of Bamberg) and a Master's degree in Marketing and Channel Management. He is currently founding www.xrealitylab.com, a platform for XR managers and scientists.

Part V

Applications in Engineering and Science

Augmented Reality Uses and Applications in Aerospace and Aviation

20

Maryam Safi and Joon Chung

Contents

20.1 History of Augmented Reality in Aerospace 473

20.2 Applications in Navigation and Guidance 474
20.2.1 Uses of AR in Manned Navigation and Guidance 474
20.2.2 Unmanned Navigation and Guidance 479
20.2.3 Air Traffic Management 480

20.3 Applications in Engineering, Manufacturing, and Maintenance 482
20.3.1 Engineering Design and Visualization 482
20.3.2 Manufacturing and Maintenance Support 483

20.4 Applications in Space Operations 486

20.5 Future Application Concepts 487
20.5.1 Airport Security 488
20.5.2 Crew Task Support 488
20.5.3 In-Flight Entertainment and Communication 489
20.5.4 Augmented Reality and Artificial Intelligence 489
20.5.5 Augmented Reality and Haptic Feedback 490
20.5.6 Augmented Reality Applications for COVID-19 Support 491

20.6 Concluding Remarks 491

References 492

Abstract

Augmented reality's (AR) first aerospace application came around at the end of the First World War in the form of the aircraft gunsight. Over the last century, this technology evolved to serve different areas of need, including engineering, training, space operations, and flight crew support. The most notable AR applications are the head-up display (HUD) and the helmet-mounted display (HMD), which are used to provide navigation assistance to pilots flying high-profile missions. In recent years, various aerospace corporations have turned to AR technology to enhance their prototyping, manufacturing, and maintenance operations. Similarly, airlines and airports are investigating the use of AR to support daily tasks of operating crew and personnel. These efforts are supported by research groups studying the potential of AR in air traffic control and management, airport operations and security, crew training, aircraft cabin visualization, and In-Flight Entertainment and Communication (IFEC). This chapter presents an overview of AR's history and applications in the aerospace industry throughout the years.

Keywords

Aircraft simulation · Aircraft navigation · Airport operations · Aerospace engineering · Aircraft manufacturing · Crew training

20.1 History of Augmented Reality in Aerospace

The introduction of augmented reality (AR) into aerospace is not well defined. A common thought is that the technology was first used in the mid-twentieth century in the form of navigation assistant technology such as the head-up display (HUD), a non-wearable AR device that allows elements of the real environment to be viewed along with augmented information. The earliest record of augmented information is found in Irish telescope maker Sir Howard Grubb's patent no. 12108, which was awarded in 1901 [5]. Sir Grubb's patent described a device that would help users aim firing weapons. This invention would later become known as the gunsight. Though the gunsight combines real-world views with additional information to supplement user knowledge of the real environment, it only displayed static information. As such, the gunsight does not fit the definition of AR presented by Azuma in 1997, which specifies that the technology must provide interactive information in real time.

M. Safi · J. Chung (✉)
Department of Aerospace Engineering, Ryerson University, Toronto, ON, Canada
e-mail: maryam.safi@ryerson.ca; j3chung@ryerson.ca

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_20

During the First World War, German optics manufacturer Optische Anstalt Oigee was inspired by Sir Grubb's invention to use the gunsight on board the Imperial German Army Air Service's Fokker DDS1 and Albatros D.III fighter aircraft [5, 46]. Optische Anstalt Oigee developed the Oigee Reflector Sight in 1918, paving the way for AR use in the aerospace and aviation industry. For years after this occurrence, AR would be used in the aerospace and aviation industry in various ways, though the technology was not known as "augmented reality" until the late twentieth century.

In the early 1960s, the first HUD with dynamic display capabilities was used as an aircraft navigation tool on board a British carrier-borne combat aircraft called the Royal Navy Buccaneer [46]. The HUD became a standard device on board military aircraft in the 1970s. Some years later, it revolutionized into the helmet-mounted display (HMD) through Thomas A. Furness III's design for the US Air Force's Super Cockpit program [24, 25]. The fundamental difference between the HMD and the HUD arises in that the HMD is a wearable device. As a result, the HMD introduces new capabilities such as head tracking and almost 360-degree field of view (FOV).

AR developments in aerospace were initially inspired by the need to improve weapon aiming accuracy on military aircraft [46]. The discovery of AR's ability to improve overall pilot performance led the civilian aviation sector to adopt the technology. American aerospace manufacturer, The Boeing Company, also known by its short name Boeing, was inspired by Furness III's HMD to research AR as a method of enhancing aircraft assembly operations. The goal was to create a tool that would assist technicians and engineers in wiring harnesses in aircraft fuselage [22]. The potential of Boeing's project was presented by Thomas Caudell and David Mizell at the Twenty-Fifth Hawaii International Conference, whereby the term "augmented reality" was coined [12].

Today, AR is used daily in aerospace and aviation to support various tasks. Many civilian aircraft such as the Boeing 737 and the Embraer 190 offer optional HUD systems for the pilot-in-command or captain [46]. Moreover, the Boeing 787 was the first aircraft to add a HUD as standard equipment rather than add-on tools. AR is further used to support manufacturing, maintenance, and space operations. Many companies are also studying new applications in less conventional fields related to aerospace. These were surveyed and presented by Safi et al. [70]. Emerging applications of AR for aerospace and aviation include In-Flight Entertainment and Communication (IFEC), airport security, and crew operation support.

20.2 Applications in Navigation and Guidance

The beginning of AR in aerospace and aviation took place in the navigation and guidance sector. Military research efforts in the mid-twentieth century created opportunities for AR to be used for more than just gunsights. The invention of the HUD and the HMD allowed the use of AR to support aviation and navigation tasks. These tasks are carried out by pilots flying both manned and unmanned aircraft. In simple terms, AR supports aviation and navigation tasks by improving the pilot's understanding of the surrounding environment, airspace, and terrain. Applications for AR in air traffic control and management have been studied for a few decades. These studies assist the air traffic control officer (ATCO) by augmenting key airport and weather information on a window view of the airport runway or taxiway. The developments of AR for aerospace navigation and guidance can be divided into three categories: manned navigation and guidance, unmanned navigation and guidance, and air traffic management (ATM). Figure 20.1 summarizes the developments of AR devices used for aerospace navigation and guidance.

20.2.1 Uses of AR in Manned Navigation and Guidance

Early work in AR for aerospace focused on using the technology to assist pilots in navigating and aviating manned aircraft. The tasks of navigating and aviating hold explicit definitions in the aerospace field. All pilots are responsible for three primary tasks when in control of an aircraft: aviate, navigate, and communicate. The task of aviating describes operating and controlling the aircraft. In everyday language, this is often referred to as the act of "piloting." One can appreciate the complexity of this task by understanding that aviating includes all maneuvers required to bring the aircraft into the air, keep the aircraft in steady level flight, and land the aircraft safely on the ground. Modern aircraft are equipped with systems capable of automating large parts of the aviating task; however, pilots are still expected to remain diligent and ready to take over aviation control should any problems arise.

The task of navigating includes planning and directing routes that will allow the pilot to aviate from a given start point to their destination. Correct planning of navigation routes is critical for air safety. Similarly, maintaining preplanned navigation routes is critical to help avoid mid-air collisions and accidents. However, it must be noted that

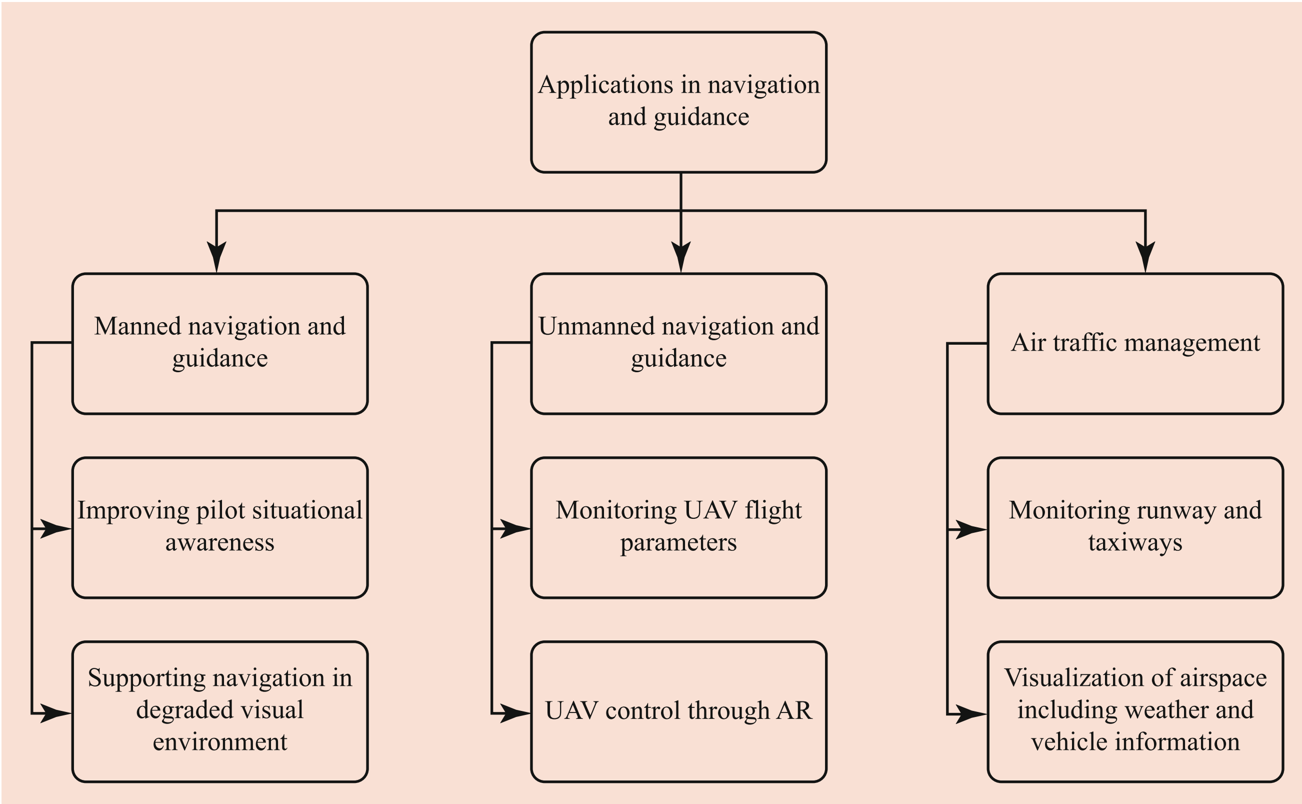


Fig. 20.1 Summary of AR applications in aerospace navigation and guidance

sometimes rerouting becomes essential for safety, especially when weather conditions change undesirably. In such instances, communication becomes increasingly critical. Communicating is sharing information between pilot-in-command and co-pilot, pilot and other aviation crew, or pilot and control tower. Communication takes place through all phases of flight and is essential for maintaining safety on the runway, taxiway, and air.

In addition to the tasks mentioned above, pilots must monitor their vehicle and the surrounding environment. This is typically done by monitoring cockpit instruments which display aircraft data, navigation routes, environmental terrain, and more. Instruments that present flight data are of most importance and so are more often monitored by pilots. These instruments are referred to as “the six pack” by many aviators, a nickname derived from the fact that there are six instruments in total. The six pack is found on every manned aircraft and includes the airspeed indicator, attitude indicator, altimeter, vertical speed indicator, heading indicator, and turn coordinator. In glass cockpits, the six pack is found in a consolidated single screen device called the primary flight display (PFD).

The importance of the six pack can be explained through an automotive example. In an automotive vehicle, drivers will monitor the speedometer more often than other dashboard instruments. A driver may glance at their fuel indicator once or twice during a drive, typically when they first enter the vehicle. They may glance at the fuel indicator again after a particularly long drive or when they are conscious of the fact that the vehicle is low on fuel. Conversely, a driver will glance at the speedometer every few minutes while driving. This is mainly due to the fact that vehicle speed is an essential measure for road safety. On an aircraft, pilots will monitor the six pack in a similar fashion to the speedometer. They will maintain awareness of current readings and settings on these six instruments. They will also monitor engine data like fuel quantity and engine power, but less frequently.

Much like automotive vehicles, aircraft instruments are located in the cockpit on a dashboard-like display panel. The panel is easily viewable but requires the pilot to actively look away from the surrounding airspace found outside the window. Once more, this can be likened to the driving experience, whereby a driver will look away from the road and at their dashboard instruments or warning lights. While this is typically not a problem, especially if done for short periods, drivers sometimes find themselves near collisions caused by their lack of attention paid to the road ahead. Pilots face similar situations, except that the higher speeds and larger aircraft size make it even more critical for pilots to react quickly. However, as pilots need to monitor six or more instruments at any given moment, the amount of time they spend with their heads down is longer than that of automotive drivers.

The monitoring of flight instruments and surrounding airspace allows pilots to receive information about their current situation. This information is then understood through cognitive processes used by the human mind to form mental models that shape decision-making and problem-solving. This mental process is not exclusive to pilots. Humans are constantly receiving and processing information from their surroundings. However, pilots engage in deliberate information gathering and processing, decision-making, and problem-solving while in the aircraft. The long duration through which pilots continuously carry out these mental efforts can lead to mental exhaustion and fatigue [68]. Fatigue is a major concern in aerospace as it directly influences safety.

Exhaustion and fatigue can increase the chances of pilot error. These can occur due to decreased situational awareness (SA), cognitive slowing, impaired judgment, or lapsing. Pilot error is classified into slips and mistakes [48]. Slip errors occur when wrong actions are performed, leading to undesired consequences. Pilots are said to slip when they identify the correct goal but carry out the incorrect action. Mistakes are errors that occur due to the wrong goal selection. A mistake occurs when a planned action is correctly taken but the plan was erroneous.

The measure of knowledge of one's surroundings can be described through SA. SA is most comprehensively defined by Endsley [18] as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future." In essence, SA is composed of three major parts: perceiving the environment and the elements within it during a specific amount of time, understanding these elements and their behavior or effects on the environment, and forming logical expectations about the status of the said elements in the near or immediate future. Pilot SA can be affected by many factors, including workload. As mentioned previously, aircraft require pilots to switch between viewing instruments located within the cockpit and the sky viewed from the cockpit window, which forces pilots to look down in order to obtain readings. The constant switching between head-up and head-down positions adds to the pilot workload, leading to exhaustion. It is an inherent challenge of both traditional aircraft equipped with electromagnetic instruments and newer aircraft featuring glass cockpits. The long head-down times found in pilots reading flight instruments resulted in the creation of the HUD, which was used to increase SA by decreasing pilot head-down time.

The HUD was first introduced in military aviation, where SA is critical for pilot survival. Military pilots need to be constantly aware of their surroundings, as they must track enemy aircraft, allied aircraft, and weapon targets. Although it is essential for the military pilot to monitor their flight instruments, the act of looking down at flight instruments can pose a risk to their safety because it takes away from their ability to see their surroundings. For this purpose, the HUD was created to mimic the information provided on the PFD using a combination of symbols and numerical values called symbology. As shown in Fig. 20.2, HUD symbology is traditionally presented in a standardized green color, though more recent technologies have allowed for full-color HUD devices. Aircraft data is displayed onto a transparent screen positioned in front of the cockpit window. Hence, the HUD combines primary flight information with out-the-window visual references onto a single visual scene [46]. In this way, the HUD allowed pilots to view aircraft flight data while looking out the window. This reduced head-down time decreases workload and improves SA.

While the HUD improved the flight experience for many aviators, it was limited in FOV and only provided flight information while the pilot looks out the front cockpit window. Militaries began using the HMD to view flight information regardless of the pilot's head position [2, 35]. The HMD continuously retains generated information in the pilot's line of sight, allowing augmented information to be accessible to the pilot across a 360-degree FOV. This is possible as the HMD is a wearable device, with all components of the HUD, mounted to a wearable object such as a helmet. Continuous access to flight information is critical during dogfights where pilots need to look around to gather sufficient data and SA actively. Modern developments enable military aircraft to fuse onboard sensor images with flight information on the HMD. Cueing systems further assist pilots by enabling the use of head motion to select targets and aim sensors or weapons at desired locations. Aircraft, like the F-35, rely entirely on the HMD as the aircraft's PFD, integrating a full view of the aircraft's surroundings into the flight deck [35]. The HMD is better able to supply pilots with flight information while maintaining an out-the-window view, which makes it more suitable for military aviation where high-performance aircraft, demanding maneuvers, and flights with extensive air-to-air or air-to-ground missions are commonly paired up with a need for high SA [2].

The advantages of AR technology in military aviation prompted civilian aviators to use the HUD to increase pilot SA [35, 75]. Studies on HUD usefulness conclude that the technology can improve overall pilot performance across different flight maneuvers such as track and glideslope main-

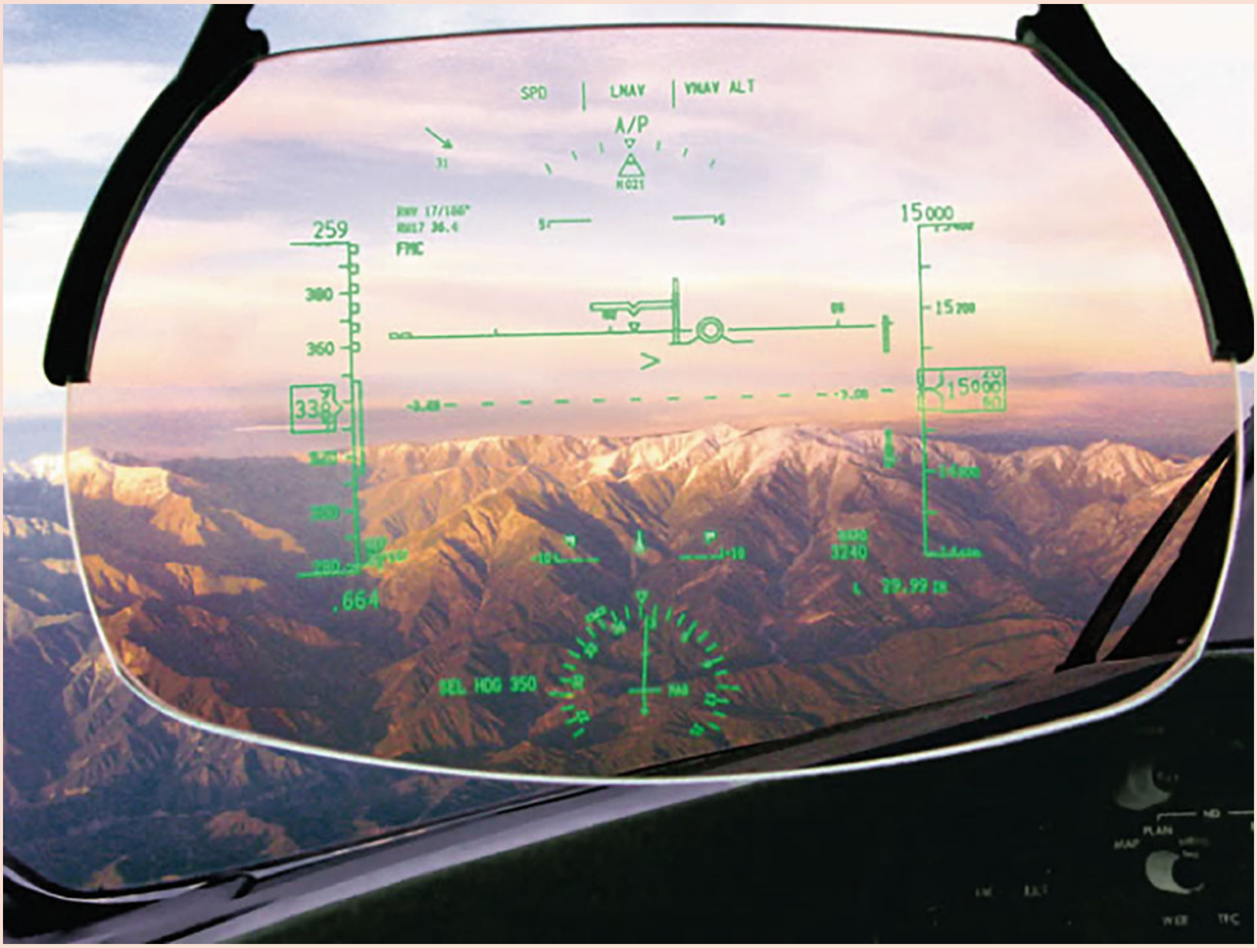


Fig. 20.2 An example of a HUD found on the B787 aircraft. (Image © Boeing)

tenance on approach, landing precision, and other demanding tasks or high-profile missions [46]. It is generally agreed that the HUD can increase SA, routine flight control, and detection of expected events compared to traditional instrument panels or the head-down display (HDD). The HDD is an AR device placed within the cockpit instrument panel. It requires pilots to look away from the window, so it does not solve the problems associated with long head-down times.

AR is also used on board rotorcraft and helicopters to assist in navigation in degraded visual environments (DVE). DVE are characterized by low out-the-window visibility and may be caused by brownouts or whiteouts. In DVE, the visibility of the surrounding airspace and terrain decreases tremendously, making it very difficult for pilots to maintain SA. Flights in DVE are associated with a higher risk of collisions, mainly because the pilot cannot see the environment outside the cockpit window. Navigation in DVE cannot be done through visual reference and requires instrument readings. In a traditional rotorcraft, the vertical situation display (VSD) and the horizontal situation display (HSD) are used to assist pilots navigating in DVE [29]. Pilots use the HSD and VSD to maintain an understanding of their position relative to the terrain around them. Data presented on the HSD is often thought of as a "top-down" view of the airspace and so is used to understand the position of the rotorcraft relative to the surrounding space. For simplicity, this can be thought of as a position left, right, ahead, or behind another object. The VSD gives an understanding of the vertical position relative to the terrain. This can be thought of as the high of the rotorcraft relative to a given object or terrain surface.

Recently, simulations have been conducted to test for new solutions for DVE navigation methods using AR technology. AR-based navigation for DVE uses on-aircraft sensors to obtain data of the surrounding terrain. Data collected from these sensors is used to create a map of the aircraft's surroundings, shown to the pilot on a screen placed within the cockpit. Flight information is then augmented onto this map by onboard computers, giving pilots access to their aircraft data and a synthetic image of the surroundings. Terrain maps typically highlight ground elevation and nearby obstacles like trees or buildings. The map helps pilots to formulate an idea of where their aircraft is situated relative to these obstacles, which allows them to avoid terrain collisions better. The Technical University of Munich conducted a simulation on navigation in DVE and showed that the use of an AR-based HMD could decrease pilot workload and increase SA. AR can display flight paths and obstacles that are otherwise not easily viewable by the pilot flying in DVE [78]. The research found that the use of AR improved situational and collision awareness, allowing pilots to better navigate environments with low out-of-window visibility [61]. Collision awareness refers to the pilot's alertness of the risk of collision with objects in the airspace.

Despite the use of AR devices to support navigation tasks, drawbacks of the traditional HUD and HMD exist in different forms. One of the most important forms is attention capture or cognitive tunnelling [46]. Cognitive tunnelling refers to the potential for highly saturated information to control attention. With a large amount of information present at any given moment, the HUD can draw the attention of pilots away from the environment located out the window. Under the changing workload, pilots may have difficulty processing co-located information on the HUD and in the surrounding airspace, hence hindering performance. Cognitive tunnelling can be managed in numerous ways, including increasing HUD brightness and decluttering HUD information, which led to the introduction of a declutter mode on most HUD devices used today. HMDs can further introduce ergonomic concerns for pilots [2]. Due to the weight of the technology, HMDs are not viable solutions for long flight durations commonly found in civilian aviation. This is one reason why the HUD, but not the HMD, has been widely accepted within civilian aviation. Civilian aviation often requires pilots to fly routes that last many consecutive hours of aviating. For example, a common flight route is the New York City to Los Angeles flight, which takes approximately 6 h to complete. Such long flight routes make wearing bulky or uncomfortable devices unrealistic for many pilots.

In the 1990s, the National Aeronautics and Space Administration (NASA) developed an HMD at their Langley Research Center that weighed just under 3 kg [2]. Other HMD developed around that time weighed around 4.5 kg, which made NASA's HMD revolutionary for its time. The device could be worn by most pilots who participated in the corresponding study for about 1 h before discomfort was reported. Hence, it is clear that the HMD is not a viable option for civilian pilots. The HMD also brings physiological challenges into consideration, as it may cause visual and vestibular interactions that can negatively affect user comfort. Phenomena such as latency and occlusion can affect the user's perception, hindering performance by causing confusion, motion sickness, or further discomfort [2].

Research teams are exploring the benefits and drawbacks of using AR in civil aviation through simulation efforts. An extensive research area is motivated by HUD FOV limitations that arise from the design and construction of these devices [2, 5]. Another research area focuses on HMD drawbacks, such as visually induced motion sickness and temporary optical or neurological damage to users. Other research efforts are looking to improve existing systems through the enhancement of symbology shown on AR displays. Current HUD symbology standards use 2D symbols and data to augment real-world views. However, visual information found in the real world is often three dimensional in form, which makes 3D visual information more intuitive to humans.

There is an increasing interest in using a combination of 2D and 3D symbology on aviation HUD and HMD devices. Gorbunov [27] first suggested that the use of 3D symbology may reduce common HUD problems like attention capture and cognitive switching. Presenting pilots with 3D symbology is thought to give more intuitive cues closer in representation to pilot natural perception, making them more suitable for navigation devices. Researchers have tested the use of 3D symbology in the form of stereoscopic markers and symbols or a tunnel-in-the-sky on pilot performance [27, 28]. A tunnel-in-the-sky is a 3D cue that marks desired flight paths using a series of symbols, often rectangular, through which a pilot is to guide the aircraft. Figure 20.3 shows a tunnel-in-the-sky concept developed by NASA's Langley Research Center. This concept uses magenta-colored rectangular sections to mark the "correct" path through which the aircraft should be flown. The idea is that as long as the pilot's view of these rectangular sections is centered, they can rest assured that they are on their desired flight path. Studies concluded that the addition of 3D symbology could enhance pilot performance. However, concern for overload and attention capture remains a leading reason why 3D symbology is not widely used in the industry.

The demand for the HMD and similar worn displays within the civil aviation market has not reached a threshold needed to drive further research and development. Although new head-worn displays and smart glasses such as the Microsoft HoloLens are in development, many are the result of overall interest in AR technology and are not specific to aerospace applications. These devices offer reduced weight and superior performance, making them ideal candidates for a wide variety of HMD-based uses. Smart glasses have revolutionized the industry by creating lighter and smaller wearable AR devices, which can solve the existing HMD weight problem. One example of smart glass use for manned aircraft is Aero Glass, a navigation assistance software developed by Akos Maroy and Jeffrey Johnson [5]. Aero Glass combines sensor-collected information with automatic dependent surveillance – broadcast to display important navigation cues to pilots on the Epson Moverio wearable AR glass. Aero

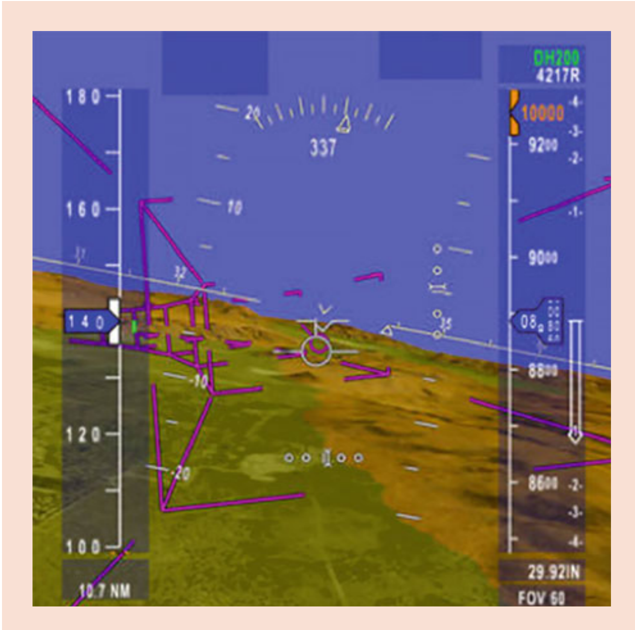


Fig. 20.3 Tunnel-in-the-sky concept developed by NASA's Langley Research Center. (Image © NASA)

Glass provides pilots with navigation features in 3D with an advertised 360-degree FOV. The system is designed to be compatible with a variety of aircraft and simulators. In essence, Aero Glass works as a multi-aircraft HMD, which can be used by leisure pilots in smaller aircraft. General aviation aircraft flown for leisure purposes seldom equip AR devices on board due to weight restrictions. Aero Glass enables general aviation pilots to use AR navigation devices without concern about weight limitations. The low device weight also allows pilots to use the glasses for extended periods compared to the conventional HMD.

Other examples of innovative AR use in aerospace navigation is the Bell Helicopters initiative that combines AR and artificial intelligence (AI) to create a flight control model for manned vehicles [62]. The project uses the Microsoft HoloLens for visualization of flight parameters. In 2018, BAE Systems announced plans for developing an AR cockpit equipped with fully interactive and reconfigurable displays [52]. The company investigates ways to increase the amount of information provided to fighter pilots without hindering performance through overload. Smart glasses are also frequently used to provide support for navigation and guidance of unmanned vehicles like quadcopters and drones.

20.2.2 Unmanned Navigation and Guidance

AR systems are used to provide flight information to remote pilots operating an unmanned vehicle. The unmanned aerial vehicle (UAV) is characterized by the lack of onboard cockpit, pilot, and crew. UAVs are used for a wide variety of missions, including cargo transport, military operations, recreational flying, and filming. This makes UAVs particularly useful to various industries, like medicine, commerce, and mail delivery in highly urban areas. UAVs have been used extensively in atmospheric, geophysical, and oceanographic research, mineral exploration, telecommunications relay, traffic and accident surveillance, emergency monitoring, search and rescue operations, and much more [50]. The use of UAVs to deliver medical kits to disaster sites and transplant organs to hospitals has also been explored. The aerospace industry is also developing methods of using UAVs, coupled with AR, to inspect aircraft structures easily.

Historical challenges with UAV use include the difficulty in navigation and guidance associated with remotely piloted aircraft. The UAV pilot operates from outside the vehicle, meaning they are physically away from the aircraft they control. Pilots may be near the vehicle or in a geographically distant area which makes maintaining visuals on the UAV impossible. The result of this separation between pilot and aircraft is a total reliance on flight instruments in order to complete the flight mission. With only flight instruments to guide the pilot, UAV navigation becomes cognitively challenging. The inability to see the UAV and its surrounding environment makes it difficult for pilots to form a good mental model of the situation. The consequence is a decrease in SA, which can endanger the success of the flight mission. Much like manned navigation, UAV navigation is dependent on pilot cognition. High SA is essential for pilots to maintain a good understanding of the aircraft's situation, position, and the surrounding environment. As discussed in Sect. 20.2.1, high SA is essential for pilot decision-making.

Much like manned navigation, supporting pilot SA is a priority for unmanned navigation. New technology in the form of first-person view (FPV) displays enabled pilots to maintain high SA by providing UAV surroundings. FPV displays stream live video footage obtained from cameras mounted to the UAV exterior, hence giving pilots an insight on the aircraft's surroundings. This feed simulates the out-the-window view available to pilots on manned aircraft. Coupled with flight instruments, FPV displays provide UAV pilots with the ability to understand the current flight conditions. This understanding can be further enhanced by the use of AR. FPV displays can be augmented with flight data to create a HUD, HDD, or HMD device for UAV pilots. HMD-type devices are commonly used for UAV navigation today. Smart glasses such as the Epson Moverio BT series are well known for their FPV display capabilities, especially when paired with DJI Phantom multirotor drones [83]. Using the Epson Moverio smart glass, pilots can view flight information of their DJI Phantom drone augmented onto an FPV feed. The FPV is obtained through a camera mounted onto the UAV, allowing the pilot to perceive the vehicle's surroundings and flight parameters even when they are not locally present within the same environment. Flight information such as UAV altitude, attitude, vertical speed, airspeed, power, and energy reserves can be augmented on the FPV feed. This allows pilots to access all necessary information in a single glance, removing the workload required to switch attention between traditional flight instruments and the FPV feed.

In addition to supporting pilot SA, AR is being used to control aircraft and UAVs. The Immersive Semi-Autonomous Aerial Command System (ISAACS) is an AR-based, open-source project centered around drone control [55]. ISAACS was launched by the University of California, Berkeley, and DJI to explore new methods of navigating and commanding UAVs. The project works to create intuitive human-machine interfaces [20]. While the project focuses on UAV control, the interface solution provided by ISAACS can be applied to control of robotics in general [20]. ISAACS uses the Microsoft HoloLens as a human-machine interface coupled with simultaneous localization and mapping (SLAM) algorithm to guide UAVs in performing various actions specified by a remote pilot. ISAACS improves human-machine interfaces by enabling pilots to control their UAV through gesture and voice commands [55]. The system also provides augmented FPV of the UAV. Information such as drone battery power and current task is augmented through the Microsoft HoloLens to provide better pilot understanding. Through ISAACS, pilots can control an aerial fleet and are not limited to interacting with a single UAV at any time [20].

In 2017, the Drone Research and Application Centre (DRAC) at the University College of Technology Sarawak developed a flight simulation based on the AR-to-gamepad interface [77]. The simulation deploys the prefab model of a multirotor drone to the Microsoft HoloLens. Similarly, Y Media Labs developed an AR application for UAVs in conjunction with DJI and Epson Moverio [54]. The application offers training and test modules for UAV pilots through an AR-based game. Game data is shown on a pair of Epson Moverio BT-300FPV smart glasses, while the UAV is controlled through a gamepad. The application augments a UAV vehicle to create a full AR simulation. In 2017, a new AR application for DJI drones was developed by Edgybees. The application is called Drone Prix AR and uses the Epson Moverio BT-300 smart glasses to provide AR gaming experiences for UAV pilots. Drone Prix AR augments virtual racetracks and collectibles through the Epson smart glasses [53].

The popularization of UAVs for recreational flying led to the development of numerous AR applications for drone guidance. In 2012, the Parrot company developed and sold the AR Drone quadrotor as a toy for AR games [40]. Smartphones or tablets control the AR Drone through the wireless Internet connection. The low cost of the quadrotor and the open-source application program led to the use of the drone for research purposes. The AR Drone has many open-source programs, including the MATLAB AR Drone Simulink Development Kit [40]. Epson and DJI introduced a new AR application in 2018 named Drone Soar, which allows pilots to control their DJI brand UAVs through an HMD-style application [56]. UAVs are provided input through a handheld controller. A pair of Epson Moverio BT-300 smart glasses is used as the accompanying HMD. DJI offers many popular commercial drone models, which are supported by Drone Soar. The application is equipped with FPV feed overlaid with AR navigation symbology, including flight status, distance, altitude, and battery level. Through GPS, the application makes maps and restriction warnings regarding nearby geographical areas accessible to pilots.

20.2.3 Air Traffic Management

ATM applications of AR work minimize the ATCO head-down time [30, 70]. While these applications are not widely used worldwide, studies have been conducted to test the feasibility of such technology in assisting ATCO. The work of an ATCO is highly stressful. The expectation is that the ATCO will carry out all responsibilities and tasks perfectly and faultlessly [39]. These tasks include monitoring aircraft in the surrounding airspace, communicating with pilots, and solving conflicts that threaten air safety. ATCO solves issues like maintaining minimal separation between multiple aircraft in a given airspace, takeoff, landing priority, and ground vehicle directions.

In general, each ATCO is assigned a section of the surrounding airspace to monitor and control. The increase in air traffic forces the ATCO to monitor a larger number of vehicles within a given space or location [39]. It also increases the probability of error, scenario difficulty, and ATCO workload. An ATCO will alternate between gazing at computer screens located within the tower and looking out the tower window [37]. Conventionally, an ATCO has roughly 12–49% of its operating time to look up into the airspace they control [63]. Hence, 51–88% of their time is spent looking at HDD-based screens located within the control tower, a phenomenon that is commonly called head-down time. The switching between viewing out-the-window scenes and the HDD requires the ATCO to shift their visual depth. This was found to increase the risks of mission-critical events in airspace [31]. During morning operations or good weather conditions, out-the-window visibility is high, enabling the ATCO to view the airspace and build an understanding of the situation. However, in low-visibility conditions caused by poor weather or low light levels, the out-the-window scene becomes difficult to view, making it challenging for an ATCO to understand the situation within their allocated airspace [37]. Without sufficient aid, ATCO performance can decrease, putting the crowded airspace's safety in jeopardy [39].

In such cases, AR or virtual reality can become useful tools for helping the ATCO perform their operational tasks efficiently and safely [37]. Replacing HDD screens with AR screens positioned within the ATCO FOV allows the ATCO to view both the airspace and the information needed at the same time [23, 32, 65, 66]. This can increase the ATCO's awareness of operational airspace, reducing the risk for mission-critical events and accidents. AR can be used to minimize ATCO head-down time by displaying relevant information of the surrounding space onto a screen mounted directly ahead of the ATCO. Flight data of aircraft in the nearby airspace, runway and taxiway occupancy, ground vehicle position, and trajectory are critical information for an ATCO on the job. The ATCO needs them for organizing and ensuring safe operations. This information can be provided to an ATCO on a transparent display positioned directly in front of the tower window. Effectively, the ATCO workstation will be equipped with a HUD showing essential airspace information. AR can also help the ATCO by increasing SA in low-visibility conditions. As mentioned in Sect. 20.2.2, visibility is vital for humans to gather cues about the surrounding environment. ATCOs use the out-the-window scene and available radar data obtained from the airport surveillance radar (ASR) and the surface movement radar (SMR) to form a mental model of their allocated airspace. The formed mental model is critical for problem-solving and decision-making, which are two cognitive processes found at the core of the ATM. In low-visibility conditions, sound decision-making and problem-solving become more difficult, mainly due to the ATCO's inability to form complete mental models due to insufficient visual cues. Using AR, information gathering can be made easier for an ATCO by augmenting real-world objects and obstacles with computer-generated visuals. For example, nearby buildings, runway lines or numbers, and parked aircraft can be highlighted on the ATCO's workstation window. This does not only enable the ATCO to visualize the airspace during low-visibility conditions better, but it can be used to provide more information through the use of color to highlight empty runways, fallen debris, parked ground vehicles, and more. An implementation of this idea is already underway. Hungary-based start-up company 360world became one of the first commercial implementors of AR for ATM [59]. The company created a mixed reality (MR) system called Clairity and a corresponding AR application named HoloTower that displays real-time information like weather reports, radar, and camera feed to the ATCO. The project is capable of recognizing aircraft taking off or landing within the surrounding airspace, making ATCO's work easier. Clairity offers a low-visibility mode that renders 3D visuals of nearby aircraft and airport settings. The company is also offering a Clairity SmartBinocular device capable of automatically tracking aircraft and providing alerts within the ATCO FOV [1].

AR has not been as commonly applied in ATM when compared to manned and unmanned navigation. At this time, the majority of AR applications for ATM are at a conceptual or testing level. In the late 2000s, investigations into the feasibility of using AR to assist an ATCO with their tasks were carried out. Pinska and Tijus [63] conducted a study to test the viability of AR use for ATCO assistance. Their study found that ATCO dependence on head-down tools increased during high traffic loads, although the out-the-window view from the control tower is essential for maintaining SA during these times. The study concluded that replacing information found on the HDD with a HUD will decrease head-down time significantly. This swap is also predicted to eliminate concerns raised by Hilburn [30] regarding fixation switch between near and far locations. The expectation is that the use of a HUD will increase SA, decrease workload, and improve the performance of ATCOs. Research suggestions for AR use in ATM have pointed mainly to the use of a HUD but not an HMD. This is because, unlike the HUD, the HMD was found to be unsuitable for ATM operations. Hofmann et al. [32] state three major issues that arise from HMD use for ATM. First, the device can only allow the ATCO to view a small amount of information at a given time, due to the small waveguide lenses used in the HMD. Second, many HMD devices are tethered to a computer, so they demand that the ATCO remain in a fixed position. This limits the range of motion of the ATCO, forcing them to remain near their workstation. The result is difficulty accessing more out-of-window information that may not be available within the ATCO's direct FOV. Lastly, occlusion issues prominent in the HMD could lead to confusion of ATCOs. Due to occlusion, an ATCO may be misled in their understanding of aircraft location relative to nearby obstacles or other objects. Table 20.1 summarizes the advantages and disadvantages of using HUD and HMD devices for ATCO support.

Although untethered HMD devices exist at the time of writing, limitations to direct FOV make them unsuitable for ATM operations. An untethered HMD operates wirelessly and so is not connected to a computer through cables. Due to the construction of HMD devices, the visibility of augmented information is restricted to a small vertical FOV at any given time, which is not ideal for ATM use. ATCO benefits most when information is placed in a single plane located directly ahead of their neutral head position, as it allows them to view desired data while maintaining visuals on the runway or taxiway located outside the window. In addition, HMD devices induce physical discomfort when used for long durations. This discomfort is sometimes caused by visual phenomena but more often caused by the device's weight. The nature of ATM tasks will require an ATCO to wear the HMD for many consecutive hours, which can cause

Table 20.1 Advantages and disadvantages of the HUD and HMD for ATM support

Device	Advantages	Disadvantages
Head-up display (HUD)	Allows freedom of movement for the user The weight does not affect user comfort Larger field of view enabling the addition of more information Viewable by all ATCO in the area, making collaboration easier Direct attachment to computers enables faster processing and information display	No head tracking ability at this time Information only accessible at certain head positions or angles Ability to view augmented information at other ATCO's stations may distract individuals
Helmet-mounted display (HMD)	Head tracking capability built-in Augmented information always available to ATCO as a result of head tracking Information viewable by single ATCO, reducing the risk of distractions for others	The small field of view limiting the quantity of information Prone to occlusion problems that can cause confusion The weight of the device can cause discomfort or pain Information viewable by single ATCO, making collaboration difficult Untethered devices may suffer from latency issues as information is transferred Untethered devices with built-in computers do not provide sufficient processing power Devices may not be suitable for ATCO requiring corrective lenses

discomfort or pain. As a result, large holographic screens with head tracking capability may present the best solution for ATM. The technology would draw on the advantages of both the HUD and the HMD. The out-the-window view can be augmented in a reconfigurable manner, without limiting the motion of ATCOs. Moreover, physical discomfort may be lessened as the AR device will not be mounted to the operator. The addition of head tracking technology will enable information to be redisplayed for an ATCO regardless of head position.

20.3 Applications in Engineering, Manufacturing, and Maintenance

Outside of navigation and guidance, AR is used to support engineering, manufacturing, and maintenance tasks. The use of AR in these operations enhances visualization and understanding for personnel. Computer-aided design (CAD) models can be seen within the real-world environment, giving workers a better idea of what the final product will look like and how it fits within an assembly of parts. Aerospace companies use a mixture of augmented and virtual reality technology to improve the daily design, manufacturing, and maintenance operations.

20.3.1 Engineering Design and Visualization

AR can be used to enhance design and optimization tasks by allowing engineers to envision CAD models within the real environment [41, 73, 80–82]. The engineering design process is iterative in nature. Design solutions are updated and enhanced after each review stage. Typically, the engineering team will meet together to discuss design solutions, during which 2D images of the parts or designs in question are available for viewing. A 3D version of the design may be made available through a computer with screen-share capabilities. However, only one individual can control the model at a time. This makes collaboration a little challenging, as team members cannot easily manipulate and access the design. The meeting relies on communication between the team to request manipulation or change of the displayed model, which can be disruptive to the overall discussion. Engineers can also access 3D models during design review meetings by creating prototypes. Early prototypes are usually 3D printed to create quick mock-ups. The current process separates the designed part from the real-world environment until the said part is manufactured. This makes it very difficult for engineers to visualize how the design will fit within existing assemblies.

For instance, an aerospace structural engineer may need to redesign a structural part and fit it within an existing aircraft assembly. To do this, the engineer will need to understand the existing assembly, which they can do in two ways: either by studying the CAD model of the aircraft assembly or studying the existing structure of the manufactured aircraft. The problem with both methods is that not all relevant information may be visible or available at a given time. The high computational demand for loading a full aircraft CAD assembly leads engineers to access smaller components of the aircraft CAD model at a given time. The result is a gap-filled understanding of how the full aircraft structures fit. This can lead engineers to develop or approve design solutions that may clash with the existing structure, pipelines, wires, and interior installations. When such clashes are detected at the manufacturing level, assembly work is disrupted as parts are redesigned and tested by the engineering team.

The use of AR can help smooth out the design process by enabling quick visualization of parts and assemblies. Collaboration during design review meetings becomes easier with AR, as all team members can envision 3D models without the need for computer monitors or expensive prototypes [64]. HMD-type devices like the Microsoft HoloLens allow engineers to visualize and control their models during design review meetings. These models can be automatically updated onto the HMD worn by other meeting attendees through wireless or Bluetooth connection. AR also replaces the process of creating numerous physical prototypes or mock-ups, which reduces operating costs and time. This is especially true as 3D printing services can become expensive and time-consuming when outsourced. In the aerospace engineering sector, a high level of detail is often coupled with small tolerance allowances, which forces parts and prototypes to be continuously redesigned and remade. Creating a new prototype for every redesigned part is not only expensive but is also materially wasteful.

In comparison, AR models need only to be updated through CAD software already available to the engineering team. After every design iteration, engineers can update the CAD and re-deploy the model onto the AR device without needing to wait for a prototype to be printed or manufactured. The process can be completed within minutes of a new design iteration.

AR can also be used to enhance understanding of engineering analysis results. For example, finite element analysis (FEA) results can be overlaid onto an object or space to visualize stress concentrations acting on engineered components. This allows better comparison and understanding of how engineering analysis fits into the overall real-world solution [33, 34, 72]. Similarly, computational fluid dynamics (CFD) analysis results can be augmented onto the aircraft to help engineers understand physical properties like temperature, airflow, and air pressure [64]. This can be done for the aircraft interior, where airflow, pressure, and temperature variations are of high importance for engineers. The importance of these parameters is linked to passenger safety and passenger comfort, both of which profoundly affect the success of an aircraft program. For the aircraft exterior, airflow and air pressure values are of particular interest to the engineering team as they provide proper measures of the aircraft's performance.

Results of CFD analysis are traditionally shown through color contours, where different colors indicate different values of pressure, airspeed, or temperature. Understanding how such contours translate into the real world can be challenging, especially since the contours do not include object locations or 3D depth. Contour maps given by CFD analysis software are generally 2D in nature. Although 3D analysis is possible with today's technological abilities, the process is incredibly time-consuming and computationally intensive and financially expensive. As a result, 3D CFD analysis is seldom completed early on in the design process. Instead, 2D analysis is used throughout the design process to help engineers visualize the aircraft's parameters and performance.

Consequently, matching the contour color to the appropriate object within a 3D space is left up to the engineer. This task is cognitively challenging and exhausting for the majority of people. However, with AR, CFD contour results can be visualized directly onto the aircraft, helping engineers better understand their analysis results in 3D space. This reduces the cognitive demands of engineers and ensures faster and more accurate visualization of CFD results compared to both 2D and 3D analyses.

Anecdotal evidence suggests that when using AR, engineers prefer the optical see-through and video see-through HMD to handheld devices. Though both sets of AR devices can be employed in an engineering environment, HMD devices allow for seamless work by freeing up the user's hands. When wearing an HMD, engineers have the freedom to use computers, pen and paper, and other handheld tools. In this way , the AR device is not disruptive to their typical work process or cycle. Also, HMD devices capable of head tracking and spatial mapping allow engineers to place holograms and other augmented data in a full 360-degree environment. Some HMD devices allow the user to anchor holograms to a specific location within the real world. As a result, the hologram is only viewable when the user's head is turned to the particular location to which the hologram was anchored. The benefits of such ability are numerous.

On the one hand, engineers can anchor multiple holograms of different CAD parts or revisions of a specific part. They can then view these holograms separately by anchoring them to entirely different locations or simultaneously by anchoring them near each other. Engineers can also anchor holograms to nearby locations, but not directly in front of their computer monitor, enabling them to view their computers with minimal distraction and confusion. This is also applicable for manufacturing and maintenance tasks, which use AR to enhance operations by guiding workers.

20.3.2 Manufacturing and Maintenance Support

AR applications in manufacturing and maintenance are used to guide aerospace operators to perform a given task with high efficiency [3, 49, 57, 58]. In aerospace manufacturing, task efficiency is measured through product quality, process time, worker effort, and occurrence of errors. For AR devices to truly help manufacturing and maintenance operators, they must enable operators to complete their assigned tasks with the highest possible first-time quality and the lowest possible time, effort, and error. With AR, 2D and 3D instructions or animations can be overlaid onto the real-world area. It provides operators with a simultaneous view of both the working area and the task that must be carried out [22]. Manufacturing and maintenance operators can view manuals, instructions, and engineering drawings simultaneously with the working part or assembly.

Traditional manufacturing, much like engineering, separates the worker from instruction manuals. Work manuals and engineering drawings first existed in paper format. Manufacturing operators would switch between viewing these documents and working on the physical part or product. To reduce the time lost in switching focus, some operators would commit sections of the instruction manuals or engineering drawings to memory. However, human memory is an unreliable and highly limited cognitive function that can lead manufacturing operators to err more frequently. Over time, advances in computer technology allowed aerospace manufacturers to adopt paperless manuals. Instructions, manuals, standards, and drawings are accessible to operators through computer workstations located around the manufacturing workspace. The drawback of using computer workstations is that instructional documents are physically separated from operators, creating higher operator workload and inefficiencies in the manufacturing process. Aerospace manufacturing workspaces take up a considerable area due to the large aircraft that must be assembled within. At the top assembly level, a single workstation holds one aircraft anywhere from a few days to months as work is completed. This workstation often has a single computer through which all engineering drawings, instruction manuals, and other documents are accessible.

Manufacturing operators at the workstation must all share the computer while each of them carries out a different task. Operators must also travel between the aircraft and the computer station to reference or view documentation [67]. Safety precautions require all personnel to walk around the workstation perimeter as not to risk causing damage to the aircraft. Structural parts like the wings and the empennage may protrude from the aircraft fuselage at heights well above that of the average human. Rather than walk around these structures, many personnel choose to save time and effort by ducking below them and following the most direct route. The result is a potential injury for workers and damage to the aircraft structure. When working on interior structures, operators are forced to board and disembark from the aircraft frequently. This practice is dangerous and disruptive to the work cycle. For instance, during the aircraft structure assembly, the inner cabin or cockpit walls and floors are not yet installed. Operators use wooden boards in place of flooring so that they may walk through the aircraft. These boards balance on the lower aircraft stringer and are not secured to the structure, as they are meant to be easily moveable. Operators may accidentally displace aboard, resulting in falls, which can lead to worker injury and aircraft structural damage.

The computer workstation setup is far from efficient, which is why many operators carry printouts of the instructions; hence, the earlier discussed issues related to paper documents arise again. The problems posed by computerized and paper documentation can be solved with AR. AR can be used to provide operators with quick access to instructional documents and engineering drawings by augmenting images or other data in the operator's FOV. With AR, manufacturing operators can view engineering drawings, step-by-step instructions, and live tutorials overlaid in their work area. For example, the location of hidden parts or components can be highlighted to personnel through AR. The information can be obtained from existing CAD models and shown to manufacturing workers through an AR device. Airbus Military completed a comparative study in 2010 titled asseMbly Oriented authOring augmeNted reality, or Project MOON for short. The project created AR-based instructions for technicians routing electrical harnesses in the frame of the A400M aircraft. The instructions were created using information obtained from industrial digital mock-up, calibration markers, and RT video images. The resulting study found an overall preparation and maintenance time reduction of 90% using AR-based instructions compared to conventional methods [74].

One of the best-known projects in this field is Boeing's study on AR for model-based instruction (MBI). In the early 2010s, Boeing teamed up with Iowa State University to study the workload of personnel performing assembly tasks using MBI on conventional devices and AR tablets [67]. Study participants engaged in a wing assembly task through which three types of MBI were provided: a stationary computer, tablet, and tablet with AR. The study found that AR instructions can increase first-time quality and reduce the time taken to complete a task. The study also showed that when using desktop and tablet MBI, a significant amount of time was spent by participants in traveling and confirming information.

Similarly, when using handheld AR, participants required additional time to switch between the AR device and other work tools. Hence, it is anticipated that head-worn AR devices will cut back on the overall time required for task completion by eliminating the need to switch tools and travel to access information. Boeing has since worked on finalizing the Boeing Augmented Reality Kit (BARK) , an AR toolkit that will be used to improve worker efficiency for assembly and repair tasks [4]. The evident benefits of AR led American aerospace, defense, and security company Lockheed Martin to use head-worn AR devices in assembly tasks [38]. The company reports that the use of AR devices has helped cut back on assembly time, which translates to massive cost

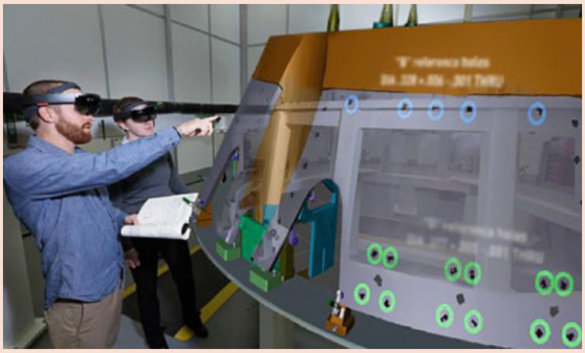


Fig. 20.4 AR images can be overlaid onto working space and parts to give engineers and technicians a better understanding of the overall end product. (Image via [38])

savings. Figure 20.4 shows a snapshot of the Orion spacecraft manufacturing process, where the part being assembled is overlaid with AR images.

In aerospace maintenance support, AR can be used in a fashion similar to manufacturing. Instructions and tutorials can be provided to maintenance operators through handheld or head-worn AR devices. Additionally, AR can highlight key areas on the aircraft, such as areas requiring repairs or check-ups. The technology may also be used to provide operators with visuals on what the aircraft, or parts, should look like when repaired. This allows operators to visually determine whether their repair work has been sufficient. Lockheed Martin also uses AR in maintenance support through collaboration with NGRAIN, a Canadian subsidiary of Universal mCloud company . The collaboration resulted in the development of AR modules to assist engineers and technicians with maintenance and assembly of the F-22 and F-35 fighter jets [45]. The module developed by NGRAIN provides Battle Damage Assessment and Repair (BDAR) to Lockheed Martin, which is then used to document and evaluate damages sustained by aircraft. BDAR enables maintenance operators to record damage information onto a 3D virtual model of the aircraft under repair. The result is a reduction in the time required to complete maintenance tasks through streamlined inspection workflows.

Remote and Tele-maintenance

A more popular use of AR for maintenance support is the remote maintenance application, also called tele-maintenance. AR applications for maintenance support can take the form of assisted reality (AsR), a technology that enables direct communication between technicians located on the field and managers who are not present on site. AsR technology is relatively new but is expected to revolutionize the manufacturing and maintenance scenes tremendously. This technology allows users to view a hands-free screen within their immediate FOV, which makes it ideal for maintenance tasks. The technology takes a hands-free form by using head-worn AR devices. In aerospace, AsR is used in maintenance tasks to provide instructions and communication capabilities to operators. With AsR, information is added to the user's peripheral vision without changing the real-world environment. This means the operator can see the aircraft on which they are working while still accessing information documents.

Remote maintenance also allows operators to communicate directly with engineers or managers through video and audio feed. This is especially advantageous in aerospace as it allows maintenance operators working at various airports or hangars to communicate with engineers who are often located at the airline's headquarters or the original aircraft manufacturer's plant. The technology reduces the total time required for maintenance by removing the need for engineers to travel to the aircraft to inspect damages. As such, aircraft repairs can be finished faster, saving airlines a tremendous amount of money by enabling them to put their aircraft back in service. With AsR, engineers can inspect the damage done to aircraft through video feed connection established between the maintenance operator and themselves. They can then provide live instructions to the maintenance operator while still viewing the work being done. Maintenance operators also gain the ability to ask questions, highlight additional areas of damage or interest to the engineer, and discuss viable solutions.

Australian aerospace company, TAE Aerospace, used AsR to send camera feed obtained from a technician's headset to an engineer who was not located in the same area [76]. The engineer would monitor the camera feed and then provide instructions that appeared as augmented markings overlaid onto the technician's AR display. The technology was named FountX AsR and was used to assist technicians in repairing aircraft engines. Technicians were able to obtain real-time information hands-free, using a voice-activated screen placed within their immediate FOV [9]. The information placed on the screen was controlled by an engineer who remained at the office for the duration of the repair. TAE Aerospace found that FountX AsR reduced the time required for maintenance operations by removing the need for engineers to travel to the worksite to aid technicians.

Similarly, Japan Airlines used Google Glass in 2014 to assist maintenance operations at their Honolulu station [7]. The airline used AR to share advice and instructions between maintenance staff and colleagues off-site, much like the FountX AsR did. The instructions could be provided through photographs, videos, or real-time information. Following the discontinuation of the Google Glass, Japan Airlines began using the Microsoft HoloLens instead. The airline further used the HoloLens to train new engine mechanics and flight crew members, using AR in place of the traditional cockpit and engine component printouts, allowing trainees to view, explore, and work on aircraft parts in 3D holographic space.

20.4 Applications in Space Operations

Space operations require astronauts to carry out simultaneous tasks that may be physically or cognitively complex. These tasks can involve thousands of steps, tools, and components, posing various limitations on the achievability of the operation [19]. Space operations are further restricted by limitations on technological resources, monetary funds, and human performance. AR is used to support spacecraft engineering and space operations by guiding human operators through work processes. Work tasks like manufacturing, assembling, and maintaining spacecraft can be well supported with AR in a fashion similar to that discussed in Sects. 20.3.1 and 20.3.2. The European Space Agency (ESA) launched a study in 2015–2016 on Augmented Reality for Product (or Quality) Assurance (ARPASS) [3]. The study evaluated AR usage in the assurance of space missions and found that AR can significantly reduce errors and increase compliance. ARPASS was not implemented at large scale due to its high customization cost and lack of interoperability. However, AR is currently used to assist astronauts with tasks on the International Space Station (ISS). On the ISS, astronauts perform various missions, including payload-specific tasks and routine maintenance tasks.

ISS operations are guided by strict procedures to ensure personnel safety and mission success [36]. These procedures are available to astronauts through operation data files loaded and viewed on the commanding crew station. As astronauts carry out payload-related tasks, they retrieve information by shifting focus between the commanding crew station and the payload. As discussed in Sect. 20.3.2, shifting focus between the instruction manual and the object on which work is done is time-consuming and cognitively demanding and can lead to a higher workload for humans. In space operations, astronauts experience higher workload, loss of intended focus, and loss of concentration when switching between the instructions provided on the commanding crew station and the payload tasks on which they are working [36]. Much like its use in aviation, manufacturing, engineering, and ATM, AR can be used to enhance space operations by integrating data into the FOV of astronauts. Research projects have shown the potential for AR to enhance space operation performance through decreasing workload. A study was conducted by Markov-Vetter et al. [36], highlighting the benefits of AR in assisting astronauts with payload activities on the ISS. The study used an HMD AR device with an optical inside-out tracking ability to create the Mobile Augmented Reality for Space Operation Procedures (MARSOP) prototype. MARSOP augments instructions found in operation data files onto the FOV of astronauts through AR capabilities. A pilot study completed at the ESA's European Astronaut Centre showed that expert users preferred the MARSOP augmented instructions to conventional methods.

The wearable augmented reality (WEAR) application is another example of AR use for space mission support. WEAR was an early AR project created to assist astronauts in performing tasks on board the ISS [15, 16]. The project uses an HMD assembled from off-the-shelf components to create an AR display over one eye [19]. WEAR superimposes 3D graphics and data onto the FOV of astronauts. The device is controlled through voice commands and is capable of tracking location and recognizing objects. ESA uses the WEAR system to give step-by-step instructions to astronauts in place of paper manuals. Using paper manuals, astronauts are required to consistently switch between the instruction manual, handheld tools, and work tasks. To avoid the increased workload associated with this shift in focus, WEAR augments instructions and provides details about the object which the astronaut is viewing. This enables astronauts to work seamlessly without paper manuals or the requirement to switch focus. WEAR further tracks the user's location from CAD models stored in the onboard mobile computer [19].

Similarly, NASA's Jet Propulsion Laboratory (NASA JPL) launched Project Sidekick with the California Institute of Technology, while ESA introduced the Mobile Procedure Viewer (mobiPV), both of which are meant to support astronaut tasks through AR technology [11, 14]. Project Sidekick was launched into space in December 2015 to give visual aid to astronauts performing maintenance tasks [43]. The project uses the Microsoft HoloLens to augment an interactive user manual experience to astronauts in space. The project operates in two modes. The first is the remote expert mode, which uses Skype to connect astronauts with ground operators. The ground operator can view what the astronaut sees on the ISS, provide real-time guidance, and draw augmented annotations onto the astronaut's FOV. The annotations are useful for guiding astronauts through maintenance processes, which they may not have expertise in. The remote expert mode functions similarly to remote maintenance technology. The second mode of operation is procedure mode, which gives astronauts stand-alone procedures with animations, holographs, and illustrations augmented onto the HoloLens. Procedure mode provides excellent support to astronauts, especially since communication between in-space vehicles like the ISS and mission control centers on the Earth is prolonged. With Project Sidekick, astronauts can perform maintenance tasks on the ISS even when engineers and other experts are offline. Procedure mode makes maintenance expertise available to astronauts, without additional training [8].

Training costs for astronauts are substantial in value, which places great importance on the development of newer technology. Currently, all astronauts are required to hold multiple specialties in different domains. This creates a multidisciplinary team capable of tackling different conflicts that can arise while in space. The complexity and cost of launching a payload into space make it impractical to send objects, or other humans, to the ISS to provide help whenever needed. Astronauts must rely on themselves to solve problems within the ISS. These problems can include mechanical, electrical, or structural issues found within the ISS or its equipment. For logical reasons, not all astronauts on the ISS have engineering or maintenance backgrounds; however, astronauts may need to perform maintenance tasks while on the ISS.

AR devices like WEAR and Sidekick are extremely beneficial, as they allow all members of the ISS team to perform maintenance tasks without needing to be trained for engineering operations. Similarly, astronauts may be required to conduct emergency medical tasks while on board the ISS. While astronaut training covers first aid procedures, it does not cover high-risk examinations or surgeries. In critical situations, astronauts may be required to perform medical procedures without access to a full team of healthcare providers. In anticipation of such cases, ESA established the Computer Aided Medical Diagnostics and Surgery System (CADMASS) project, which aimed at facilitating high-risk medical examinations and surgeries conducted by astronauts in space [44]. During long-term spaceflight, astronauts are likely to require medical expertise that is often not available within the mission team. The team may not be able to wait until the next launch or return mission to access medical expertise, which leaves them responsible for carrying out a medical task. With the help of CADMASS, astronauts without medical expertise can perform medical operations and surgeries. This is done by providing astronauts in space with augmented and virtual instructions, simulations, and telecommunications with medical doctors located on Earth. The device is similar to remote maintenance applications but provides medical instructions and communications with healthcare providers.

AR can also be used to assist with spacecraft engineering, manufacturing, and assembly. For instance, NASA JPL created an application called ProtoSpace, which allows its engineering team to avoid design integration conflicts by examining full-scale CAD models of spacecraft and their components in AR [47]. In 2015, NASA JPL developed Project OnSight, an application that helped scientists and engineers of space institutes visualize and study planet surfaces [5]. Project OnSight used the Microsoft HoloLens to view feeds provided from the Curiosity Rover. Data obtained from the Curiosity Rover was used to simulate Mars' surface, allowing engineers and scientists to study the planet's surface while remaining on Earth. The project further enables scientists to "meet" at various simulated locations to discuss rover missions and actions, as shown in Fig. 20.5. OnSight is a breakthrough in space exploration, as it enables humans to view and experience a full 360-degree setting of Mars. Humans can experience other planets by sending technology that is better capable of surviving extreme climates than humans. This solves some of the logistic issues associated with sending humans to space , including the financial burden of readying a space mission and training astronauts. It also allows humans to explore the surface of other planets without leaving the comfort of Earth or placing themselves under the risk of developing spaceflight-related medical issues.

Spacecraft lifespan and quality are affected by the electromagnetic compatibility of onboard equipment [13]. All electronic equipment found on a given spacecraft must be mutually compatible to ensure both safety and functional success. For electromagnetic compatibility goals to be achieved, all equipment must have the same reference potential. This can be done by establishing a conductive network over the spacecraft. When the network is connected to all electrical devices' grounding terminals, a consistent reference potential is realized. The conductive network is made up of many overlapping copper foil tapes. Due to the high density of onboard devices and tight space available in spacecraft, it depends on manual operation. When establishing electromagnetic compatibility of equipment, human personnel must rely on their subjective consciousness to lay, bond, and lap treat copper tapes [13]. This places great importance on worker experience , resulting in a process with low efficiency, high error, and high learning costs. AR devices can be used to supplement gaps in human knowledge or experience by providing personnel with new methods of solving conductive network problems. Chen et al. [13] presented a concept using the Microsoft HoloLens and virtual-real fusion to assist personnel in conducting network assembly. The project guides assembly personnel, hence solving problems associated with the traditional process. In this application, AR can decrease errors and improve the accuracy, quality, and efficiency of work.

20.5 Future Application Concepts

The recent popularization of AR has led the aerospace industry to pursue various application concepts to create better, safer, and more accessible user experiences. This section presents the most promising future applications of AR technology in the aerospace industry. Figure 20.6 provides a breakdown of these applications and lists some examples of projects. These applications are found mainly in security, crew support, and IFEC. Also, the efforts to enhance AR with the use of AI and haptic devices are discussed in this section.

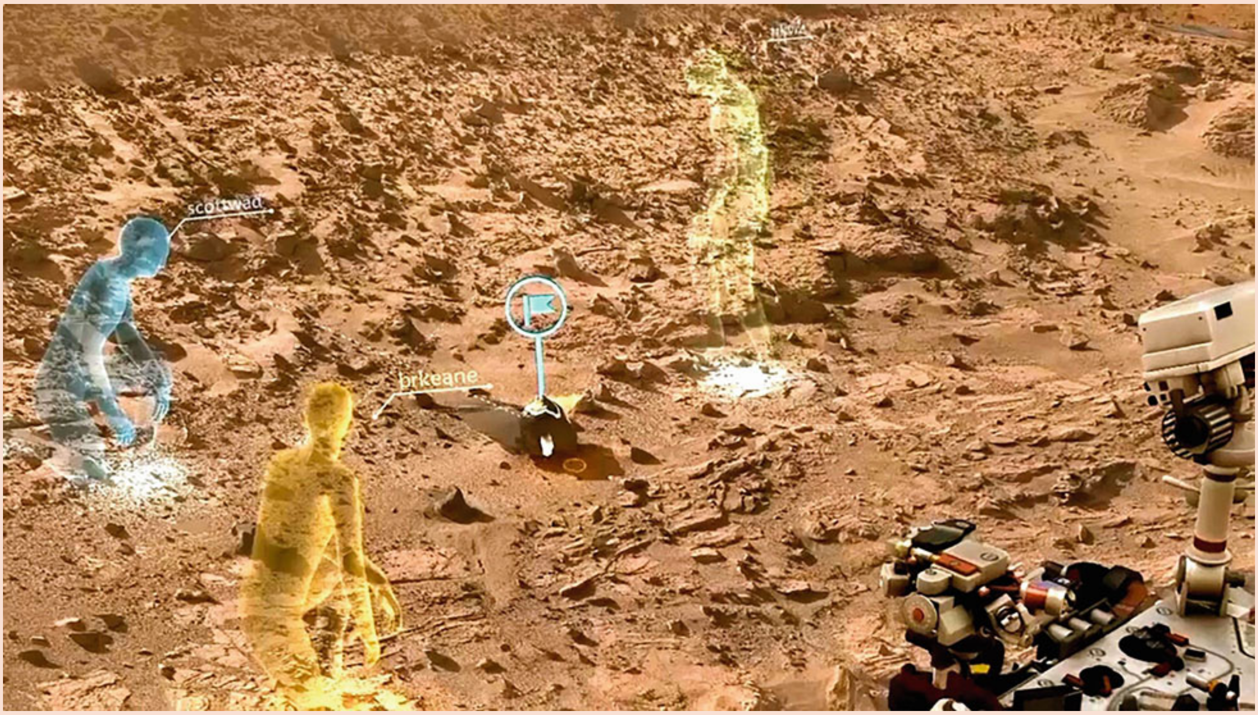


Fig. 20.5 Screen views from Project OnSight displaying 3D Martian environment simulation developed using rover data and mission scientists at the scene. (Image © NASA JPL – Caltech)

A summary of COVID-19 support applications is also presented due to the relevance of this topic to the modern-day industry and society.

20.5.1 Airport Security

A new AR application concept for the aerospace industry is that of airport security . It is thought that AR can assist with security and surveillance by monitoring airport operations, baggage handling, and overall airport settings. Head-worn devices like smart glasses can be used by airport staff to survey and monitor the surrounding personnel or baggage. For example, AR headwear can be used to visualize X-ray scans of luggage in airport security checks. Integrating security cameras, trackers, sensors, and AR devices can allow airport staff to pinpoint areas of potential hazards or security breaches [60]. Enhanced facial recognition, video analytics, and threat recognition algorithms are necessary for the success of AR in airports' security. One concept is to use AR to scan all individuals in an airport facility; record their behavioral, gestural, and kinematic characteristics; and then analyze those traits to determine a threat profile. The threat profile and passenger information of individuals will be sent to a security agent.

20.5.2 Crew Task Support

Air New Zealand introduced an AR-based flight attendant assistance application that runs on the Microsoft HoloLens [70]. The application is named Inflight and is used to provide flight attendants with important passenger information. At this time, the application remains in development. However, it is said to allow aircraft crew members to work at the galley, recognize passengers and their crucial information, access flight information, and communicate. Critical passenger information such as mood, food allergies, and destination are shown on the demonstration video published by Air New Zealand.

Helsinki Airport and SITA Labs collaborated to reproduce an airport operational control center with AR devices capable of projecting airport operation data [70]. This data can be displayed in alphanumeric or visual forms such as animation, maps, and charts. Similarly, Singapore Changi Airport and SATS ramp handling designed an AR application to help baggage handlers determine the unit number, weight, and loading sequence. They allocated the in-the-aircraft position of luggage or cargo pieces. The application uses the Vuzix M300 AR Smart Glasses and a module named EON Reality AR Assist, which runs in coordination with AI and the Internet of Things.

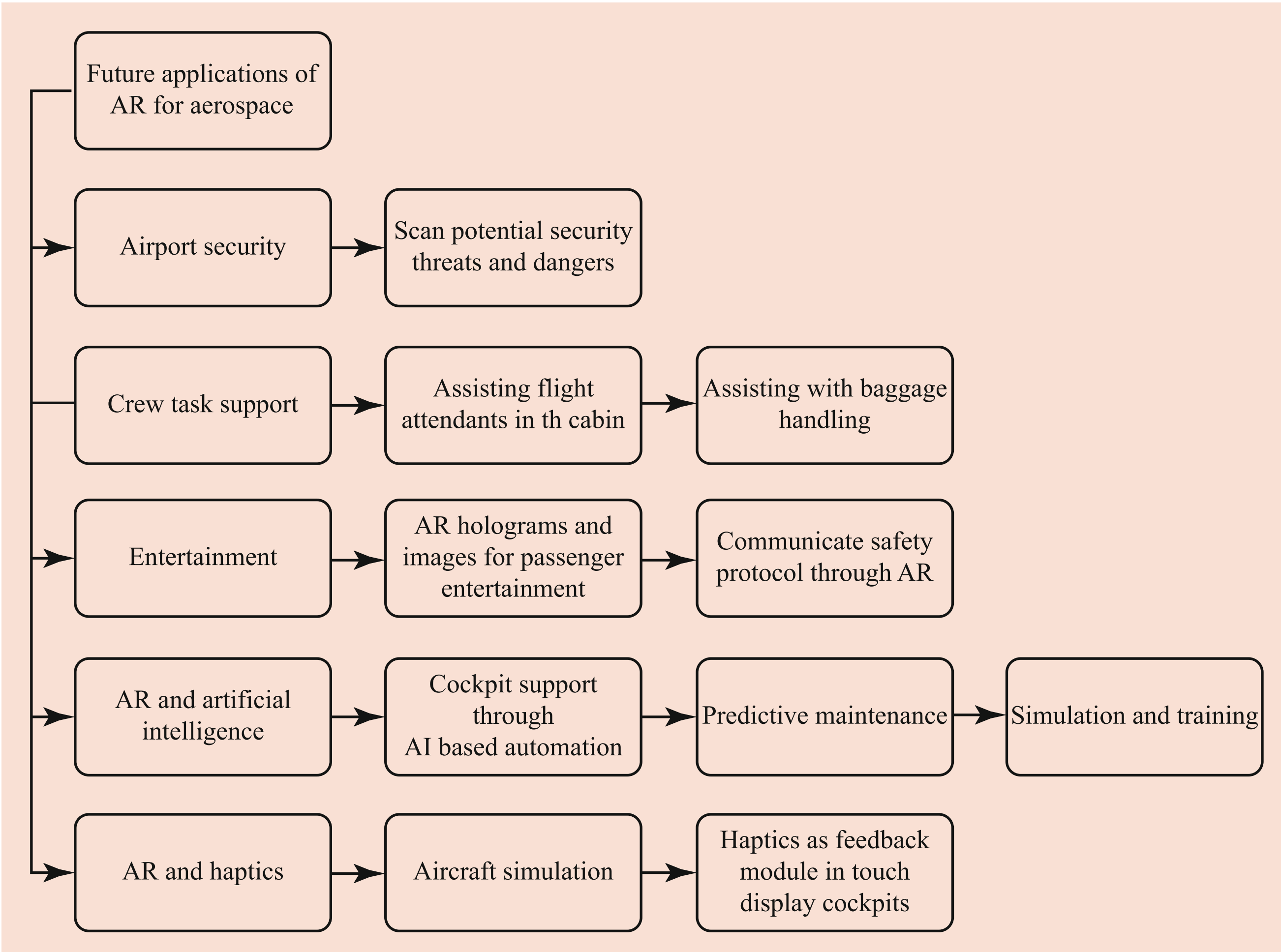


Fig. 20.6 Breakdown of future AR applications for the aerospace sector

In 2018, LATAM Airlines Group announced the upcoming release of its AR phone application to measure carry-on luggage [79]. The application will assist airline passengers in determining whether their luggage is within the allowable carry-on dimensions. Royal Dutch Airline, also known as KLM, launched an almost-identical application within their iOS app. This application transposes a virtual suitcase of carry-on dimensions onto the user's suitcase to allow users to envision whether their luggage will be accepted upon boarding.

20.5.3 In-Flight Entertainment and Communication

AR glasses and similar wearables can be used to provide IFEC in the form of 2D and 3D holograms [70]. Images, videos, and augmented animations can give passengers essential safety instructions such as fasten "seatbelts" cues or entertainment modules. AR devices can also be used to allow passenger interaction with the passenger service unit (PSU) and communication with the flight crew. A HoloLens application initiated by Boeing and the University of Cincinnati, named Virtual Assistant: Boeing Onboard, enables passengers to browse information, entertainment, and the Internet while in the cabin [42]. A similar application is Acti-Vision smart interactive window, launched by Vision Systems. This application uses a transparent screen with AR capability to augment the cabin window with interactive maps, meal plans, flight information, news, and games.

20.5.4 Augmented Reality and Artificial Intelligence

The growth of AI technology into a general-purpose utility opens pathways for numerous applications within the aerospace industry. AI can reduce costs of aerospace engineering programs by assisting in various segments of the design cycle, including optimization, simulation, prototyping, maintenance, and manufacturing. AI use within aerospace is limited at this time; however, companies such as Boeing and

Airbus are integrating AI into various aspects. Over the next few decades, AI has become a commonly used tool in the aerospace and aviation industries. Contrary to media presentation, AI models do not inherently remove human operators from the loop. They can be employed to assist and ensure safer, faster, and more accurate human performance [6, 21].

Merging AR with AI can open many new applications in aerospace by supporting routine or difficult tasks. Most uses of AI in aerospace can be improved when combined with AR displays. Areas such as flight performance, supply chain management, generative design, aircraft maintenance, and quality control can benefit from both AI integration and AR technology [6]. One concept that uses AI to support flight performance was presented by the US Air Force. The plan is to incorporate AI into the cockpit of the F-35 fighter aircraft to help pilots complete the ever-growing list of tasks. Future fighter jet concepts require pilots to aviate as they control a host of UAVs. The use of AI and AR in such scenarios becomes critical, as these modules can assist pilots in safely completing the required tasks [51]. For instance, AI can assist in decision-making, pilot-UAV communication, and aircraft data monitoring. Meanwhile, AR can be used to display critical information and feedback to the pilot.

A rising AI application for the aerospace sector is predictive maintenance, which uses AI to analyze and predict repair times for various aerospace vehicles. The module inputs a large amount of vehicle data, which is time-consuming for a human to look through or analyze. The AI module output is in the shape of data points and meaningful insights [6]. These are presented to a human operator, who carries on the maintenance and repair jobs. The results of the predictive maintenance module can be displayed to the worker in AR mode to maximize worker efficiency. As discussed in Sect. 20.3.2, Boeing's studies have shown that AR enhances worker efficiency in maintenance and manufacturing tasks. Using a handheld or head-worn AR displays can help the worker visualize the results of the AI analysis relative to the aircraft or physical part. AR and AI have also been employed in numerous simulation and training exercises in the aerospace industry, though they have often been used separately. AI allows for real-time simulations with complex data collection and analysis. The simulation environment can be changed to introduce unexpected aircraft errors, weather phenomena, or situational changes using AI. When coupled with AR or virtual reality, the simulation can be quickly and cheaply displayed. Changes to the simulated environment can be done in a short amount of time, often at no extra cost. For example, to change the layout of cockpit controls in AR mode, one simply needs to update the program, changing the virtual components and then reuploading the new layout onto the device. A similar change on a physical simulator would require a lot of work and money, since the entire panels may need to be replaced.

AR can also replace or support tablet-based or printed pilot operating manuals on board aircraft. Traditional aircraft practice requires the pilot to search manually through the documentation for correct recovery procedure upon recognition of a fault. A computer algorithm utilizing AI can recognize or anticipate faults and system failures experienced by the aircraft, search through a database of recommended practices or an operating manual, and then display a step-by-step procedure to the pilot onto a transparent screen mounted within the pilot's FOV. In this scenario, the use of AR is essential to increase pilot SA by reducing the need to switch between the instruction display and cockpit controls. This reduces workload and stress and speeds up information access during critical situations. It also enables the AI system to provide visual cues to point the pilot's gaze toward critical instruments or controls. For instance, a virtual arrow can be displayed onto the AR screen to point toward the throttle when power should be cut back.

20.5.5 Augmented Reality and Haptic Feedback

Haptic interfaces are actuated computer-controlled devices that allow humans to touch and manipulate objects within a virtual or real environment [71]. The technology creates bilateral interactions between the operator and the virtual environment, hence providing both input and feedback. Haptic technology is already being used in automotive and entertainment applications but has not yet been implemented within the aerospace industry [26]. However, the technology shows much promise for aerospace applications such as simulation. By coupling haptic technology with AR, reconfigurable and reprogrammable simulators can be created, reducing the cost of full-scale flight simulators which are expensive to manufacture, maintain, and operate [10, 26]. Current full-scale flight simulators are limited to one aircraft type, which forces airlines and flight schools to purchase simulators for different types of aircraft. With AR and haptics, a single simulator is sufficient. The cockpit interface can be displayed visually onto an AR display, while haptic technology can be used to simulate a real-life flying experience.

At this time, haptic technology is being studied for different aerospace uses. One example is the haptic display for flight envelope protection systems introduced by Ellerbroek et al. [17]. The display was created to design more human-centered automation, as haptic feedback is known to allow for flexible sharing of information and control [17]. The technology showed promise in addressing the lack of SA issues regarding flight envelope limitations as it captured the study participants' attention effectively. However, the participants could not successfully utilize the information they received from the haptic display, leading to questions regarding the suitability of the haptic display in this setting. Further research is needed in order to optimize haptic interface use for aerospace. The technology is not yet ready to be implemented on a large scale. However, it is expected to be used within future generation aircraft, which will implement touch-display or full AR display cockpits.

20.5.6 Augmented Reality Applications for COVID-19 Support

In light of the COVID-19 pandemic outbreak, the aerospace industry has searched for methods of using AR to screen and sanitize airports and aircraft for possible contamination. The novel coronavirus spread was linked mainly to international travel using air transport. In response, the aerospace sector is developing methods of using AR devices like the Microsoft HoloLens to scan for symptoms in travellers and contamination of surfaces. Various concepts that utilize AR devices along with sensors have emerged. While none of these applications is ready for large-scale production at this time, research efforts continue to create viable solutions that can be useful not only for the COVID-19 outbreak but also for future pandemic cases. AR is also being used throughout the COVID-19 outbreak to support medical professionals, provide remote assistance, and enable no-touch interaction through virtual interfaces [69]. AR is providing individuals and companies worldwide with more accessible and safer services, all while maintaining physical distancing protocols required to curb the spread of the virus [69].

20.6 Concluding Remarks

AR is becoming an increasingly essential tool in aerospace and aviation. Global forecasts predict a tremendous increase in the overall number of passengers traveling on board aircraft by the mid-twenty-first century. To meet such demand, an increase in speed, efficiency, and reliability is needed across the entirety of the aerospace and aviation sector. The aerospace industry will need to train a large number of professionals quickly to keep up with the rising demand for its services. It will also need efficient day-to-day operations in engineering, manufacturing, maintenance, and aviation. As a result, AR can be a useful tool for aerospace and aviation employees and professionals.

With the advent of urban air mobility and personal air vehicle concepts, the aircraft cockpit will require significant changes. Aircraft cockpits are currently cluttered with high-performance instruments that are only easily understood by individuals with extensive training and piloting experience. Navigation devices like the HUD and HDM use symbology that is not familiar to the everyday individual. However, personal air vehicle success depends on the ability of everyday individuals to operate an aircraft successfully. This can only be done through rigorous redesign of current aircraft cockpits. AR can assist designers by allowing rapid prototyping of cockpit layouts, with interactive capabilities for easier testing and verification. This can cut back on prototyping time and cost while allowing end-user involvement in earlier phases of the design process. Prototypes can be made available through devices with cameras or AR support.

AR can also help future pilots and personal air vehicle users train and fly their aircraft through augmented cockpits. For example, smart screens can be used within the cockpit to augment different segments of flight. WayRay showcased a smart screen capable of imitating HUD functionality in 2019. Their smart screen provides larger FOVs and is being readied to replace automotive windshields. Aircraft manufacturers can use similar concepts in place of current cockpit windows and HUD combinations, integrating AR devices into the aircraft structure. A substantial drawback of both HUD and HMD is the weight associated with both devices. HUD weight affects overall aircraft weight. While this is not a prominent issue on large aircraft like the Airbus A320 or the Boeing B787, it becomes a significant concern on smaller, general aviation aircraft like the Cessna 172. Smart transparent screens in place of a traditional HUD and windshield combination may reduce the number of parts installed on the aircraft, which can translate to cost or weight reductions, allowing AR to become standardized equipment on board smaller aircraft. It also readies the aircraft for a wide range of AR application support that can be beneficial to the pilot or user. It introduces a wider FOV and a more substantial area with augmentation capability.

AR has growth potential in the field of aerospace and aviation. Numerous projects, use cases, and research efforts over the last century have enabled AR to be used in and around aerospace in a multitude of manners. The recent development in computer systems, hardware and software, has furthered AR's ability to be used in this field. Aerospace and aviation are largely restricted in the use and introduction of new technology due to high standards, a necessity to ensure the worldwide safety of personnel, passengers, and crew. As a result, the development of technological advances in aerospace is seemingly slow compared to other industries, where new systems are introduced daily. Years of research and validation are necessary before any system, AR or otherwise, is allowed on board an aircraft in day-to-day operations. The applications introduced in this chapter, both in use and futuristic, remain open to criticism as the industry tries to better itself in all aspects, including safety. Rigorous research is needed on any applications to be used in aerospace and aviation, making the collaboration of academic, legal, and industrial members critical for continued growth and advancement of this industry.

References

1. 360World: https://360.world/clairity/ (2017)
2. Arthur III, J., Bailey, R., Williams, S., Prinzel III, L., Shelton, K., Jones, S., Houston, V.: Review of head-worn displays for the next generation air transportation system. J. Opt. Eng. **55**(5), 051405 (2017)
3. Augmented Reality for Enterprise Alliance: AREA webinar augmented reality in the aerospace industry. https://www.youtube.com/watch?v=u8rPFMGzOrw (2016)
4. Augmented World Expo AWE: Paul Davies (Boeing): BARK–the Boeing augmented reality kit. https://www.youtube.com/watch?v=zVsjjzWVEkw (2017)
5. Aukstakalnis, S.: Practical Augmented Reality A Guide to the Technologies, Applications, and Human Factors for AR and VR. Pearson Education Inc., Crafordsville (2017)
6. AXISCADES: Machine learning & artificial intelligence in aerospace industry. https://www.axiscades.com/blog-resources/whitepaper/Aerospace-whitepaper.pdf (2019)
7. Bellamy III, W.: 9 Companies using augmented reality and virtual reality in aviation. Avionics International. http://www.aviationtoday.com/2017/08/24/9-companies-using-augmented-virtual-reality-aviation/ (2017)
8. Blakemore, E.: There's augmented reality in space now. https://theweek.com/articles/607794/theres-augmented-reality-space-now (2016)
9. Blyth, C.: Immersive technologies and language learning. Foreign Language Annals. **51**(1), 225–232 (2018)
10. Borro, D., Savall, J., Amundarain, A., Gil, J., Gracia-Alonso, A., Matey, L.: A large haptic device for aircraft engine maintainability. IEEE Comput. Graph. Appl. **24**(6), 70–74 (2004)
11. Boyd, A., Fortunato, A., Wolff, M., Oliveira, D.: MobiPV A New, Wearable Real Time Collaboration Software for Astronauts Using Mobile Computing Solutions, SpaceOps Conference. AIAA, Daejeon (2016)
12. Caduell, T., Mizell, D.: Augmented reality an application of heads-up display technology to manual manufacturing processes. In: IEE Proceedings of the International Conference on System Sciences, pp. 659–669. Kauai (1992)
13. Chen, H., Sun, G., Chen, C., Wan, B.: Augmented reality based assembly guidance for spacecraft conductive network. In: IOP Conference Series: Earth and Environmental Science 252, Xi'an (2019)
14. Chintamani, K., Lierde, B., Maloney, S., Kiernan, P.: Wearable crew support technology on the international space station: the mobile procedure viewer (mobiPV). In: Proceedings of the Human Factors and Ergonomics Society, pp. 1–11 (2014)
15. Di Capua, M.: Augmented reality for space applications. Ph.D. Thesis, University of Maryland, College Park (2008)
16. Di Capua, M., Akin, D., Brannan, J.: MX-3 design and technology development of an advanced all-soft high mobility planetary space suit. In: International Conference on Environmental Systems. AIAA, Reston (2010)
17. Ellerbroek, J., Rodriguez y Martin, M., van Paassen, M., Mulder, M.: Design and Evaluation of a Haptic Display for Flight Envelope Protection Systems, Int Symp Aviat Psychol 416–421 (2015)
18. Endsley, M.: Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors Society, pp. 97–101. Santa Monica (1988)
19. ESA: Augmented reality to help astronauts make sense of space. https://www.esa.int/Enabling_Support/Space_Engineering_Technology/Augmented_reality_to_help_astronauts_make_sense_of_space (2009)
20. FHL Vive Center for Enhanced Reality: Immersive Semi-Autonomous Aerial Command System (ISAACS). https://vivecenter.berkeley.edu/research1/isaacs/ (2020)
21. Ford, M.: Martin Ford. In: Atitkar, R. (ed.) Architects of Intelligence, p. 11. Packt Publishing, Birmingham (2018)
22. Frigo, M., Silva, E., Barbosa, G.: Augmented reality in aerospace manufacturing. J. Ind. Intell. Inf. **4**(2), 125–130 (2016)
23. Fulbrook, J., Ruffner, J., Labbe, R.: An augmented reality binocular system (ARBS) for air traffic controllers. In: Proceedings of SPIE Defense and Security, Orlando (2008)
24. Furht, B.: Handbook of Augmented Reality. Springer, New York (2011)
25. Furness III, T.: The super cockpit and its human factors challenges. Proc. Hum. Factor. Ergon. Soc. **30**(1), 48–52 (1986)
26. Girdler, A., Georgiou, O.: Mid-Air Haptics in Aviation – creating the sensation of touch where there is nothing but thin air. ArXiv. **abs/2001.01445** (2020)
27. Gorbunov, A.: Stereoscopic augmented reality in visual interface for flight control. J. Aerosp. Sci. Technol. **38**, 116–123 (2014)
28. Gursky, B., Olsman, W., Peinecke, N.: Development of a tunnel-in-the-sky display for helicopter noise abatement procedures. CEAS Aeronaut. J. **5**(2), 199–208 (2014)
29. Hicks, J., Jessee, M., Durbin, D.: A summary of visual gaze and dwell times for army pilots during mission simulations. In: Technical Report. Army Research Laboratory, Aberdeen Proving Ground (2012)
30. Hilburn, B.: Head-down time in aerodrome operations a scope study. In: Technical Report. Centre for Human Performance Research, The Hague (2004)
31. Hilburn, B.: Head-down time in ATC tower operations real time simulations results. In: Technical Report. Centre for Human Performance Research, The Hague (2004)
32. Hofmann, T., König, C., Bruder, R., Bergner, J.: How to reduce workload – augmented reality to ease the work of air traffic controllers. Work. **41**(1), 1168–1173 (2012)
33. Huang, J., Ong, S., Nee, A.: Real-time finite element structural analysis in augmented reality. J. Adv. Eng. Softw. **87**, 43–56 (2015)
34. Huang, J., Ong, S., Nee, A.: Visualization and interaction of finite element analysis in augmented reality. Comput. Aided Des. **84**, 1–14 (2017)
35. Lim, Y., Gardi, A., Sabatini, R., Ramasamy, S., Kistan, T., Ezer, N., Bolia, R.: Avionics human-machine interfaces and interactions for manned and unmanned aircraft. Prog. Aerosp. Sci. **102**, 1–46 (2018)
36. Markov-Vetter, D., Millberg, J., Staadt, O.: Mobile augmented reality for space operation procedures: a generic approach of authoring and guiding on-board payload activities. In: 64th International Astronautical Congress, Beijing (2013)
37. Masotti, N., Persiani, F.: On the history and prospects of three-dimensional human-computer interfaces for the provision of air traffic control services. CEAS Aeronaut. J. (2016). https://doi.org/10.1007/s13272-016-0185-4
38. Maw, I.: How Lockheed Martin is using augmented reality in aerospace manufacturing. http://new.engineering.com/story/how-lockheed-martin-is-using-augmented-reality-in-aerospace-manufacturing (2019)
39. McClung, S.N., Kang, Z.: Characterization of visual scanning patterns in air traffic control. Comput. Intell. Neurosci. **2016** (2016). https://doi.org/10.1155/2016/8343842
40. Merheb, A.-R., Noura, H., Bateman, F.: Emergency control of AR drone quadroter UAV suffering a total loss of one rotor. IEEE/ASME Trans. Mechatron. **22**(2), 961–971 (2017)
41. Mizell, D.: Augmented reality applications in aerospace. J. Technol. Stud. **40**(2), 96–107 (2000)
42. Morrison, M.: Inside stories from the crystal cabin finalists. Flight AIXDaily News (2018)
43. NASA: NASA, Microsoft collaborate to bring science fiction to science fact. https://www.nasa.gov/press-release/nasa-microsoft-collaborate-to-bring-science-fiction-to-science-fact (2015)

44. Nevatia, Y., Chintamani, K., Meyer, T., Blum, T., Runge, A., Fritz, N.: Computer aided medical diagnosis and surgery system: towards automated medical diagnosis for long term space mission. In: Advances in Space Technology Robotics Autonomy. ESA, Neuchâtel (2011)
45. NGRAIN: https://www.ngrain.com/lockheed-martin/ (2017)
46. Nicholl, R.: Airline head-up display systems human factors considerations. J. Econ. Manag. Sci. (2015). https://doi.org/10.4172/2162-6359.1000248
47. Noor, A. K.: The HoloLens Revolution. Mechanical Engineering. **138**(19), 20–25 (2016)
48. Norman, D.: The Design of Everyday Things. Doubleday/Currency, New York (1990)
49. Novak-Marcincin, J.: Augmented reality aided manufacturing. In: International Conference on Virtual and Augmented Reality in Education, pp. 23–31. Procedia Computer Science, Tenerife (2013)
50. Office of the Privacy Commissioner of Canada (OPC): Drones in Canada. https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2013/drones_201303/#fn25 (2013)
51. Osborn, K.: Artificial intelligence is going to make America's F-35 even stronger. https://nationalinterest.org/blog/buzz/artificial-intelligence-going-make-americas-f-35-and-b-2-even-stronger-104967 (2019)
52. Osborne, T.: Augmented reality is key to future fighter cockpits. https://aviationweek.com/defense/augmented-reality-key-future-fighter-cockpits (2018)
53. Palladino, T.: Edgybees extended drone prix AR game to Epson Moverio smartglasses. https://next.reality.news/news/edgybees-extend-drone-prix-ar-game-epson-moverio-smartglasses-0177817/ (2017)
54. Palladino, T.: Epson launches AR app to control DJI drones via Moverio smartglasses. https://next.reality.news/news/epson-launches-ar-app-control-dji-drones-via-moverio-smartglasses-0186770// (2017)
55. Palladino, T.: UC Berkeley team commands drone fleet with HoloLens, next reality. https://hololens.reality.news/news/uc-berkeley-team-commands-drone-fleet- with-hololens-0176671/ (2017)
56. Palladino, T.: Epson's AR drone simulator lets you get your virtual wings before crashing the real thing. https://next.reality.news/news/epsons-ar-drone-simulator-lets-you-get-your-virtual-wings-before-crashing-real-thing-0180997/ (2018)
57. Palmarini, R., Erkoyuncu, J., Roy, R.: An innovative process to select augmented reality (AR) technology for maintenance. In: International Conference on Through-life Engineering Services, pp. 23–28. Procedia CIRP, Cranfield (2017)
58. Palmarini, R., Erkoyuncu, J., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. In: Robotics and Computer-Integrated Manufacturing, pp. 215–228 (2018)
59. Park, J., Kim, H.: Research on GUI (graphic user interface) design styles in mixed reality service – focusing on the way to distinguish between reality and virtual object. J. Korean Soc. Des. C. **23**(2), 271–282 (2017)
60. Peddie, J.: Augmented Reality Where We Will All Live. Springer, Gewerbestrasse (2017)
61. Peinecke, N., Schmerwitz, N., Döhler, H.: Review of conformal displays more than a highway in the sky. Opt. Eng. **56**(5), 051406 (2017)
62. Petrescu, R., Aversa, R., Akash, B., Corchado, J., Berto, F., Mirsayar, M., Petrescu, F.: About bell helicopter FCX-001 concept aircraft – a short review. J. Aircr. Spacecr. Technol. **1**(2), 91–96 (2017)
63. Pinska, E., Tijus, C.: Augmented reality technology for control tower analysis of applicability based on the field study, CEAS Euro. In: Air Space Conference, pp. 573–580. Berlin (2007)
64. Regenbrecht, H., Baratoff, G., Wilke, W.: Augmented reality projects in automotive and aerospace industry. IEEE Comput. Graph. Appl. **25**(6), 48–56 (2005)
65. Reisman, R., Ellis, S.: Air Traffic Control Tower Augmented Reality Field Study, N SIGGRAPH '05 ACM SIGGRAPH 2005 Posters. ACM, New York (2005)
66. Reisman, R., Feiner, S., Brown, D., Augmented reality tower technology flight test. In: Proceedings of the Human-Computer Interaction in Aeronautics, San Jose (2014)
67. Richardson, T., Gilbert, S., Holub, J., Thompson, F., MacAllister, A.: Fusing self-reported and sensor data from Mixed-Reality training. In: Industrial and Manufacturing Systems Engineering Conference Proceeding, pp. 1–12. Orlando 2014)
68. Roscoe, A.: Workload in the glass cockpit increasing. Flight Saf. Dig. **1992**, 1–20 (1992)
69. Rygol, M.: COVID-19 how augmented reality is helping mitigate business impact. https://thearea.org/covid-19-how-augmented-reality-is-helping-mitigate-business-impact/ (2020)
70. Safi, M., Chung, J., Pradhan, P.: Review of augmented reality in aerospace industry. Aircr. Eng. Aerosp. Technol. (2019). https://doi.org/10.1108/AEAT-09-2018-0241
71. Samur, E.: Introduction. In: Ferre, M., Ernst, M., Wing, A. (eds.) Performance Metrics for Haptic Interfaces. Springer Series on Touch and Haptic Systems, vol. 5, p. 37. Springer, London (2012)
72. Seo, D., Kim, E., Kim, J.: Finite element analysis based augmented reality based mechanical product simulation platform for small-medium sized enterprise industry. J. Adv. Sci. Technol. **64**, 26–30 (2014)
73. Serdar, T., Aziz, E., Esche, S., Chassapis, C.: Integration of augmented reality into the CAD process. In: ASEE Annual Conference and Exposition, Proceedings, p. 23.784. ASEE, Atlanta (2013)
74. Serván, J.: Using augmented reality in AIRBUS A400M shop floor assembly work instruction. In: Proceedings of the Manufacturing Engineering Society International Conference, pp. 633–640. AIP, Cadiz (2012)
75. Spitzer, C., Ferrell, U., Ferrell, T.: Digital Avionics Handbook, 3rd edn. CRC Press, Boca Raton (2015)
76. TAE Aerospace: Award-winning innovation ready to shift aircraft maintenance paradigm. www.taeaerospace.com/media/TAE-Aerospace-Innovation-Award-media-release-3.pdf (2017)
77. UTSC School of Computing and Creative Media: Microsoft HoloLens: drone simulator (extended version). https://www.youtube.com/watch?v=Ua9ifkpd3L0 (2017)
78. Vietler, F., Hajek, M.: Evaluation of visual augmentation for rotorcraft pilots in degraded visual environment. J. Am. Helicopter Soc. **62**(1), 1–11 (2017)
79. Vlean, K.: LATAM set to launch augmented reality tool that measures carry-on luggage. https://apex.aero/2018/09/11/latam-augmented-reality-measure-carryon-luggage (2018)
80. Wang, X., Ong, S., Nee, A.: A comprehensive survey of augmented reality assembly research. J. Adv. Manuf. **4**(1), 1–22 (2016)
81. Wang, X., Ong, S., Nee, A.: Multi-modal augmented-reality assembly guidance based on bare-hand interface. Adv. Eng. Inform. **30**(3), 406–421 (2016)
82. Wang, X., Ong, S., Nee, A.: Real-virtual components interaction for assembly simulation and planning. J. Robot. Comput. Integr. Manuf. **41**, 102–114 (2016)
83. Younes, G., Asmar, D., Elhajj, I., Al-Harithy, H.: Pose tracking for augmented reality applications in outdoor archaeological sites. J. Electron. Imaging. **26**(1), 011004 (2016)

Maryam Safi is a Master of Applied Science (MASc) student at Ryerson University working at the Mixed-Reality Immersive Motion Simulation Laboratory. She received her Bachelor of Engineering in Aerospace Engineering in 2018. Her research work focuses on developing advanced aircraft cockpits using mixed reality technology. She is also studying the interactions of pilot-cockpit systems caused by these modules.

Dr. Joon Chung received his Ph.D. from the University of Toronto. He is a professor at the Department of Aerospace Engineering at Ryerson University, a member of the Professional Engineers of Ontario (PEO), a member of American Institute of Aeronautics and Astronautics (AIAA), and a member of Canadian Aeronautics and Space Institute (CASI). His research focus includes multidisciplinary design optimization, flight test data analysis systems, extended reality (XR) application in the aerospace systems, and human-machine interface within aircraft cockpit and interiors.

Augmented Reality for Building Maintenance and Operation

21

Lorena C. S. Moreira, Regina C. Ruschel, and Amir H. Behzadan

Contents

21.1 The Building Maintenance Instruction Manual 496
21.2 Augmented Reality and Facility Management 498
21.3 Systematic Literature Review 499
21.3.1 Existing Studies 499
21.3.2 Artifacts Categorization 500
21.4 Methodology 501
21.4.1 Identification of the Problem 502
21.4.2 Design and Development 502
21.4.3 Artifact Evaluation 503
21.4.4 Explicitness of Learning 505
21.4.5 Generalization to a Class of Problems 505
21.5 Design and Development 505
21.5.1 Artifact Design 505
21.5.2 Artifact Development 505
21.5.3 Artifact Publication 510
21.6 Evaluation and Findings 511
21.6.1 General Workload Measurement: NASA TLX 511
21.6.2 Workload Analysis Considering Perception Filters 522
21.7 Generalization 525
21.8 Conclusion 526
21.9 Data Availability 529
References 529

L. C. S. Moreira
Laboratory of Advanced Studies in City, Architecture and Digital Technologies (LCAD), Faculty of Architecture, Federal University of Bahia, Salvador, Brazil
e-mail: lorenasm@ufba.br

R. C. Ruschel
School of Civil Engineering, Architecture and Urban Design, University of Campinas, Campinas, Brazil
e-mail: ruschel@unicamp.br

A. H. Behzadan (✉)
Department of Construction Science, Texas A&M University, College Station, TX, USA
e-mail: abehzadan@tamu.edu

Abstract

On average, annual building maintenance and repair cost 1–5% of the initial building cost, which, accumulated over the building life, could even exceed the initial construction cost. New advancements in architecture, engineering, construction, and operation (AECO) are transforming the current practice of building maintenance operations using visualization and sensing technologies. This chapter will describe a use case for the application of augmented reality (AR) visualization in building maintenance and repair. AR enhances user's perception of the surroundings by overlaying virtual objects on real-world views and can lead to new forms of user interaction. For instance, AR visualization embedded in the building maintenance instruction manual (BMIM) can be used to guide facility managers and repair personnel. We present the design and evaluation of an AR-integrated BMIM to improve the quality of building operation and maintenance. We adopt the design science research (DSR) methodology to carry out a systematic literature review, characterization of AR features that can be applied to BMIM, development of AR artifacts for incorporating into BMIM, and assessment of user performance through experiments with measurements taken with the NASA TLX protocol. Results are implemented in two applications, namely, Living Augmented Reality (LAR) and Manual Augmented Reality (MAR), with different visualization scales. Analysis of users' workload data indicates that a majority agree that BMIM can be effectively enhanced with a high degree of acceptability using AR on mobile devices. It is also found that this integration will be helpful to future evolution of the BMIM and integration with the Internet of Things paradigm.

Keywords

Facility management · Augmented reality · Building maintenance instruction manual · User performance · Mobile device

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_21

21.1 The Building Maintenance Instruction Manual

Building maintenance and repair costs are influenced by several factors such as hours of operation, use type and intensity, location, and overall condition. While numbers vary from one facility to the next, on average, these costs could range anywhere from 1% to 5% of the initial building cost per year [1]. When accumulated over the building lifetime, the cost of maintenance and repair may even exceed the initial construction cost. Recent technological advancements in architecture, engineering, construction, and operation (AECO) domains have created new opportunities for transforming the current practice of building maintenance using new visualization and sensing technologies.

Evidently, a product is not considered complete and ready to use if not accompanied with proper information or documentation. Information is indispensable for utilizing all the features any product has to offer and if consumers are expected to put that product in good use. Traditionally, an instruction manual is supplied with a manufacturing product or commercial good, which contains key information on how to use the product and what to do in case repair or maintenance is needed. For example, the European Union (EU) legislation titled "Usable and safe operating manuals for consumer goods: guideline" specifies that a product is complete only when accompanied by an instruction manual [2]. Delivering or selling a product without an instruction manual is against the law in many parts of the world, and in such cases, the user is entitled to full assistance or cost reimbursement. Besides, the distribution of industrial products within the EU region requires a declaration of conformity for the product, and the distributor should bear responsibility for any issues or losses arising from the lack of conformity [2].

In turn, the British Standard BS 8210:2012 titled "Guide to facilities maintenance management" is aimed directly at building owners, operators, and facility managers [3]. In this standard, the manual is characterized as a maintenance document which provides technical instructions for preserving an item or restoring it to a state where it can perform its function. This same publication highlights that preparing a manual offers significant advantages by providing a clear statement of intent and necessary actions.

According to this standard, the maintenance and repair steps taken by the product (i.e., building, facility) manufacturing firm should be formalized in a manual, which should be updated periodically, and may pertain to broader documentation incorporated into a facility handbook [3]. In addition, BS 8210:2012 cites other related standards including (i) BS EN 82079:2012 titled "Preparation of instructions for use – Structuring, content and presentation – Part 1: General principles and detailed requirements," which aims to provide the general principles and specific requirements for the design and creation of all types of use instructions for all types of products' users, and (ii) BS EN 13460:2009 titled "Maintenance – Documentation for maintenance," which deals with the operational phase and equipment life cycle and describes a list of documents required for maintenance [3–5].

In Brazil, the 1990 Law No. 8078, better known as the Consumer Protection Code, provides for consumer protection and other standards (Brazil, 1990). In this law, a product is characterized by any property, movable or immovable, material, or immaterial. Inserted as fundamental rights of the consumer is the knowledge of the appropriate information about products and services, with correct specification of quantity, characteristics, composition, quality, taxes, and prices, in addition to the risks they present. Article 50 of the referred law stands out for establishing parameters to be followed by the contractual guarantee. The warranty term should be standardized and adequately state what the warranty consists of, as well as its form, term, place, and burden on the consumer. The installation and product instruction manual must be delivered in a simple, didactic language and with illustrations.

Although the concept of instruction manual has evolved over time, its format and delivery mode have to the most extent remained the same (i.e., traditional printed form, computer disk, or hypertext). Regardless of the type of media used, the manual is often filled with textual or tabular information or, at best, basic illustrations. Recent advancements in visualization technologies have created new opportunities to change how people with different levels of knowledge and/or learning styles interact with and understand contents or objects in their surroundings. What makes this new approach even more practical is the ubiquity of personalized technologies which has given the new generation a natural ability to use mobile and intelligent tools and a strong affinity to integrate new devices into daily tasks [6].

In this chapter, the authors focus on the "building" as the object (i.e., product) to be clarified by the instruction manual. Zevi [7] states that the traditional method of building representation (floor plans, elevations, sections, and photography) does not completely represent the architectural space because of fragmentation and ambiguity in the two-dimensional (2D) drawings used. This same author opines that:

> The plan of a building is nothing more than an abstract projection on the horizontal plane of all its walls, a reality that nobody sees except on paper, whose only justification depends on the need to measure distances between the various elements of the building, to the workers who are to perform the work materially. [7]

The research challenge addressed in this chapter can be best described as follows: the current delivery method of building maintenance instruction manual (BMIM) in textual format and technical language does not cater to the needs and level of understanding of the end users (i.e., building owners, facility managers), who are, in turn, not motivated enough to use it. As a result, it does not fulfill its role of guiding building maintenance and use activities.

The way we, as humans, perceive reality is continuously changing, as digital technology revolutionizes our view of the world and our interface with the surrounding environment. New information and communications technologies (ICT) constantly challenge the status quo by creating new ways of thinking and living, and human relationships depend on constant metamorphosis of informational devices of all kinds [8]. Augmented reality (AR) is one of these technologies that can be used to enhance the understanding of specific contents. AR is a fast-evolving field of research and development with significant growth potential.

AR combines virtual and real-world scenes in an effort to increase user perception and interaction [9]. It can also be understood as the overlapping of computer-generated information about the real environment through the user's view, differing from virtual reality (VR) in that objects in AR coexist with real environment objects [10]. In a typical AR application, at least one computing device (computer or mobile device) equipped with a visual display (e.g., monitor, projector, head-mounted display) and an image capturing device (e.g., camera) is used. In more robust computational applications, or when involving multiple users, researchers have also used web systems [11]. Regardless of the type of AR application, the unique advantage of AR is to use the real world as background, allowing the possibility of interaction with the virtual environment.

Research involving AR in AECO is a current trend. Rankohi and Waugh [12] conducted a literature review involving 133 articles from AECO journals that contained the keyword "augmented reality" and identified, among other aspects, the following types of applications for AR: (i) visualization or simulation, (ii) communication or collaboration, (iii) information modeling, (iv) access to information or evaluation, (v) monitoring, (vi) education or training, and (vii) safety and inspection. Other studies have proposed innovative ways of using AR to enhance construction education by combining written content with 3D viewing through AR [10, 13]. According to these studies, although students often have excellent theoretical knowledge, they do not know how to apply them in practice. In a different study [14], AR was used for operation and assembly tasks by comparing the assembly of parts through the isometric design of hydraulic installations and AR visualization, and it was found that using AR could lead to improved productivity and assembly performance by reducing cognitive workload. Similarly, Hou et al. [15] analyzed the effect of using AR on the user's cognitive load in assembly tasks and discussed how this approach could shorten the learning of new assembly workers. The study also compared the printed assembly guidebook with the designed AR system, and results indicated a positive impact demonstrated by the increase of the users' learning curve and the reduction of mistakes.

21

Wang et al. [16] addressed the gap between building information modeling (BIM) and AR and proposed a conceptual framework that integrates the two so that the physical context of each building activity can be viewed in real time. They suggest that for this junction to be effective, AR must be ubiquitous and act in conjunction with tracking and detection technologies. Similarly, Olbrich et al. [17] studied the problem of information visualization in AR based on BIM models and stated that the challenge is to give the agents involved in the building's life cycle access to the management system through on-site inspections and seamless exchange of information.

Considering BMIM within the context of the existing literature, the potential of creating a more interactive, accessible, dynamic, and instructive BMIM can be explained. This association can extrapolate the visualization limits and move toward an approximation with the constructed object (i.e., building). Thus, this chapter seeks to enhance the performance and adoption of the BMIM, through a closer relationship with the end user that, in most cases, is not familiar with construction terms. Using AR in facility management and building operations is relatively new and underexplored. Facility managers regularly visit managed spaces. Using mobile devices with information visualization in AR has shown to improve the quality of maintenance and operation tasks [18]. Timely information support plays an essential role in problem containment and diagnostics [19]. In this sense, a facility manager as well as the owner and the project manager may adopt the BMIM as a management tool for building maintenance and operation activities. Thus, the authors envision an innovative application of expressing concepts related to the way users interact with the BMIM.

The presented work in this chapter is multidisciplinary involving areas of design, construction, facility management, and computer science. The contribution of this research is to clarify the most appropriate ways to incorporate AR in the building maintenance instruction manual (BMIM) and qualify the performance improvements in the use of the AR-enhanced manual following NBR 14037 [20]. Therefore, the general objective of this study is to present the design and evaluation of an AR-integrated BMIM to improve the quality of building operation and maintenance. In order to guide the research and achieve the proposed objective, as described throughout this chapter, the design science research (DSR) method, also known as constructive research [21], is applied.

21.2 Augmented Reality and Facility Management

Visualization has gained increasing credibility among construction researchers and has been considered as one of the top four IT domains in this field [22]. Although every construction project is unique, several tasks (including periodic inspections and preventive and corrective maintenance) are repeated throughout the life of many constructed facilities. Visualization is a powerful tool to enhance the user/stakeholder experience and along the lifecycle of a construction project. Conflicts and other essential aspects in project execution can be understood in the pre-construction phase using a variety of visualization and simulation techniques. Visualization tools have also been used to improve and revolutionize current design approaches, jobsite safety [23], and on-site diagnostics in combination with real-time sensor data [24]. In AECO, physical objects often need to be related to their information, making AR a great candidate to achieve this by assisting users to view the environment complemented with the necessary information, united in a single interface [25].

In the past two decades, the AECO community has examined different approaches, including VR and AR, to improve communication, visualization, and coordination among different project stakeholders [24]. For Nee et al. [26], most collaborative visualization systems in AR are systems based on design visualization. In these systems, virtual models are presented in AR to the designers in order to facilitate the decision-making process. Parallel to the virtual model, knowledge about objects such as metadata, design information, and annotations can also be viewed in AR to facilitate decision-making in these collaborative systems.

There is also a growing tendency for users to interact directly with the information associated with the production process. AR can integrate these modalities in real time into the work environment, which is useful for manufacturing, assembly, training, and maintenance activities. Moreover, AR can provide users with a path of direct and intuitive interaction with information in the manufacturing process [26]. Several studies prove the use of AR in this regard. For instance, AR was used in the construction planning phase to assist in decision-making [27] and has been applied to training and simulation of the machining process of CNC machines [28]. AR has also been deployed to assist the design and intuitive assembly through gestures and manipulation of virtual objects [29], as well as for inspection and instruction in construction [30].

Wang et al. [16] presented a conceptual approach with several examples of how AR can be incorporated into BIM and listed potential use cases such as connecting design information with physical environment, mental model synchronization for communication and project control, monitoring and feedback by comparing as-built with as-planned information and visualizing discrepancies between design and production, and finally site and storage planning. In assembly, operation, and maintenance tasks, experiments have shown that the use of AR can improve operator's understanding and process control [28, 31].

Overlaying digital information on the views of the real environment using AR can assist workers to implement the correct assembly procedures with greater accuracy and error reduction. In a particular study that used laboratory experiments, researchers observed the reduction of task completion time by 55% and the reduction of assembly errors (by reducing rework) by 46% [14].

Maintenance activities, such as preventive and corrective activities, are almost always established according to predefined procedures and accepted protocols. Workers in this area need to be trained to carry out specific maintenance-related tasks. These professionals sometimes need to seek assistance from support systems and specialists in the field. Training on maintenance tasks can be done using 2D printed materials and VR simulation systems. However, VR technologies are rarely applied to maintenance, where interaction with actual physical equipment is required. AR visualization has an advantage in these applications in that the user interface can be ubiquitously designed to allow for an unobstructed view of the real physical object while accessing necessary instructions and maintenance data [26].

Graf et al. [32] add that with the increasing use of BIM in AECO, new opportunities arise to assist facility maintenance and operation professionals in taking advantage of the building's lifecycle-related BIM information and real-time environment simulation. In particular, the combination of BIM, computer vision, and tracking technologies will enable future applications for as-built capture and viewing of "just-in-time" (JIT) operation and maintenance information.

A study by Irizarry et al. [33] investigated the situational awareness integrated into the context of FM, BIM, and AR. First, the as-built BIM model of space was conceived, and the geometry of this model was simplified to be exported, in an appropriate format, to 3D visualization software. Next, 360° panoramic images were generated (a camera was positioned within the BIM model). The simplified geometry was then imported into Google SketchUp to geo-reference and position physical markers called Information Surveyed Point for Observation and Tracking (InfoSPOT). The usability and the interface quality of the developed prototype were successfully tested in a user study, presenting InfoSPOT as a low-cost solution that utilizes AR-integrated BIM for facility management.

21.3 Systematic Literature Review

Systematic literature review (SLR) is part of DSR (aka constructive research) method, previously explained. According to Ref. [34], SLR is a scientific investigation that was first adopted in the late 1980s, due to a large number of publications produced and the absence of an appropriate literature review methodology. SLR is a means of identifying, evaluating, and interpreting all available research relevant to a particular research question, topic, or phenomenon of interest [35]. Studies that contribute to the systematic review are called primary studies, and a systematic review is a secondary study. Systematic reviews should be performed according to a predefined search strategy. The search strategy should allow the completeness of the search to be evaluated. In particular, researchers who conduct systematic reviews should make every effort to identify and report research that supports the research hypothesis, as well as identify and report research gaps [35].

SLR is different from conventional literature review in several aspects. In particular, SLR begins by defining a protocol that specifies the research question and the methods that will be used in the study. In addition, SLR is based on defined search strategies intended to detect as much relevant literature as possible. These search strategies should be documented, so readers can access their credibility and completeness. Furthermore, SLR requires explicit inclusion and exclusion criteria and specifies the information to be obtained from each primary study, including the evaluation criteria. Finally, SLR is a prerequisite for a quantitative meta-analysis [35]. An important aspect of SLR is the protocol validation by a subject matter expert. If this evaluation yields unsatisfactory results, the protocol should be reformulated [34].

In this study, SLR is performed to identify existing work in AR applications related to building assembly, maintenance, and operation along with instruction manuals. The following questions guide the review of literature: (a) what prototypes are produced in AR applications? and (b) what tracking techniques are used in AR applications?

The time range used in conducting the SLR is from 1997 until September 2019. The starting point of 1997 is based on the article *A Survey of Augmented Reality* by Ronald Azuma which was published in *Presence: Teleoperators and Virtual Environments* of the Massachusetts Institute of Technology (MIT). This is one of the pioneering studies in AR which systematically introduces the concept of AR and presents an extensive bibliography in this field of research [36]. In the presented research, the developed SLR protocol consists of planning, conduction and information extraction, analysis, and synthesis of results. The steps are summarized in Table 21.1.

During the synthesis of results, timelines are elaborated considering the number of publications and the tracking method used. Also, a hierarchical categorization of activities in AECO is used [39], which is presented in Table 21.2. The artifacts are categorized according to the first three levels, namely, Area (AA), Application (AP), and Activities (AC).

21.3.1 Existing Studies

Figure 21.1 shows the temporal evolution of AR studies in the areas of building assembly, operation, and maintenance or instruction manuals. According to this figure, while initial work in these fields began in 1999, the SLR revealed no relevant literature in the years 2000, 2001, 2004–2008, and 2016. Additionally, as of 2011, a significant growth among the selected articles is observed, with the highest numbers occurring from 2013 (7 articles), reaching a peak in 2018 (11 articles).

This quantitative analysis suggests that AR applications used in the building assembly, maintenance, and operation are part of a relatively new theme and are experiencing a slow growth. According to Ref. [40], areas such as mechanics and aeronautics started to draw attention at the same time that the term “augmented reality” was first coined in the early 1990s. In short, the identified sources covered a period of 23 years, and the sample analyzed found a late appearance of these applications, specifically as related to building maintenance and operation. However, this area of research is on the rise and has an emphasis on creating functional prototypes.

An analysis of the tracking methods used in existing AR applications developed for building assembly, maintenance, and operation and/or instruction manuals identifies three different techniques that include using markers, markerless method (natural feature tracking), and sensors. As shown in Fig. 21.2, among the 73 articles analyzed, 55 explicitly characterize some form of tracking systems. It is found that the use of markers is predominant in 58% of the studies, followed by the use of sensors and markerless methods with 22% and 20%, respectively. It is also found that as of 2019, the use of newer tracking methods (i.e., sensors and markerless) is on the rise (Fig. 21.2).

Lee and Akin [19] used markers as a tracking technique for the development of an AR application in operation and maintenance of equipment and justified their approach by citing the reported inaccuracy of other tracking methods when utilized inside complex buildings and indoor spaces. In other studies, researchers have used sensing techniques such as global positioning system (GPS) and six degree-of-freedom (6DOF) tracking systems to locate objects for visualization applications. For example, Schall et al. [41] developed a

Table 21.1 Summary of SLR protocol

Intervention		DSR products from AR applications
Control		[15, 17, 37, 38]
Population		Augmented reality applied to building assembly, maintenance, and operation and applications with instruction manuals
Results		Summary of applications and type of tracking used in building assembly, maintenance, and operation together with instruction manual
Applications		Studies in the field of AR for AECO
Search strategy	**Source**	Compendex (www.engineeringvillage.com) Scopus (https://www.scopus.com) Web of Science (https://apps.webofknowledge.com) American Society of Civil Engineers (ASCE – http://ascelibrary.org)
	Idiom	English
	Search terms	"Augmented reality" with (building and assembly), (building and operation), (building and maintenance), and (building and manual)
	Where	Title, abstract, and keywords
	Types of articles	Article published in peer-reviewed journals
Filters	**Inclusion criteria**	Full text available in scientific databases Works published since 1997 up to when search was performed (September 6, 2019)
	Exclusion criteria	Works that only present the application algorithm and equipment calibration; works with approaches outside the context of applications for assembly, maintenance, and operation or instruction manual; works that do not refer to VR and/or AR
Strategy		After applying the inclusion filters, the selected papers were read, and forms were filled out for each article. Papers were reviewed considering the exclusion criteria understanding and application for each article
Synthesis of results		Both quantitative and qualitative analyses were elaborated. The publications were analyzed and the following information extracted from each article: (i) tracking method used in AR for the proposed applications and (ii) prototype characterization

Table 21.2 Hierarchical categorization of AECO activities [39]

Level	Description	Examples
1	Area (AA)	Architecture, engineering, construction, operation, education
2	Application (AP)	Safety, maintenance, repair, construction, inspection, design, coordination, collaboration, logistics, etc.
3	Activities (AC)	Assembly, visualization, planning, monitoring, robotics fabrication, equipment control, manufacturing, handcrafting, etc.
4	Composite tasks (CT)	Measure, connect, organize, select, align, annotate, inform, etc.
5	Primitive tasks (PT)	Move, reach, catch, etc.

GPS-based method for operation and maintenance that uses AR for various verification tasks in urban installations. In this study, the accuracy achieved was less than 30 cm. In markerless tracking method, Olbrich et al. [17] presented the visualization of information based on BIM models, using feature point-based tracking (based on the characteristics of the points). In this study, a system for creating mobile AR applications was proposed, in which 3D models coexist with semantic information. The combination of BIM with AR provides a visualization of construction-related data on-site and provides support for documenting content via mobile devices.

21.3.2 Artifacts Categorization

Among all the studies that were gathered in the SLR process, those that resulted in the creation of AR application prototypes for assembly, maintenance, and operation of the building and/or instruction manuals are listed and characterized in Table 21.3. This categorization is based on the mixed reality taxonomy for operation and maintenance tasks at AECO [39], presented in Table 21.2.

Visualization is the most recurring activity being the subject of the largest number of studies (24) and covers all areas of application listed in Table 21.2. Besides, other activities observed are assembly and monitoring which are covered in 12 and 10 studies, respectively, followed by equipment control (6 studies), planning (4 studies), and, finally, handcrafting and robotics fabrication (each in 1 study) activities.

Among the prototypes found, monitoring applications [17, 19], visualization [33], and editing [14, 15] are closely related to the topic of this study. Lee and Akin [19] present an AR prototype aimed at equipment maintenance and highlight the following information types: (i) maintenance information (that includes specifications, subsystems, and components, as well as agents involved and maintenance history); (ii) operating information (that includes performance data or data flow); and (iii) geometric representation (that helps the maintenance professional to better understand the equipment

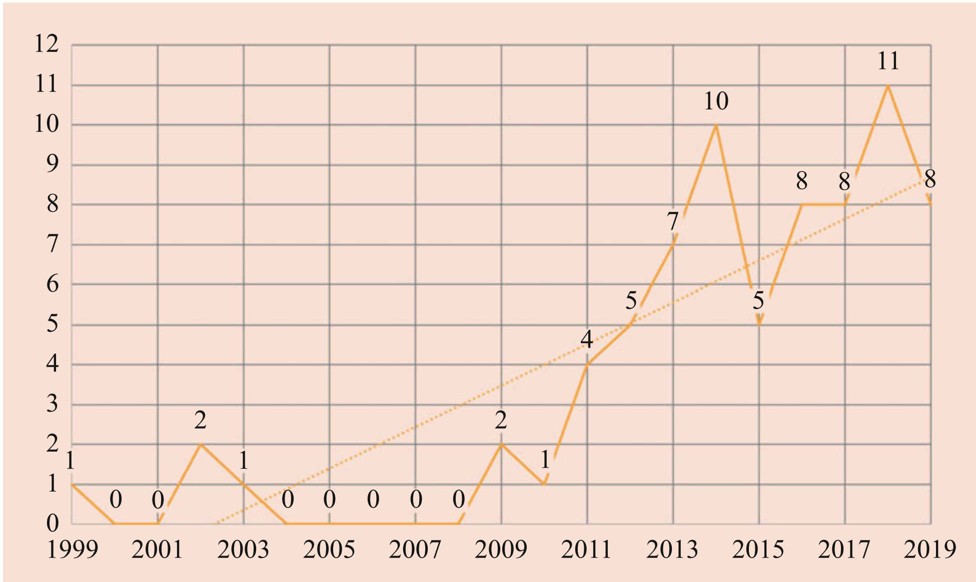


Fig. 21.1 Temporal distribution of AR studies applied to building assembly, operation, and maintenance or instruction manuals by year of publication

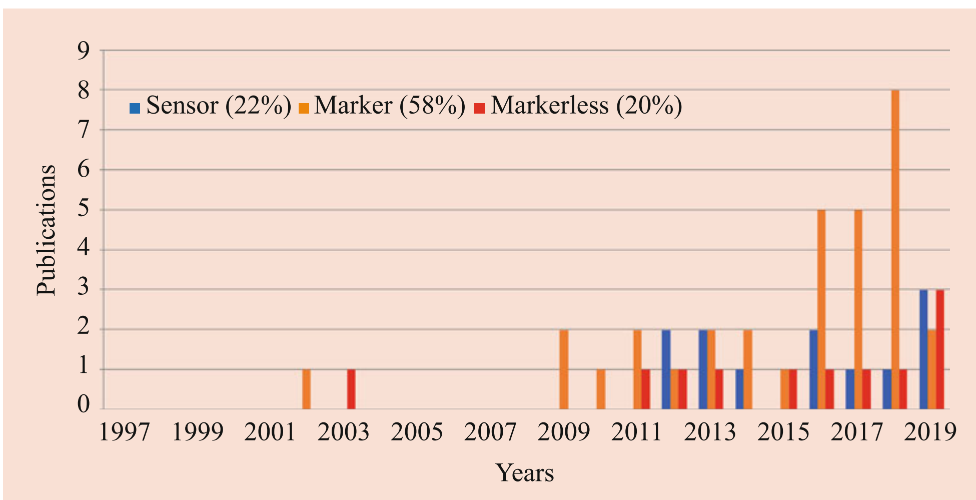


Fig. 21.2 Tracking types used in AR applications by year of publication

status and its location). The study by Olbrich et al. [17] explores the visualization of information, through AR, based on BIM models and a user-centered annotation mechanism. The importance of this study, within the context of FM, is that it emphasizes the use of information added to BIM using AR visualization. Moreover, the survey by Irizarry et al. [33] shows the use of BIM for FM through the creation of 360° panoramic images generated from the model. The work of Hou et al. [15] presents an assembly system for measuring cognitive work and compares traditional assembly with assembly using AR. Finally, the survey by Hou et al. [14] presents an application for the assembly of hydraulic installations assisted by AR, in which a sequence of hydrosanitary parts was assembled, and the assembly time and execution errors were checked.

21.4 Methodology

As previously described, the research methodology adopted in this study is based on the DSR method, also known as constructive research. In particular, the specific steps followed include a SLR of AR applied to building assembly, maintenance, and operation or manual, characterization of features of AR that can be applied to the BMIM, development of proposals for AR incorporation in the BMIM as well

Table 21.3 Characterization of AR prototypes applied to building assembly, operation, and maintenance or manual

Activities (AC)	Application (AP)	Area (AA)	Publications
Equipment control	Design	Operation	[42]
		Education	[43]
	Manufacturing	Education	[28]
	Inspection	Operation	[44]
	Maintenance	Operation	[45]
	Test	Education	[46]
Handcrafting	Construction	Construction	[47]
Monitoring	Construction	Construction	[48]
	Manufacturing	Construction	[49, 50]
	Inspection	Operation	[17, 51, 52]
	Maintenance	Education	[53]
		Construction	[19]
		Operation	[54]
	Safety	Operation	[55]
Assembly	Collaboration	Operation	[56]
	Construction	Education	[57]
		Operation	[58]
		Construction	[59]
	Manufacturing	Construction	[60, 61]
		Engineering	[15, 62]
		Education	[63]
	Maintenance	Operation	[64]
		Engineering	[14]
	Safety	Education	[65]
Planning	Design	Education	[66]
	Construction	Engineering	[67]
		Operation	[41]
	Logistic	Construction	[68]
Robotics fabrication	Manufacturing	Construction	[69]
Visualization	Collaboration	Operation	[70]
	Design	Architecture	[71, 72]
		Engineering	[73]
		Education	[74–77]
		Operation	[78]
	Construction	Architecture	[79]
		Construction	[80, 81]
		Engineering	[82]
	Coordination	Operation	[33]
	Manufacturing	Engineering	[83]
	Inspection	Construction	[84]
		Operation	[85]
	Localization	Education	[86]
	Maintenance	Construction	[87]
		Operation	[88–90]
	Safety	Construction	[91]
		Education	[92]

as proposals to incorporate the BMIM in the environment, and, finally, a comparison of user performance when presented with different forms of visualization (tablet computers and smart glasses) through experiments with measurements following the NASA TLX protocol. We adopt the outline presented in Ref. [93] according to Fig. 21.3 and explained in the following sections.

21.4.1 Identification of the Problem

The first step of the designed methodology is the identification of the problem in the form of addressing the following research questions: Does the incorporation of AR into BMIM stimulate gains with respect to the tasks being visualized? Does the choice of information delivery device used for AR visualization influence gains? Does the scale of AR visualization (i.e., small scale on paper vs. real scale in the object environment) influence gains? In this study, gain is defined as a measure of workload perceived by the user when completing a task.

21.4.2 Design and Development

The design of the artifact must consider its internal characteristics and the external context in which the artifact operates. Development corresponds to the process of the constitution of the artifact itself. For a better understanding of the perception of the BMIM by users, a satisfaction survey with apartment owners (in Goiânia, Brazil) was carried out [94]. The satisfaction survey included four questions regarding the BMIM:

(i) Did you obtain the necessary clarifications when consulting the manual?
(ii) How satisfied were you with the manual?
(iii) If you have read or consulted the manual, what were the reasons?
(iv) Suggest possible improvements to the manual.

Considering the results of this survey and the outcome of the SLR for identifying AR prototypes, one can observe how various stakeholders deal with the BMIM.

From the user's point of view, the satisfaction survey carried out revealed that the respondents rarely consulted with the manual (were indifferent to the manual), showing disinterest in its use and the technical information it presented. Another observation from this survey was that when users chose to use the manual, it was merely for checking items for use and maintenance and to understand equipment operation and verify accident prevention. In addition, it was found that building owners were in favor of a new form of presentation of the manual. There was a general tendency toward delivery methods that are interactive, present visual content (e.g., 3D model of the building), and make use of mobile devices.

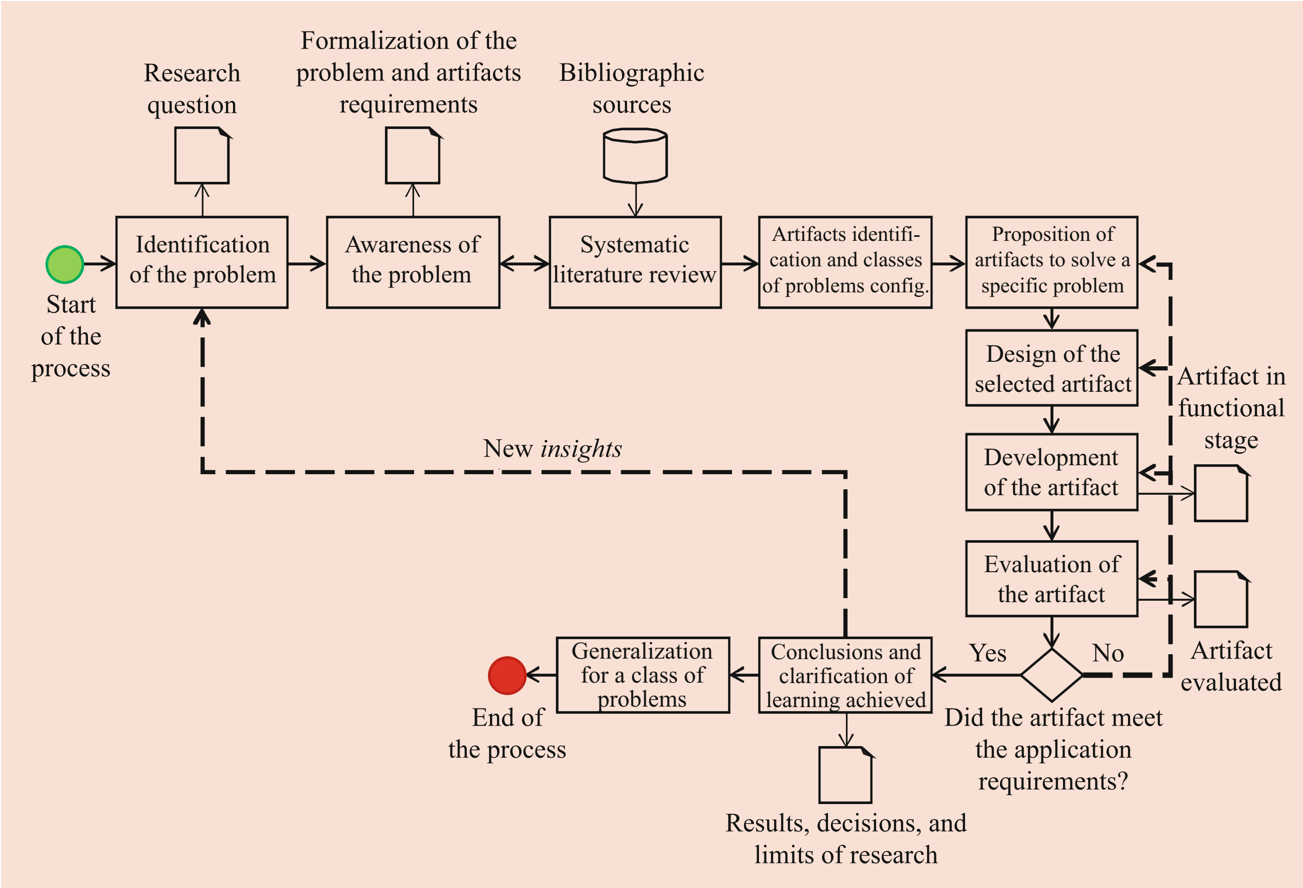


Fig. 21.3 Designed research methodology based on the DSR method

Furthermore, the survey found that not all builders and developers create manuals, rather they hire other professionals/companies to do the job, and the resulting manual conforms to the building's use and standard [94]. The preferred manual format is, in the vast majority of cases, textual, followed by digital content (computer disk or memory stick). Likewise, few companies invest in creating websites and online spaces for their BMIMs, and no builder in the sample population prepares the manual in the form of an application for mobile devices. In verifying the adherence of the manuals to NBR 14037 [20], it becomes evident that the sections of the BMIM that contain information on guarantees and technical assistance are the most complete among all sections.

As for the structuring of the content of the evaluated BMIMs, it is found that most do not follow the structure recommended by NBR 14037 [94], although the particular structure adopted is perceived to add information by component and facilitate its understanding. Still, the manuals are not very attractive because they do not use visual aids that help increase user comprehension. The structuring of the content of most of the evaluated BMIMs provides further evidence that information is often aggregated by building component. This premise is therefore followed in the proposed AR artifact in this research.

In designing the AR artifact in this research, the marker-based tracking method stood out as an option for its accessibility (easy to print and install) and scalability as, unlike sensor-based techniques, it does not require sophisticated sensors and calibration steps [19, 41].

Two prototypes are introduced and evaluated in this research: (i) traditional BMIM plus AR and (ii) incorporating BMIM within the physical environment using AR. Markers are assigned to the printed BMIM to facilitate user interaction with both textual and media contents in AR. The scale of visual data presented in AR is chosen to be 1:100 (when visualizations are overlaid on the printed BMIM) or 1:1 (when visualizations are inserted in the physical environment).

21.4.3 Artifact Evaluation

Dunser and Billinghurst [95] suggest four types of assessment to evaluate AR applications, which are (i) experiments that study human perception and cognition, (ii) experiments that examine user performance in a task, (iii) experiments that examine user collaboration, and (iv) usability and system design evaluation. In this research, evaluation is performed with end users through experiments that examine the user's performance in a task, as well as contingency heuristics that explain the artifact limits and its conditions of use [93].

We adopt the NASA Task Load Index (NASA TLX) method [96], which is a multidimensional assessment process that provides an overall workload index based on a weighted average rating of six factors, namely, mental demand, physical demand, temporal demand, performance, effort, and frustration. Three of these factors relate to the demands placed on the individual (mental demand, physical demand, and temporal demand), while the other three factors measure an individual's interaction with the task (performance, effort, and frustration). Each factor is described below:

Mental demand: Amount of mental activity and perception required, for example, thinking, deciding, calculating, remembering, looking, searching, etc. Was the task easy or difficult? Simple or complex?

Physical demand: Amount of physical activity required, that is, pushing, pulling, turning, controlling, activating, etc. Was the task easy or difficult? Slow or fast? Slow or strenuous? Restful or laborious?

Time demand: Time pressure felt due to the rate or pace at which task elements occurred. Was the pace slow and slow or fast and frantic?

Performance: Level of success in accomplishing the task objectives. How satisfied was the user with his/her performance in achieving these goals?

Effort: How hard was it to work (mentally and physically) to achieve performance?

Level of frustration: How insecure, discouraged, angry, stressed versus safe, grateful, contented, relaxed, and complacent did the user feel while performing the task?

Pairwise factor comparisons determine the degree to which each of the above six factors contributes to the perception of the workload of each task by the user. Magnitude assessments on each subscale are obtained after each performance of a task. Rankings of the factors considered most important in creating a workload for a task are given higher weight in calculating the overall workload scale, thereby increasing scale sensitivity [96].

The first requirement is that each individual assesses the contribution (i.e., weight) of each factor to the workload of a specific task. There are 15 possible pair comparisons of the 6 scales. Each pair is presented to the individual, who will choose one factor from that pair that contributed most to the workload of the task performed. The number of times each factor is chosen is computed. The score may range from 0 (not relevant) to 5 (most important) [96].

The second requirement is to obtain numerical ratings for each scale that reflects the magnitude of that factor in a given task. The scales are presented on a rating sheet, and the individual responds by marking at the desired location (gross rating). Each individual's overall workload score is calculated by multiplying each rating by the weight assigned to that factor by the individual himself (adjusted rating). The sum of assessments, for each task, is divided by 15 (the sum of weights) to obtain the overall workload score of the individual on that task. The value of the overall workload is obtained by averaging individual weighted workloads [96]. Table 21.4 presents the factorial experimental plan of the evaluation with the individuals of the developed solutions.

A pilot experiment is performed with university students, staff, and teachers, as well as shoppers at a building supply store as volunteers. Three manuals are evaluated: (A) traditional BMIM, (B) traditional BMIM plus AR (Manual Augmented Reality or MAR app), and (C) environment incorporating the manual with AR (Living Augmented Reality or LAR app). Two forms of visualization are applied to prototypes B and C, which are AR viewed on a tablet and AR viewed through smart glasses. In each of the five assessment scenarios, the NASA TLX measurement method is applied.

Table 21.4 Experimental plan for individual assessment of proposed solutions

Prototype	View type		
	Tablet	AR glasses	Paper
(A) Traditional BMIM	–	–	Individual assessment of solution A (manual selected after survey)
(B) Traditional BMIM plus AR (Manual Augmented Reality – MAR app)	Individual assessment of solution B with tablet	Individual assessment of solution B with AR glasses (Epson Moverio BT, https://epson.com/moverio-augmented-reality)	–
(C) Environment incorporating the manual with AR (Living Augmented Reality – LAR app)	Individual assessment of solution C with tablet	Individual assessment of solution C with AR glasses (Epson Moverio BT, https://epson.com/moverio-augmented-reality)	–

21.4.4 Explicitness of Learning

This stage aims to explain the lessons learned during the research process, considering the results observed in the evaluation stage. Success and failure points are also described. This approach ensures that research can serve as a reference and as a basis for knowledge generation [97]. In this study, we verify if the proposed solutions can be applied to any BMIM or sections of the manual, identify the limitations of this application, and suggest improvements to enhance knowledge in this subject.

21.4.5 Generalization to a Class of Problems

This step allows the advancement of knowledge in DSR. Generalization makes the acquired knowledge replicable in other similar situations through the use of inductive reasoning [93]. We further verify if the prototypes of the proposed solution for the BMIM can be used in a variety of assembly, maintenance, and instruction tasks related to building engineering, construction, and operation.

21.5 Design and Development

21.5.1 Artifact Design

The tangible outcomes of this research are two applications: (i) Living Augmented Reality (LAR) which incorporates BMIM into an AR environment and (ii) Augmented Reality Manual (MAR) which supplements a traditional paper-based BMIM with AR [98]. Both prototypes include a sample maintenance activity selected from one of the manuals collected in the cataloging and classification stage. The activity chosen for the prototypes is the replacement of a toilet float [94].

The traditional BMIM contains step-by-step instructions (in text format) to perform this activity, which are as follows: (i) carefully open and remove the cover of the coupled box; (ii) detach the floater; (iii) take it to a building materials warehouse to serve as a model for the purchase of a new one; and (iv) with the new float in hand, fit it exactly where the old one was taken from. The same instructions are also used in developing the LAR and MAR applications, each in two versions: one for tablet and one for smart glasses. For both applications, markers are used for tracking and registering the virtual objects in the real world.

In the LAR application, the virtual toilet model has a 1:1 scale and is superimposed on the real toilet where the marker is fixed. Furthermore, the step-by-step floater replacement activity is demonstrated in a transcribed AR animation. The MAR application, on the other hand, uses AR to overlay a 1:10 scaled virtual toilet model on a printed manual. For evaluation, all four prototypes (LAR and MAR, launched on tablet and smart glasses) are evaluated against the traditional BMIM (print format).

The tablet computer used in this study is an Apple iPad Air running on Apple's iOS operating system with Wi-Fi, 64 GB capacity, 9.7-inch retina display with gyroscope, 3-axis accelerometer and ambient light sensor, A7 64-bit architecture chip, M7 motion coprocessor, approximately 10-h battery life, and wireless and Bluetooth connectivity.

The second type of display device is the Epson Moverio BT smart glasses running on Android version 4.0 with Dual Core 1.2 GHz processor and 1 GB RAM and 8 GB internal memory (with expandability up to 32 GB with an SD card), Dolby Digital Plus sound system, wireless and Bluetooth 3.0 connectivity, and integrated sensors that include a VGA camera, gyroscope, GPS, accelerometer, compass, and microphone.

Figure 21.4 presents a navigation scheme of the AR application. From the "home screen," the user has the option to learn about the app by clicking on the "About" button, print and place the marker by clicking on the "Instructions" button, or select "Start" to continue to the application. For better navigation, from the "Instructions" and "About" screens, the user can access the "Start" screen, which will trigger the device's camera. A guide will then appear to help the user aim the camera at the marker, which will lead to the virtual model on the marker. On this same screen, there are the buttons, namely, "Replace floater" and "Information," and a third button that allows the user to return to the "home screen." When accessing "Replace floater," the user will see a step-by-step animation in AR that can be controlled during playtime. By accessing the "Information" button, the user will view technical specifications, warranties, supplier's contact information, and technical content about the toilet model and issues related to hydraulic installation and sewage systems.

21.5.2 Artifact Development

As shown in Fig. 21.5, the steps taken to develop the LAR and MAR applications for Apple and Android devices followed the sequence of Revit (https://www.autodesk.com/), 3DS Max (https://www.autodesk.com/), Maya (https://www.autodesk.com/), Adobe Photoshop (https://www.adobe.com/products/photoshop.html), Unity (https://unity.com/), and Vuforia (https://www.ptc.com/en/products/vuforia) programs.

The first step of the development is virtual object modeling during which the toilet and its components are first

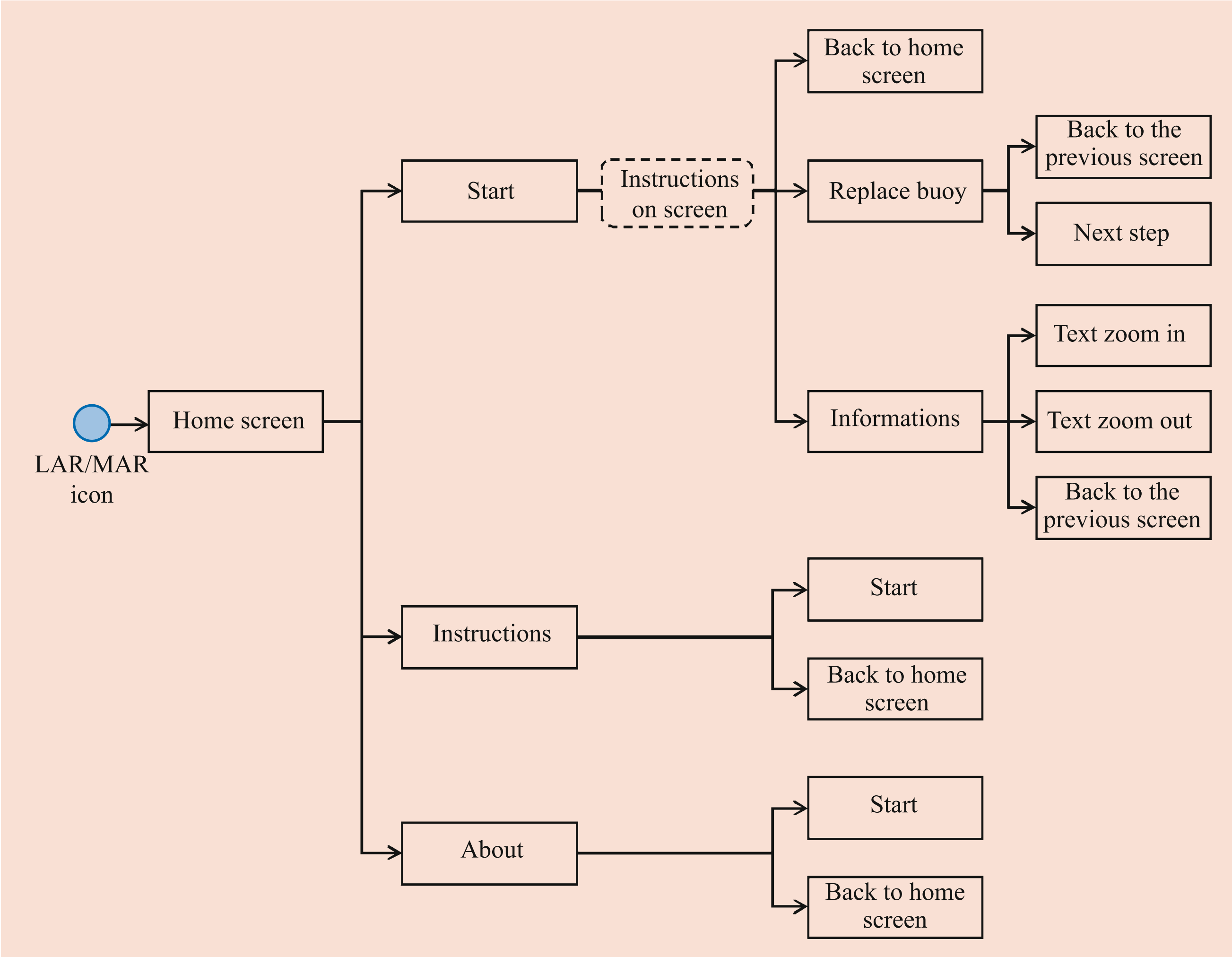


Fig. 21.4 Navigation scheme of LAR and MAR applications

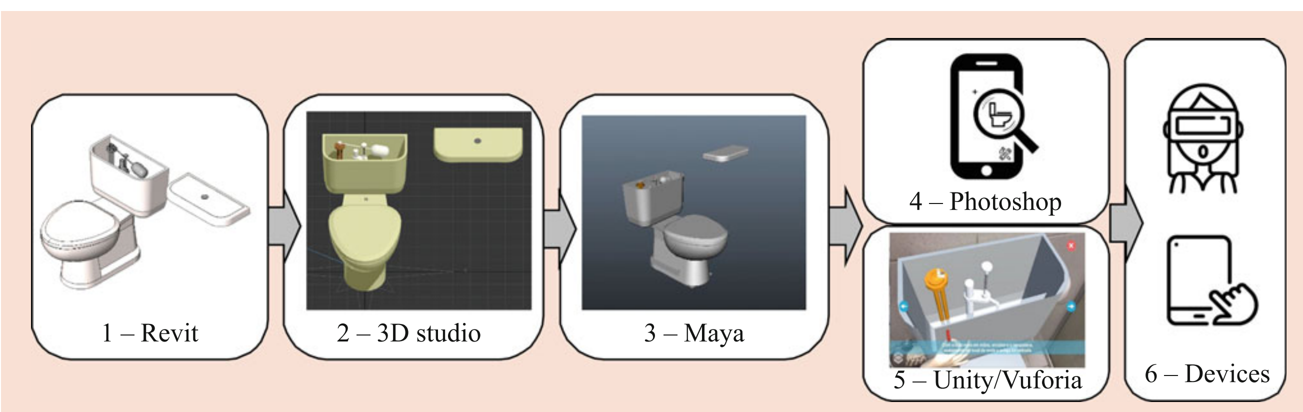


Fig. 21.5 Stages of artifact development

modeled in Autodesk's BIM Revit. The basic toilet model is obtained from the object library (Fig. 21.6) and is further enhanced by adding the coupling box model created after a commercially available prototype [99] (Fig. 21.7).

In the second step, the BIM model (toilet plus coupling box mechanism) is exported to 3DS Max in .fbx format for adding textures and materials (Fig. 21.8). It is noteworthy that exporting the BIM model to 3DS Max leads

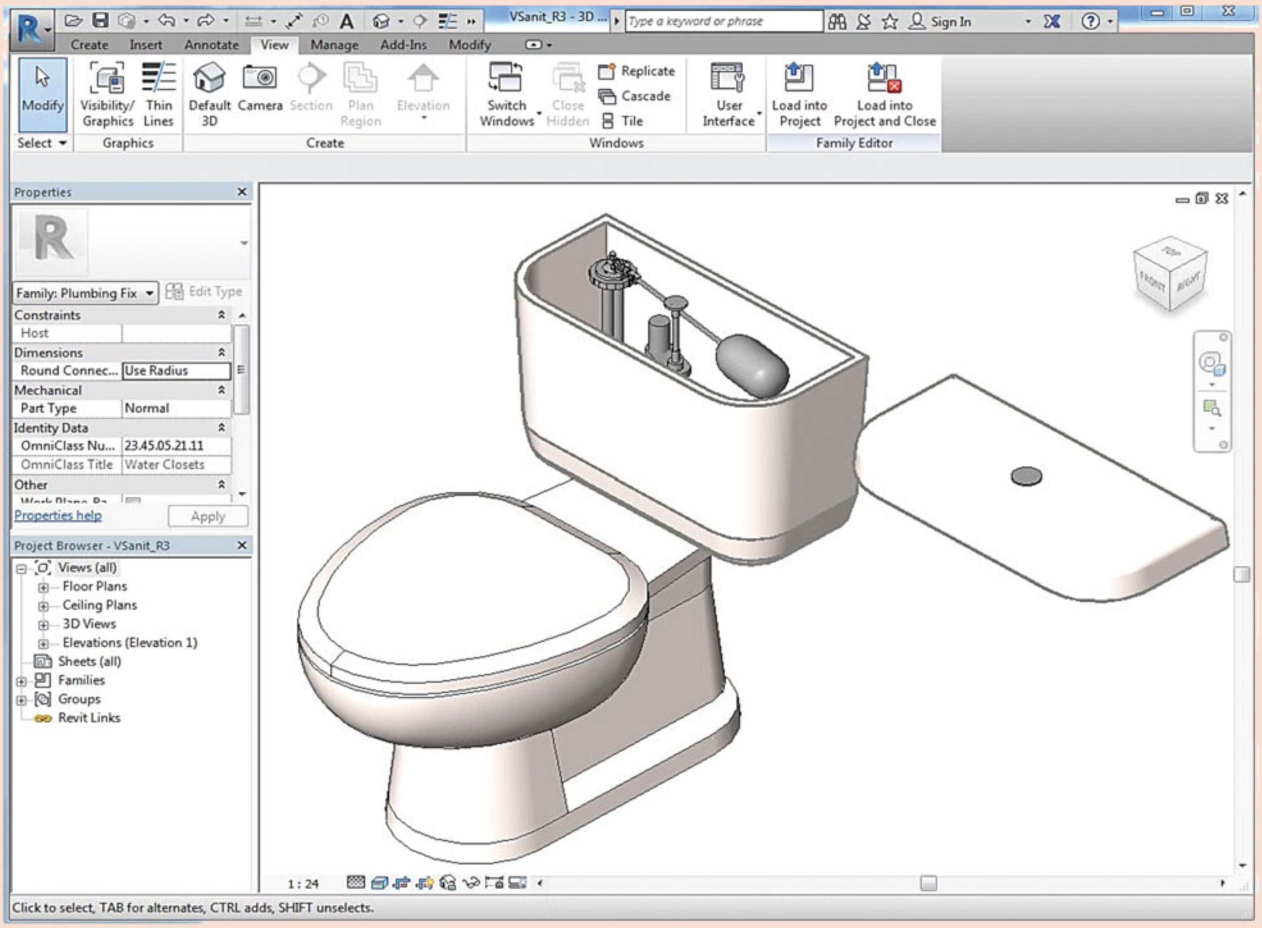


Fig. 21.6 BIM of the toilet in Revit

a b

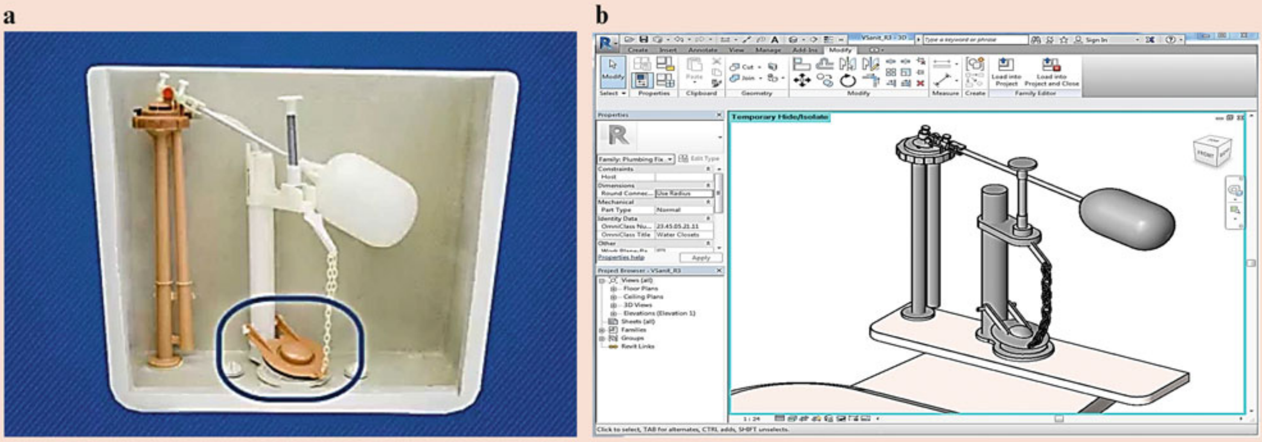


Fig. 21.7 Floater engine reference: Astra model (**a**) and internal engine breakdown in Revit (**b**)

to the removal of the model's non-geometric information. In the third step, the 3D model is exported to Maya to create the floater replacement animation in .3DS format. Here, the front face of the coupling box is made transparent in order to allow a better visualization of the internal components.

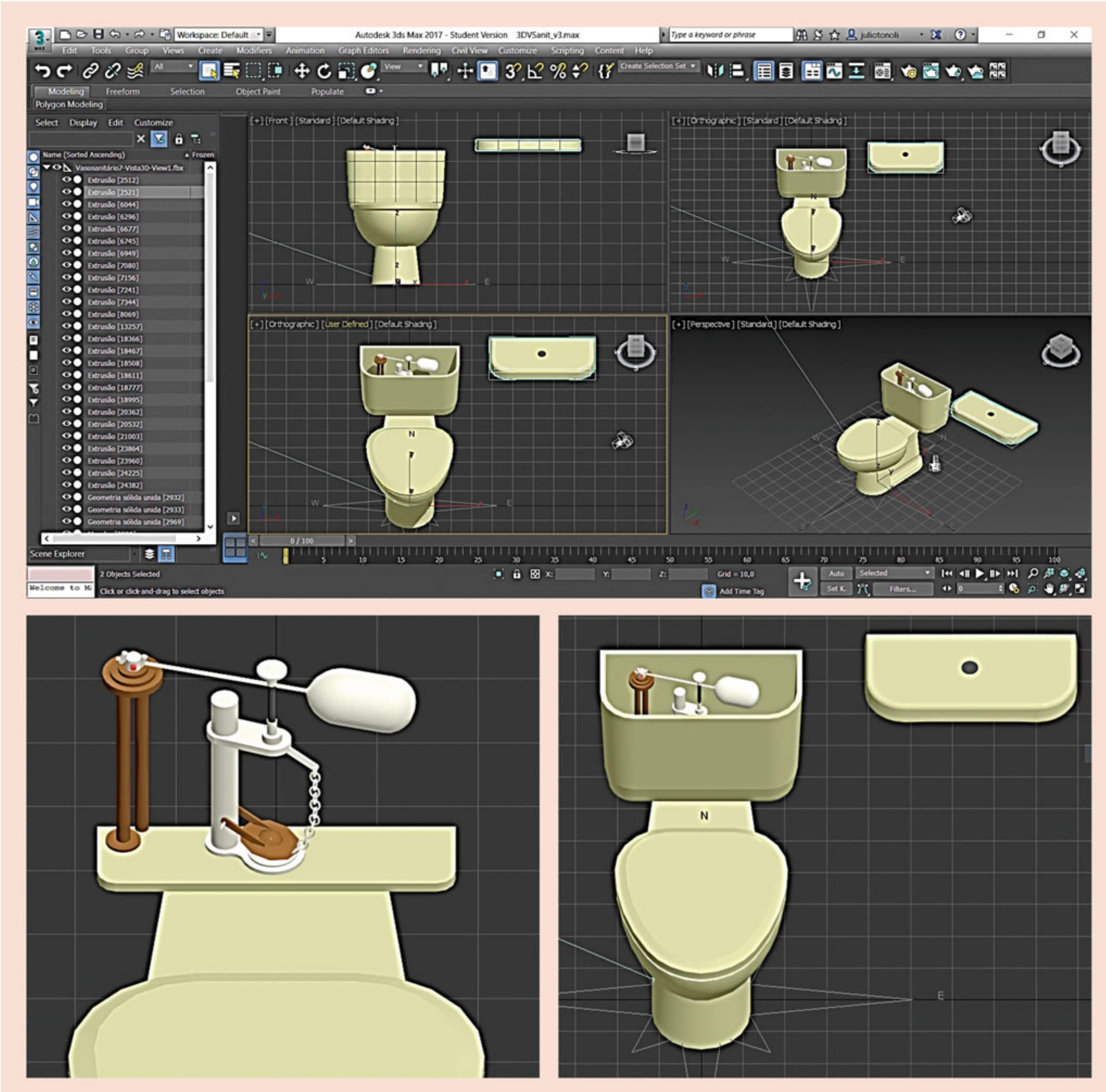


Fig. 21.8 Adding texture and materials in 3D Studio Max

In the fourth step, the user interface (e.g., interaction buttons, "Instructions" screen, icons), as well as markers, is created in Adobe Photoshop. In parallel, programming and inserting AR features into the 3D model using Unity/Vuforia is performed (fifth step). In particular, the animation is exported to Unity and Vuforia to create navigation and add AR features in .fbx format. When exporting the animation to Unity/Vuforia, some model textures and materials are lost, but new textures are inserted into the model to accommodate this loss of information.

Concerning marker design, the characteristics considered were asymmetrical pattern, high contrast, no repetition of patterns, and richness of detail. In particular, as shown in Fig. 21.9a, the image in the center of the magnifying glass represents the virtual content that is displayed in AR. On the other hand, the symbol under the magnifying glass identifies the action (i.e., maintenance, visualization). For example, the image of a hammer and a wrench corresponds to a maintenance activity, while the image of an eye corresponds only to information visualization. Similarly, the marker illustrated in

Fig. 21.9b is designed to display material or wall covering information (wall image in the center of the magnifying glass and an eye at the bottom) without requiring the user to perform any task. All markers have the same outline which resembles a mobile phone, indicating that they are intended to be viewed via mobile devices. Since, for both the LAR and MAR applications, the object of interest is toilet, the marker illustrated in Fig. 21.9a is used in both applications. The Vuforia Developer Portal Target Manager is used for marker classification using a five-point scale. In this rating system, one star corresponds to poor marker design quality, and five stars represent excellent marker quality. Figure 21.10 shows the results of marker classification, in which points used for feature tracking are highlighted in yellow color. According to the results, the designed marker is rated as five stars by the Vuforia Developer Portal Target Manager. Both LAR and MAR applications have been developed with the extended tracking feature, which utilizes environmental features to enhance crawl performance and maintain virtual object visualization, even when marker visibility is interrupted. This feature is especially recommended for architectural objects that are viewed in scale and perspective.

Figure 21.11 shows snapshots of the first LAR design test on the tablet computer. This test served the purpose of material adjustments (texture and transparency in model visualization) and refining the user interface (screens, step-by-step visualization of floater replacement, positioning of buttons). Figure 21.12 illustrates another LAR test on the tablet computer. Here, the AR model view is embedded in the real environment, and the scale of the virtual model is adjusted to match the real environment. Finally, ambient lighting interference checks and marker positioning are carried out.

Figure 21.13 shows the next step of the LAR tablet version, with a graphical interface and implemented interaction buttons inserted. In addition, to improve the visibility of the textual information at the bottom of the screen corresponding to buoy replacement, a blue banner with 70% transparency is added to the text background. Lastly, Figs. 21.14 and 21.15 show screenshots of the final version of the LAR tablet version.

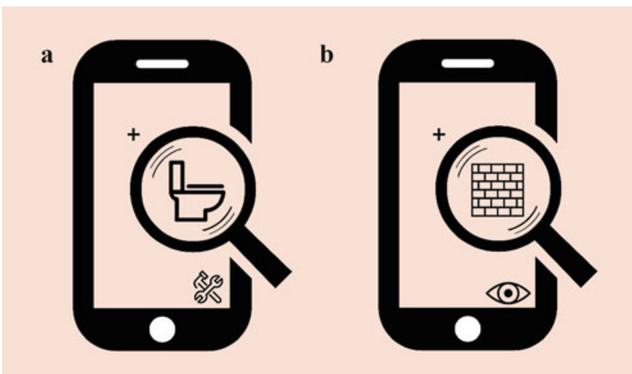


Fig. 21.9 LAR/MAR current (**a**) and future (**b**) markers

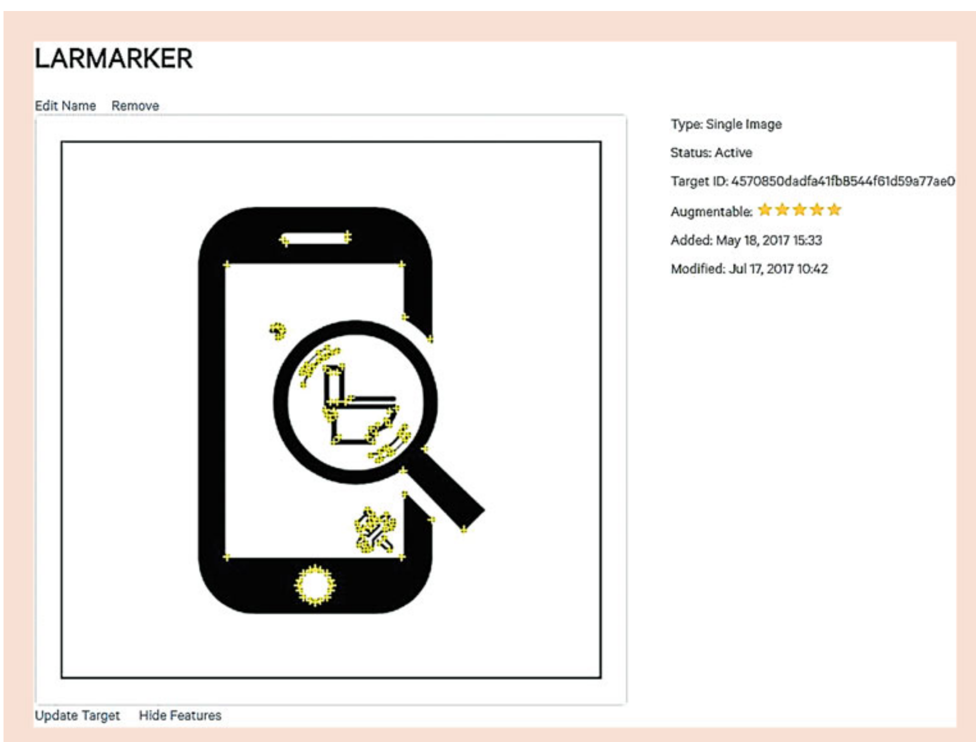


Fig. 21.10 LAR/MAR marker rating

Fig. 21.11 LAR tests, tablet version (material adjustment)

Fig. 21.12 LAR tests, tablet version (scale adjustment)

The design of the MAR application follows a similar procedure but requires adjustments to be made to some displayed textual information, scale and placement of the 3D model (model must be rotated to align with the printed marker),

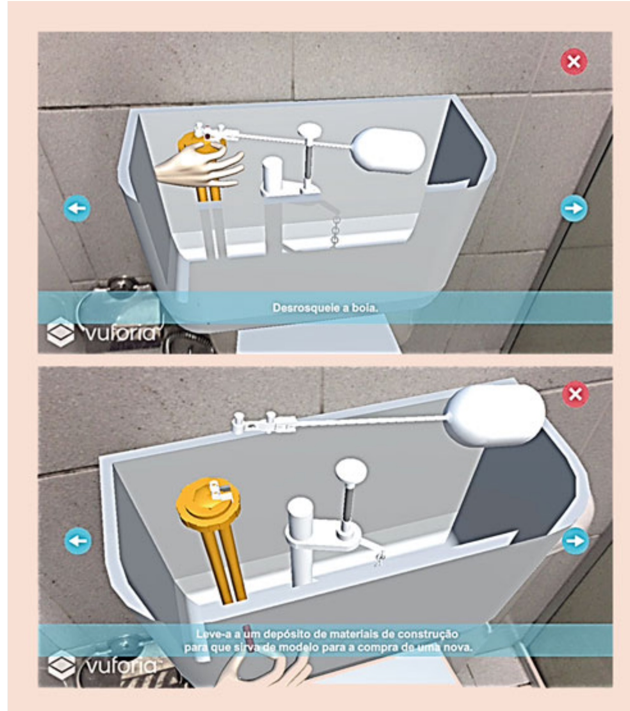


Fig. 21.13 LAR tests, tablet version (GUI adjustments)

and marker scale. The final version of the MAR application (tablet version) is illustrated in Fig. 21.16.

Similarly, the smart glasses versions of both applications are developed. Snapshots of the final version of LAR and MAR applications (smart glasses version) are shown in Figs. 21.17 and 21.18, respectively.

21.5.3 Artifact Publication

In this phase, both applications were made available for Android and iOS operating systems. To ensure that the developed applications meet the requirements of each system, they must first undergo a review and approval stage. Initially, the TestFlight tool (in iOS) was used to publish the trial (beta) version of applications by inviting select users. Upon receiving an email invitation, the user could download and run the application. The TestFlight tool was linked to iTunes Connect (Fig. 21.19). Publishing the applications in beta enabled testing, alongside making corrections and adjustments, and preliminary user evaluation.

In the Android system, the Google Play Console platform was used, which allowed application management and testing through a link sent to registered users (Fig. 21.20). A beta version of the application was published to check for errors and inconsistencies before official publication (Fig. 21.21).

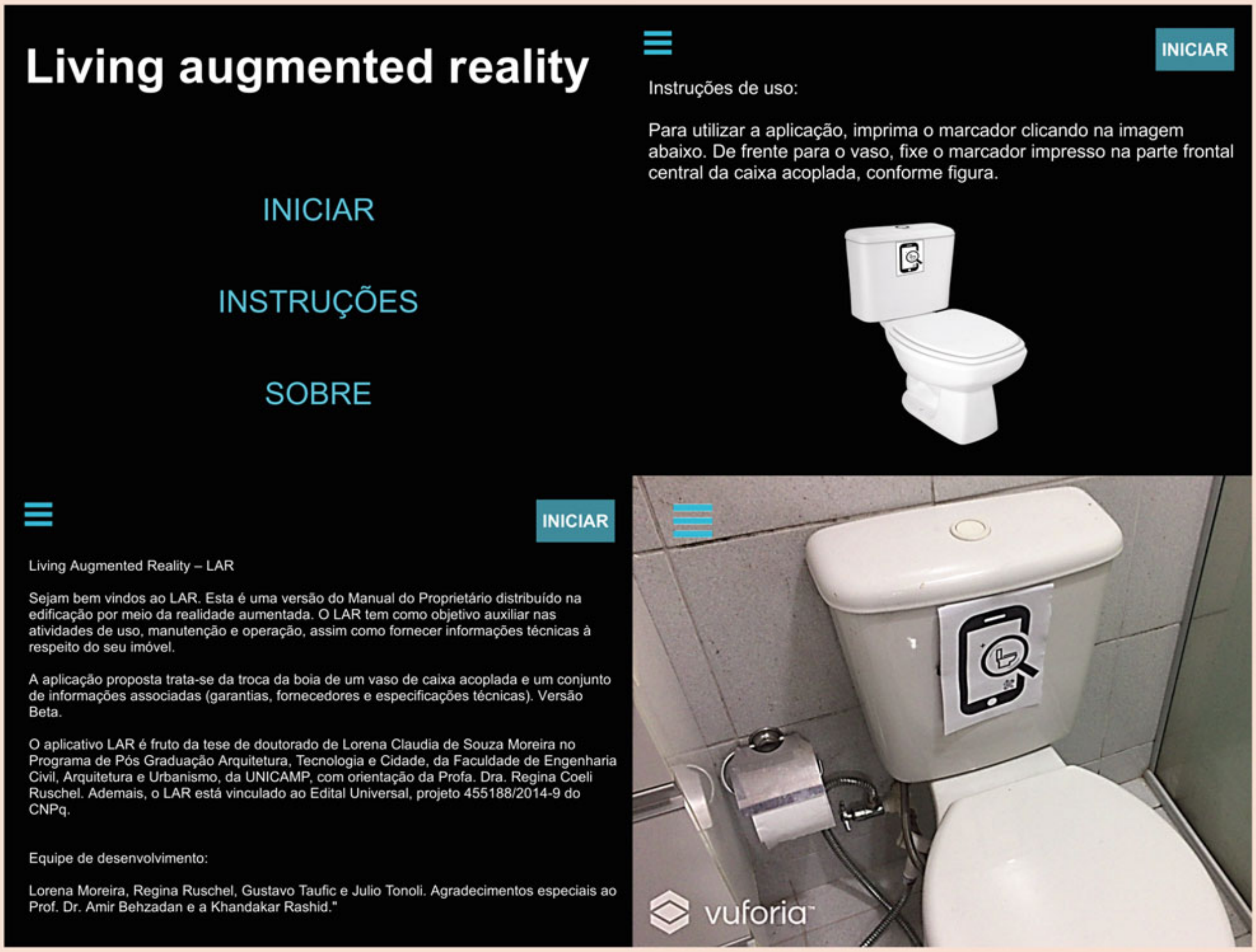


Fig. 21.14 Application screen of LAR tablet version

21.6 Evaluation and Findings

The artifact evaluation phase was carried out in a building material store (Fig. 21.22) and the School of Civil Engineering, Architecture and Urbanism – FEC (LAMPA Design Methods and Automation Research Lab) (Fig. 21.23). At the beginning of the session, each participant reviewed and signed an informed consent form and answered a profile identification questionnaire (age, gender, education level, and level of familiarity with the activity, manual, and technology). As shown in Fig. 21.24, participating individuals then completed a building maintenance activity (i.e., changing a toilet float mechanism) using information delivered by one of the five types of manual, namely, traditional (print) BMIM, MAR application (tablet version or smart glasses version), and LAR application (tablet version or smart glasses version). At the conclusion of the experiment, participants completed a NASA TLX questionnaire to report measures of mental demand, physical demand, temporal demand, performance, effort, and frustration. The activity lasted an average of 15–20 min, including the completion of the questionnaires. Each individual performed only one experiment. In total, data were collected from 100 participants, 20 individuals in each experiment (information delivery system): (A) traditional BMIM, (B) MAR application, and (C) LAR application. All individuals were able to complete the assigned activity.

21.6.1 General Workload Measurement: NASA TLX

For workload measurement, the following evaluations are performed: (i) characterization of the total sample, (ii) workload and factor analysis, and (iii) workload analysis considering the perception filters.

Total Sample Characterization

To characterize the entire sample of 100 participants, the profile identification questionnaire was adopted, which asked

Fig. 21.15 AR view screens of LAR tablet version (https://youtu.be/7rDDuKeEiwk)

for information such as age, gender, education level, and level of familiarity with the activity, manual, and technology. The age range of the participants is illustrated in Fig. 21.25a, which reveals that 42% of the sample is 18–24 years old, 15% of individuals are 25–29 years old, 10% are 30–34 years old, 8% are 35–39 years old, 8% are 40–44 years old, and the remaining are older than 44 years (including 2% who are 60 or older). Although more than half of the participants (57%) are between 18 and 29 years old, the sample covered all age groups. Also, as shown in Fig. 21.25b, of the entire sample, 42% are female, and 58% are male.

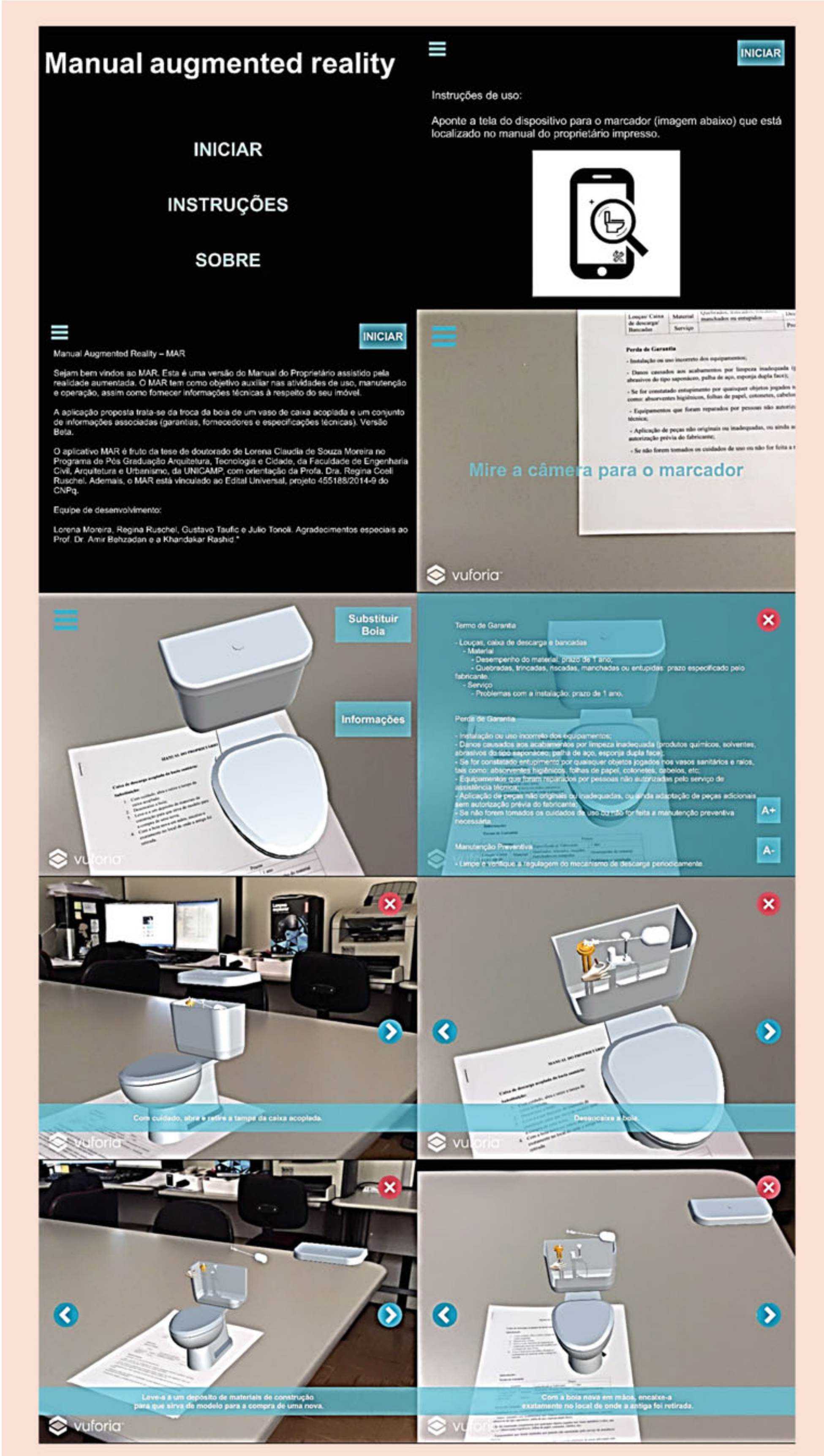


Fig. 21.16 Application screen of MAR tablet version (https://youtu.be/T90HXxrV5yU)

Fig. 21.17 Application screen of LAR smart glasses version

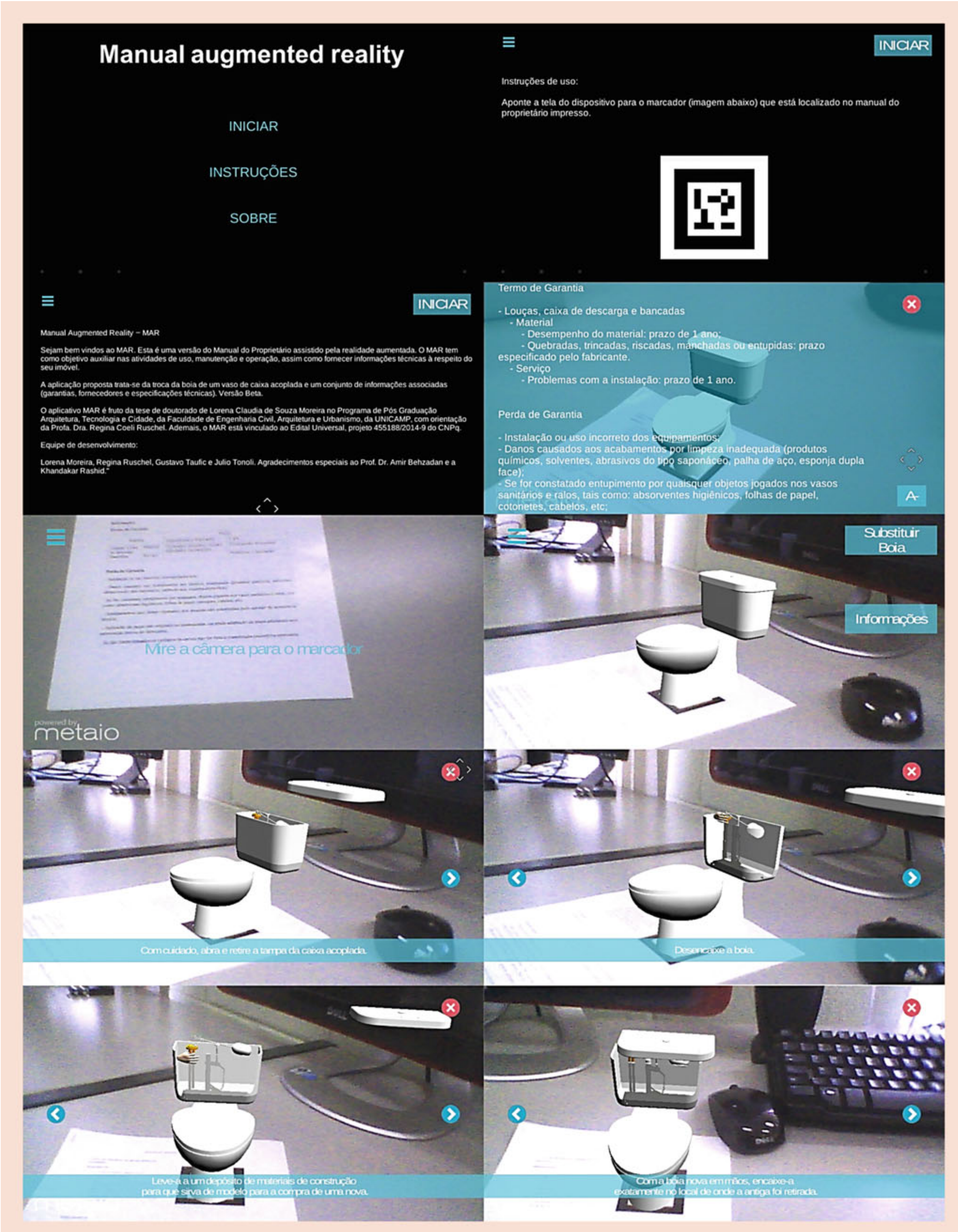


Fig. 21.18 Application screen of MAR smart glasses version

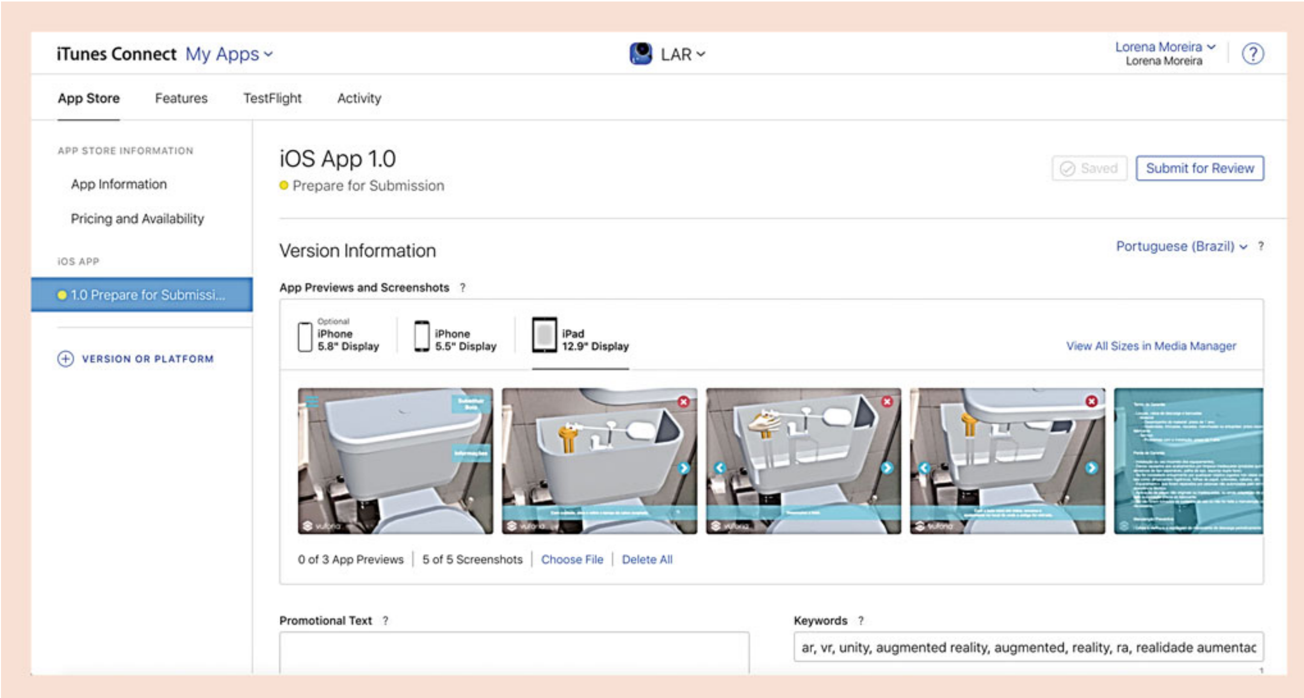


Fig. 21.19 Application publishing of LAR tablet version (for iOS)

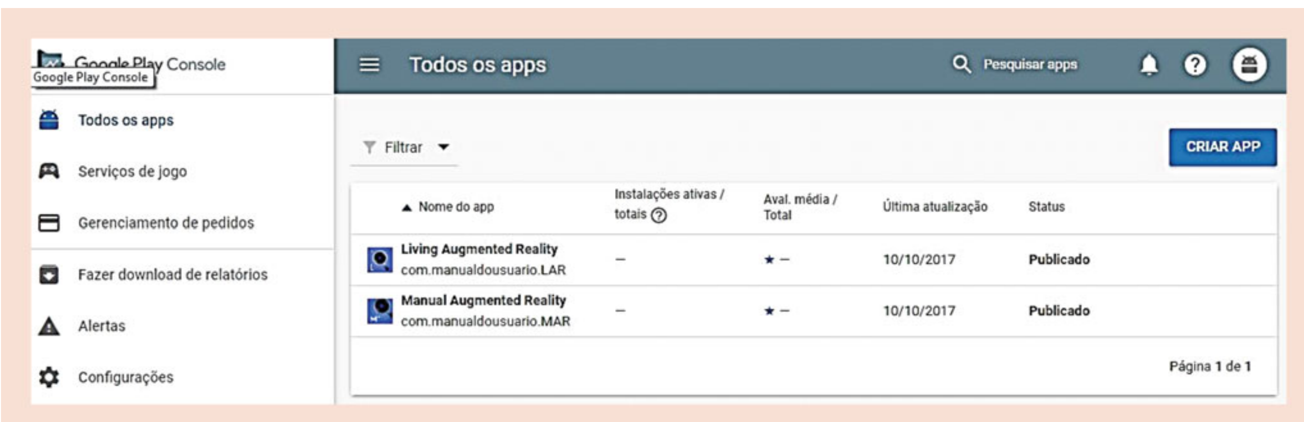


Fig. 21.20 Application publishing of LAR and MAR tablet version (for Android)

Also, according to Fig. 21.26, the analysis of participants' education levels reveals that 2% of the individuals have incomplete primary or secondary education, while 12% have completed high school. Fifty percent have attended undergraduate-level classes, but only 19% have completed this level of education. Finally, 8% answered that they have completed undergraduate studies with professional specialization, and 9% have completed graduate-level studies. Although the participants presented all levels of education, more than half have incomplete or complete higher education level.

Moreover, participants indicated their level of familiarity with the activity by answering a question about whether or not they have performed the same task before. For this question, 66% of individuals stated that they had not exchanged a toilet coupling mechanism before, and 27% indicated that they had performed this activity in the past. The remaining 7% had observed someone else performing this activity (Fig. 21.27a). Similarly, when asked if they had ever consulted a BMIM, 82% of the individuals answered that they had not done so, while only 18% had used such manual before. This data reveals that the vast majority of participants were unfamiliar with a BMIM (Fig. 21.27b). Finally, when asked about their familiarity with technology, 96% of users answered that they are familiar with a tablet computer, while 41% had familiarity with AR applications, and 24% were familiar with smart glasses (Fig. 21.27c).

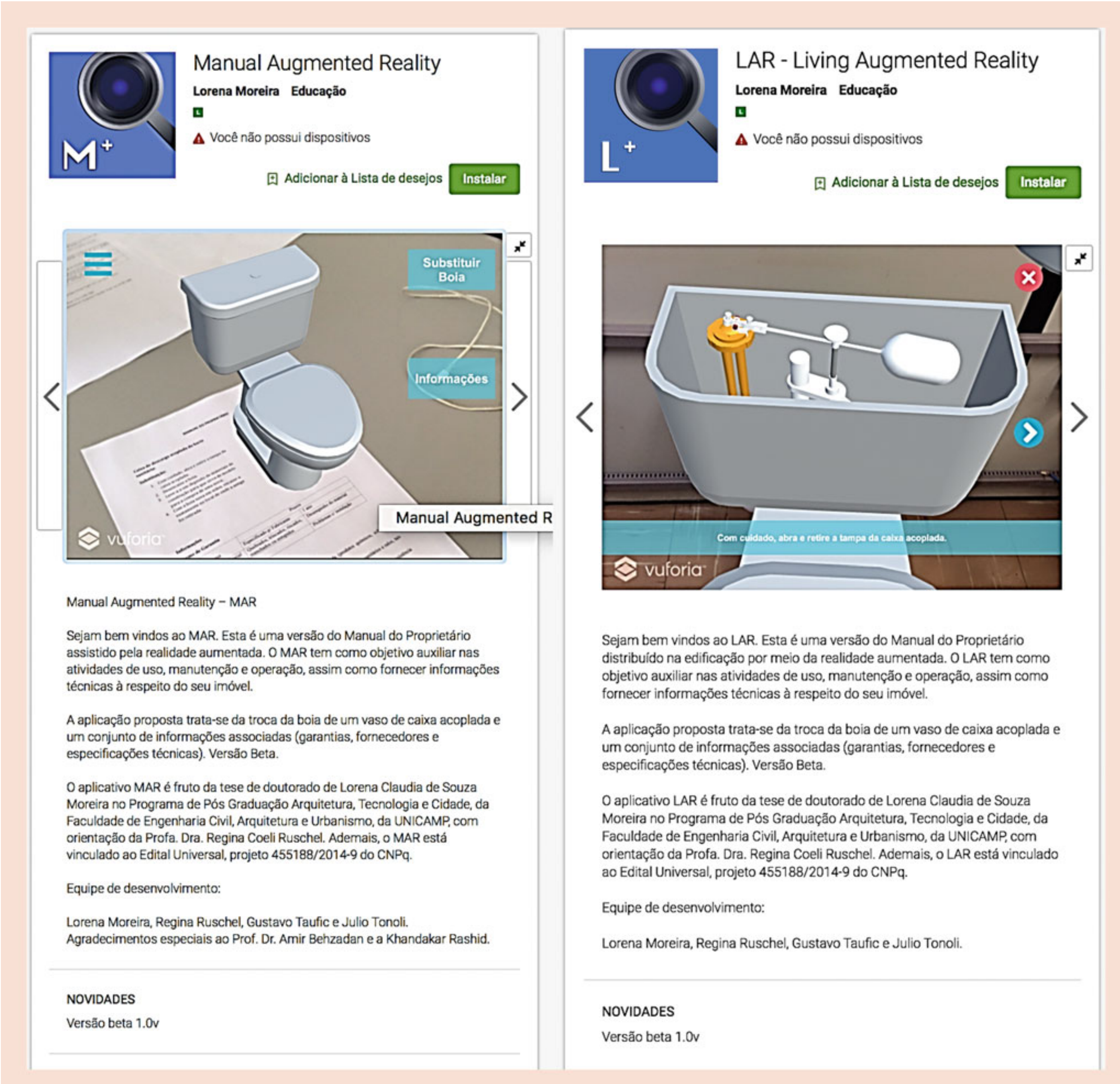


Fig. 21.21 Application testing of LAR and MAR tablet version

In summary, most of the sample is composed of 18- to 24-year-old male individuals with incomplete higher education who have never performed the activity nor have they consulted a BMIM. Also, the majority have experience with tablet devices, are relatively familiar with AR applications, and have no mastery of AR glasses.

Workload and Factor Analysis

Initially, total workload (TW) was considered comparatively among the five experiments (Fig. 21.28). The experiment that achieved the highest workload was the one with the task supported by the traditional manual with a TW of 35.7 points, followed by the paper-based BMIM plus AR (MAR application) visualized with smart glasses with a TW of 32.7 points. The manual incorporated into the environment viewed with AR with smart glasses (LAR application) achieved a TW of 29.5 points, followed by the manual incorporated into the environment with AR viewed on tablets (MAR application) with TW score of 28.5. Finally, the best performance was achieved with paper-based BMIM plus AR (MAR application) visualized with tablets with a TW of 26.0 points.

Fig. 21.22 Application evaluation at the building material store

These values indicate that experiments conducted with MAR and LAR applications using tablet devices have reported a lower workload than experiments using smart glasses. Therefore, it can be inferred that the way AR is visualized or the level of familiarity with the visualization device influences user performance.

In turn, individuals who used the BMIM on paper plus AR (MAR application) and the BMIM built into the environment with AR (LAR application), regardless of the device type used, were charged less than individuals who used the traditional BMIM. This demonstrates the better performance of the BMIM when assisted by AR.

Comparison of all factors (i.e., mental demand, physical demand, temporal demand, performance, effort, and level of frustration) of each experiment reveals that the mental demand of the BMIM on paper plus AR (MAR application) with smart glasses achieved the highest index (175.4 points) (Fig. 21.29). This could be due to the limitations of the smart glasses used. The Moverio BT smart glasses does not have a camera that produces good image quality, and the small size of the virtual model makes it difficult to view the task of changing the float.

On the other hand, the mental demand that reached the lowest value corresponds to the BMIM incorporated in the environment with AR (LAR application) viewed on a tablet device. The decisive factor is the visualization of the virtual model in 1:1 scale overlaying the real model, allowing an immediate association of the task that must be performed by the user, which confirms the influence of the type of visualization on user performance. Besides, the camera and

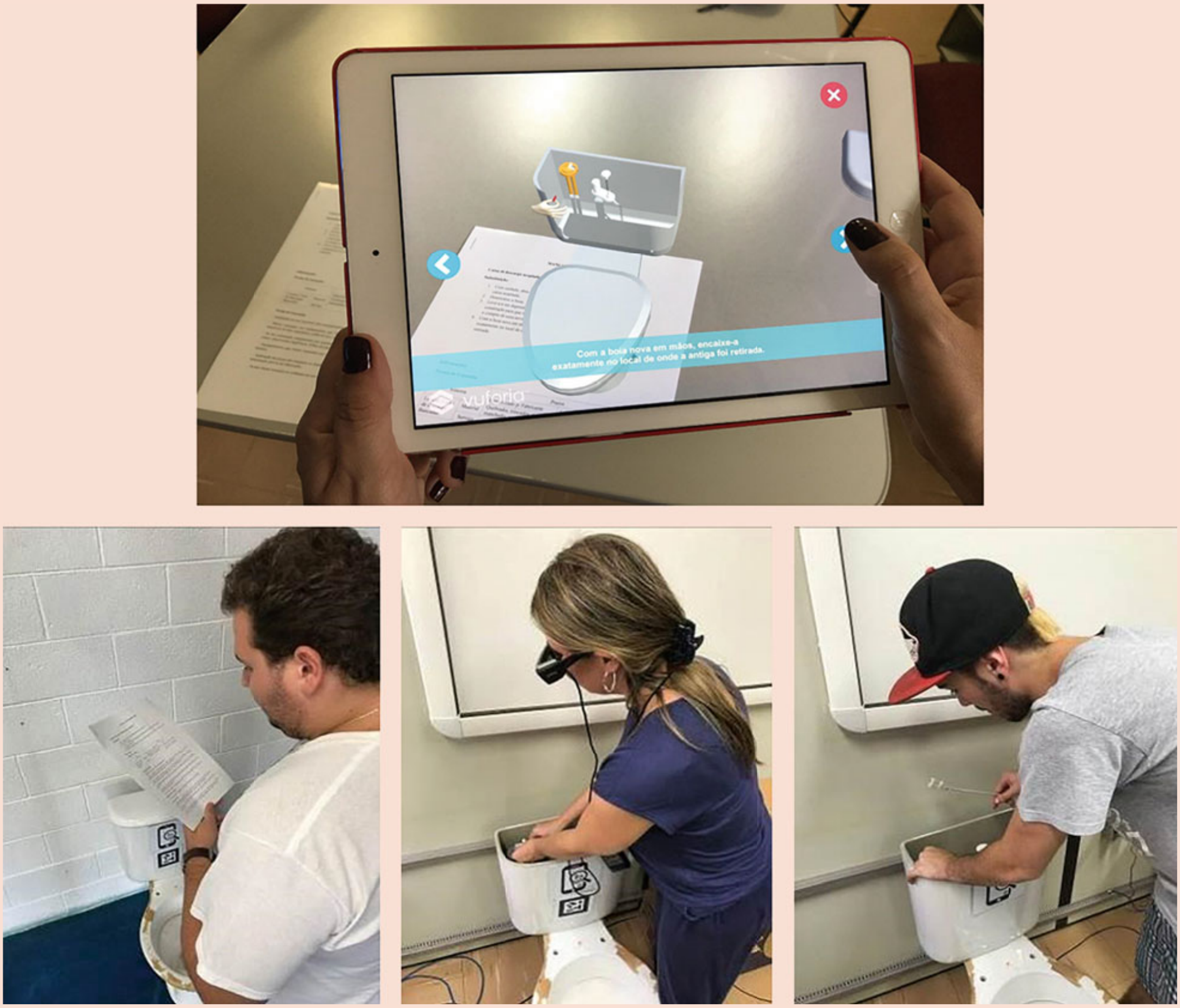


Fig. 21.23 Application evaluation at the university campus

screen of the tablet device have good quality. Comparing all five factors, it is observed that mental demand was the factor that obtained a higher degree of importance from users.

Regarding physical demand, it is observed that the paper BMIM achieved the lowest score with 46.4 points, while the BMIM incorporated in the environment with AR (LAR application) viewed on a tablet device reached 77 points. It is inferred that the paper manual performed better in terms of physical demand since it did not require users to handle a peripheral display (Fig. 21.29).

In the analysis of temporal demand, the BMIM incorporated in the environment with AR (LAR application) viewed on a tablet device reached the lowest index with 51.9 points. This demonstrates that users are faster in visualizing the activity when the virtual content is superimposed on the real object. Already the traditional BMIM plus AR (MAR application) with smart glasses reached the highest temporal demand (85.9). It can be said that viewing the object on a smaller scale and outside the physical location of the task interfered with the time spent by users (Fig. 21.29).

In the performance analysis, individuals stated how satisfied they were with performing the task. A lower score in this case corresponds to a better performance and vice versa. The traditional BMIM plus AR (MAR application) on tablet scored 41.9 points, showing a better performance. In comparison, the traditional manual achieved the worst performance index with 148.8 points. The paper BMIM plus AR (MAR application) viewed with smart glasses and the BMIM built into the environment with AR (LAR application) with tablet version scored very close, with 83.5 and 87.3 points, respectively (Fig. 21.29).

Fig. 21.24 Step of float replacement

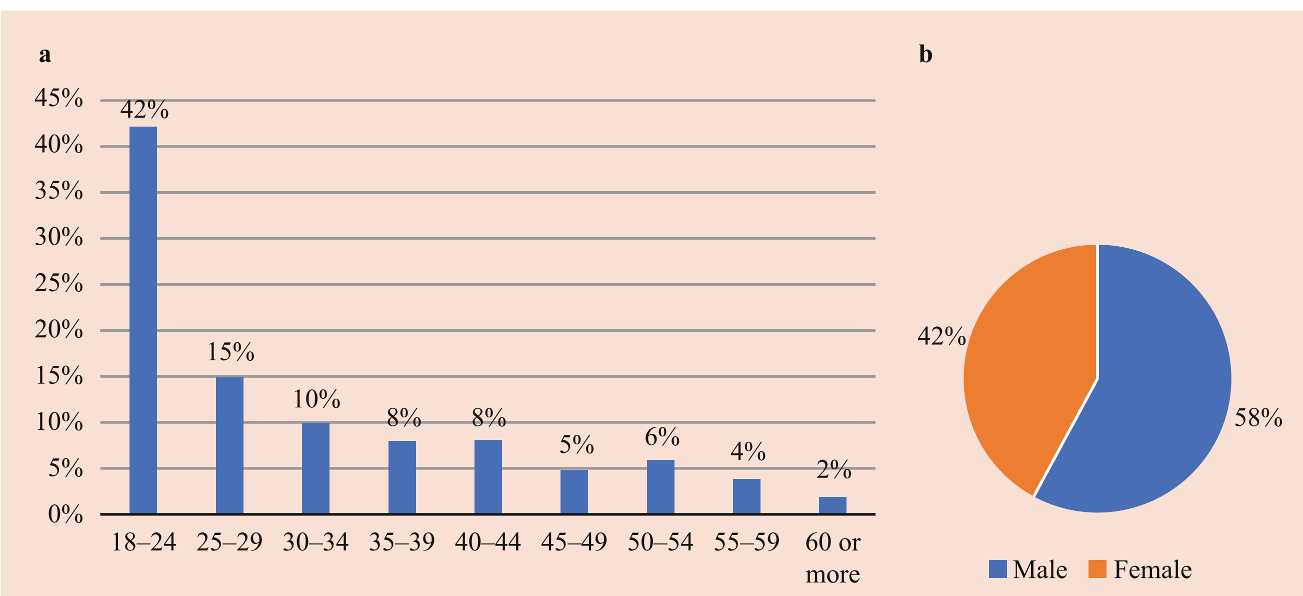


Fig. 21.25 Age range of participants (**a**) and gender distribution (**b**)

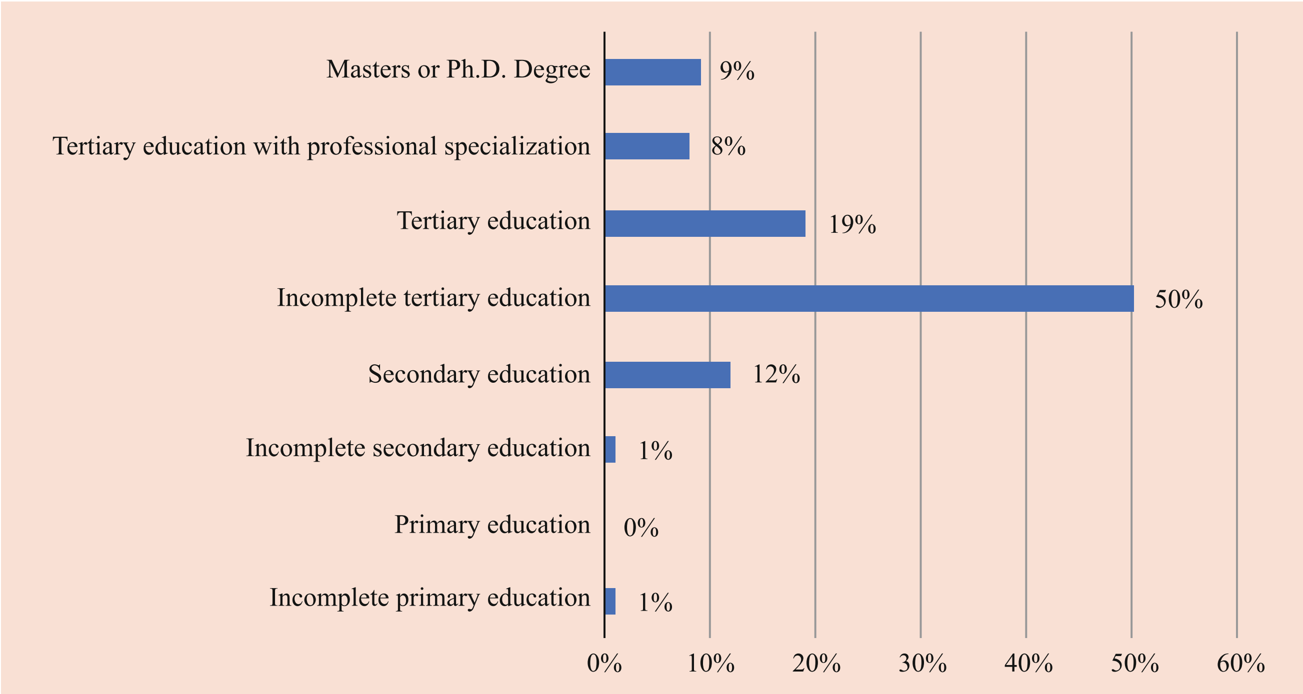


Fig. 21.26 Participants' education level

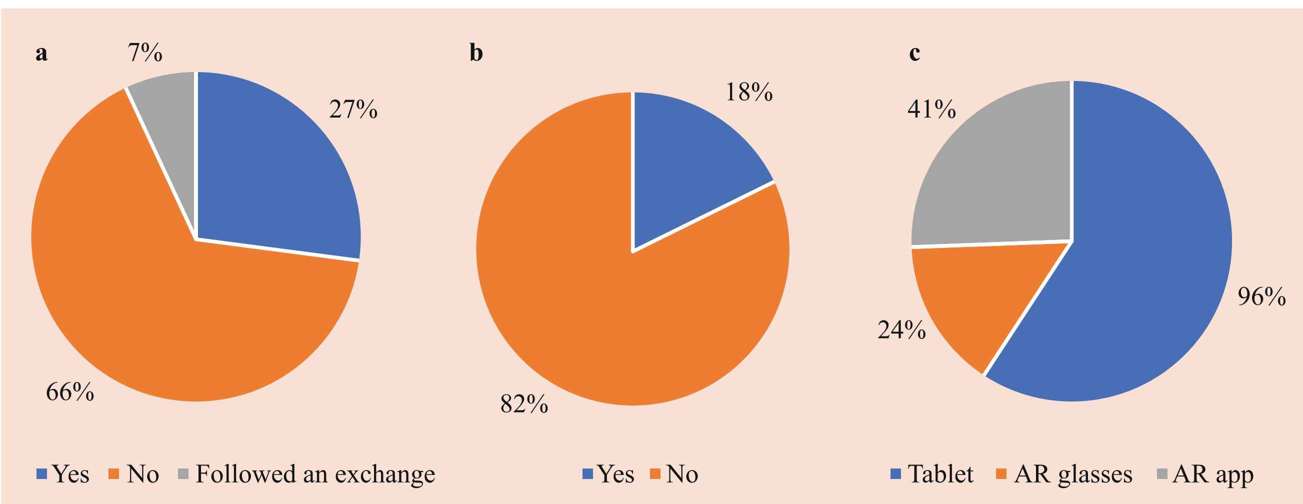


Fig. 21.27 Level of familiarity with the activity (**a**), BMIM (**b**), and AR technology (**c**)

In the analysis of the effort factor, the traditional manual obtained a score of 54.3, and the paper manual plus AR (MAR application) viewed with tablet and smart glasses scored 54.8 and 59.1, respectively. Meanwhile, BMIM built into the environment with AR (LAR application) viewed with smart glasses achieved 64.4 points. However, BMIM built into the environment with AR (LAR application) viewed with tablet presented the highest effort score of 82.5 points. As the traditional manual does not require the handling of peripheral display devices, it can be said that working with a traditional BMIM should lead to the lowest effort score. However, the difference in score among all groups was not significant (Fig. 21.29).

In the analysis of the level of frustration, the traditional manual was the one that most frustrated individuals in performing the task, as opposed to individuals who used the AR technology (Fig. 21.29). Participants who used the traditional BMIM achieved the highest score with 50.9 points, while this score for those who used AR was significantly lower (more than half, in some cases).

In summary, the analysis of the TW and individual workload factors shows that the paper BMIM plus AR (MAR application) viewed on tablet is the best solution, reaching the lowest TW score. However, it is noteworthy that the score difference for the second and third place is not significant. Thus, this analysis also indicates the potential of the BMIM incorporated into the AR environment. In contrast, using the traditional manual leads to the worst performance. Therefore, we conclude that incorporating AR in BMIM stimulates gains regarding the proper use of the manual.

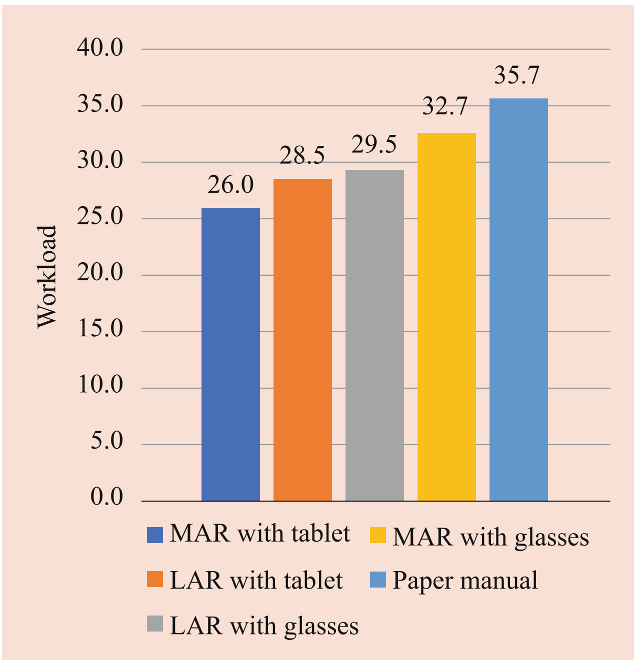


Fig. 21.28 Total workloads

21.6.2 Workload Analysis Considering Perception Filters

The third type of analysis considers the influence of perception filters on the calculated workload. Perception filters applied were age, gender, education level, and familiarity with the maintenance task performed, AR technology, and associated devices. With respect to the representativeness of the experiment, we considered the reference group to be the Brazilian population (208,317,492 people), and thus the sample size of 100 individuals and reliability of 90% impose a margin of error of 8.25% on the results. The coefficient of determination between the workload and age in a linear regression is 0.469 (Fig. 21.30). Therefore, it can be inferred that the age filter has a probable influence on workload perception. In particular, the older the individual executing the task with the support of AR, the smaller the perceived workload.

As for the gender filter, it is observed that the TW of female participants was 38.87, which is higher than the average value, while the TW of male participants was 27.97 and below the average value. Although these values are close, this difference leads us to believe that performing the task was easier for males than it was for females (Fig. 21.31), indicating gender influence on the execution of the task supported by AR.

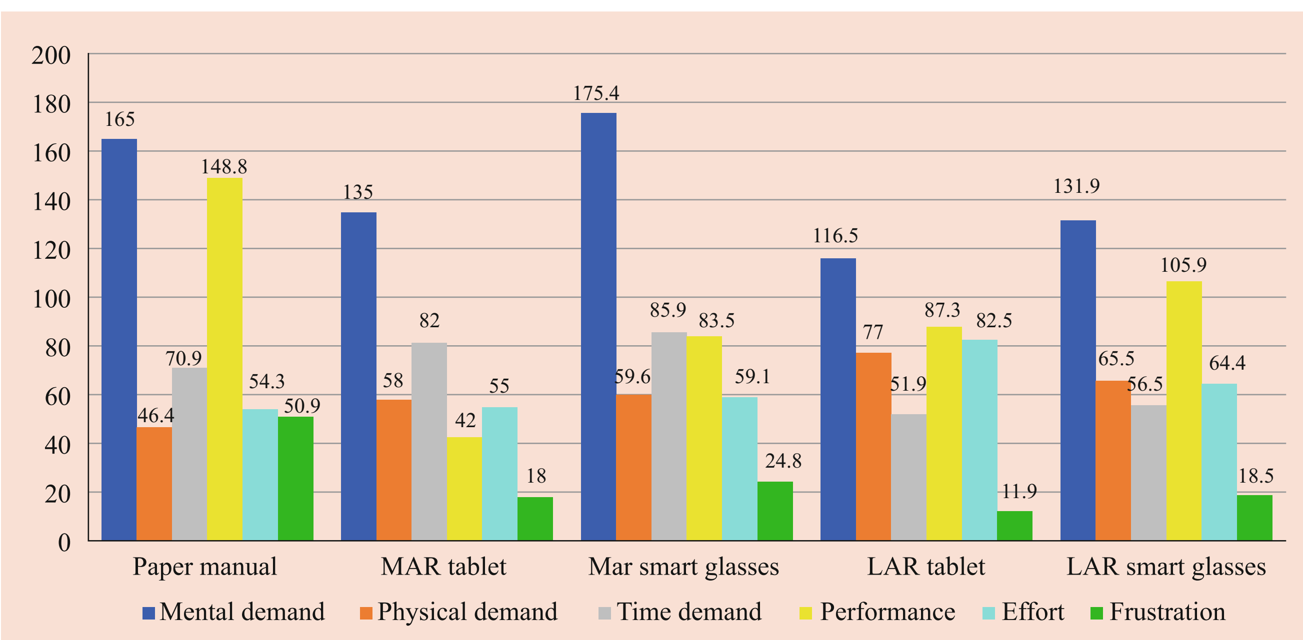


Fig. 21.29 Workload considering demand factors

Fig. 21.30 Average workload according to age filter

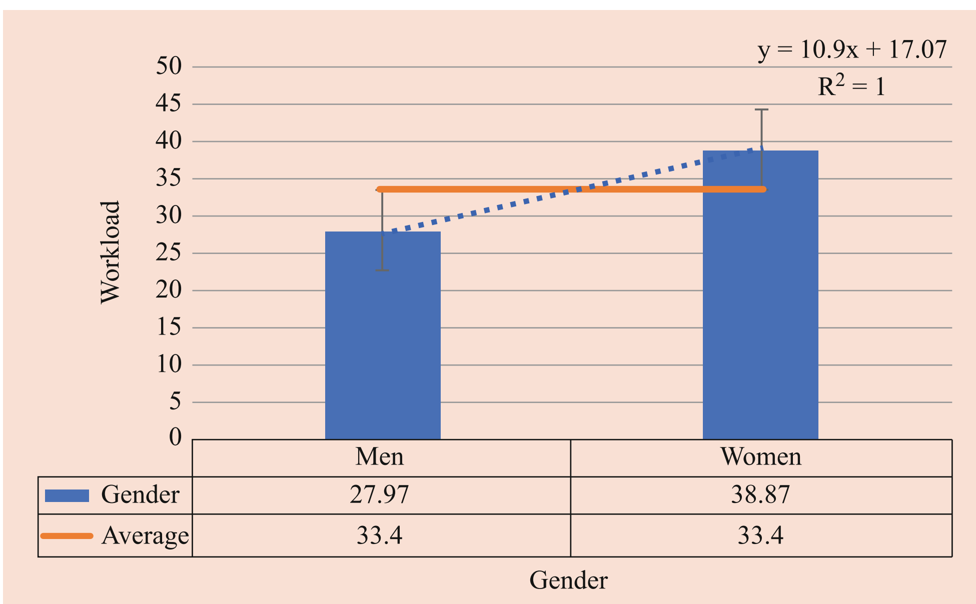


Fig. 21.31 Average workload according to gender filter

Considering the level of education, individuals with high school education presented a workload average of 34.61 which is higher than the overall sample average of 30.48. In comparison, individuals with incomplete high school education and those with incomplete elementary school obtained a TW score of 2.50 and 15.67, respectively, which is lower than the overall average of the sample. Individuals with other levels of education, however, remained close to the overall average (Fig. 21.32). Thus, this analysis allows us to infer that for individuals with incomplete high school and incomplete

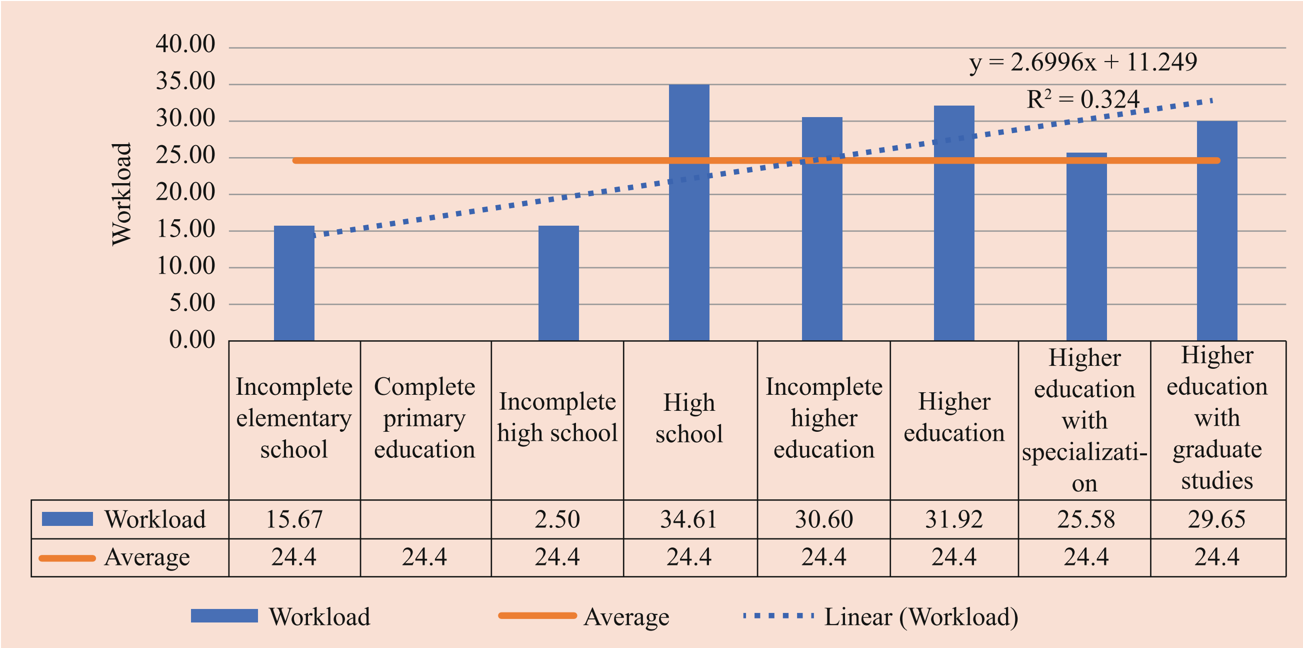


Fig. 21.32 Average workload according to level of education filter

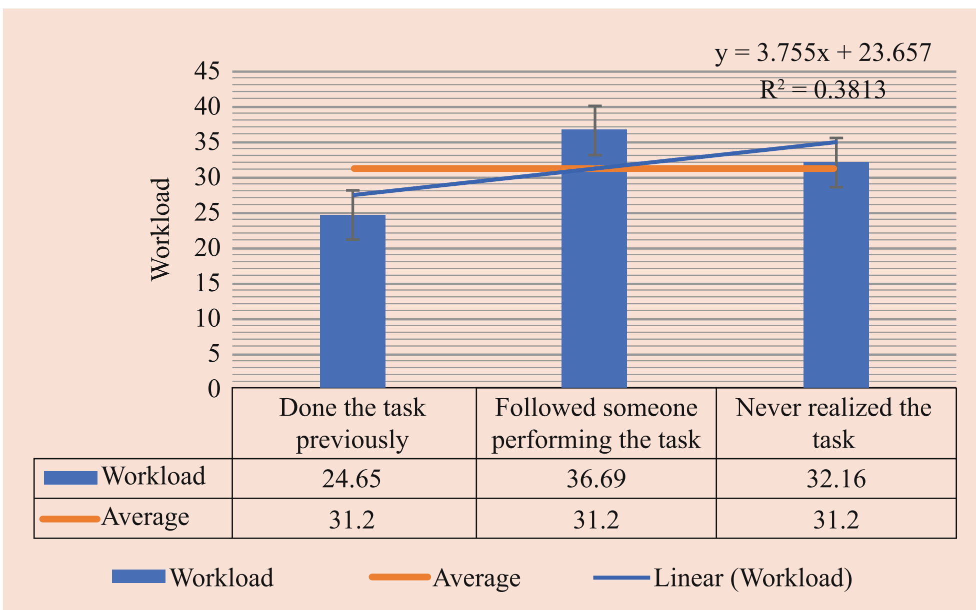


Fig. 21.33 Average workload according to previous task knowledge filter

elementary school education levels, the task presented greater ease than for other participants. However, it must also be noted that the presence of participants with this level of education was not uniform among experiments.

Regarding the level of familiarity with the activity of float exchange, it is observed that individuals who have previous knowledge of this activity have a lower workload (24.65) than those who have never performed the task before (32.16). Also, individuals who had previously observed someone else performing this activity scored above average workload (36.69). The interesting observation here is that second-hand knowledge (gained through observing the activity performed by someone else) has no positive influence on the participant's performance, leading us to believe that first-hand knowledge (gained through self-practice) has a positive influence on workload (Fig. 21.33).

Considering previous experience with BMIM and its relation to the average workload, it appears that individuals who

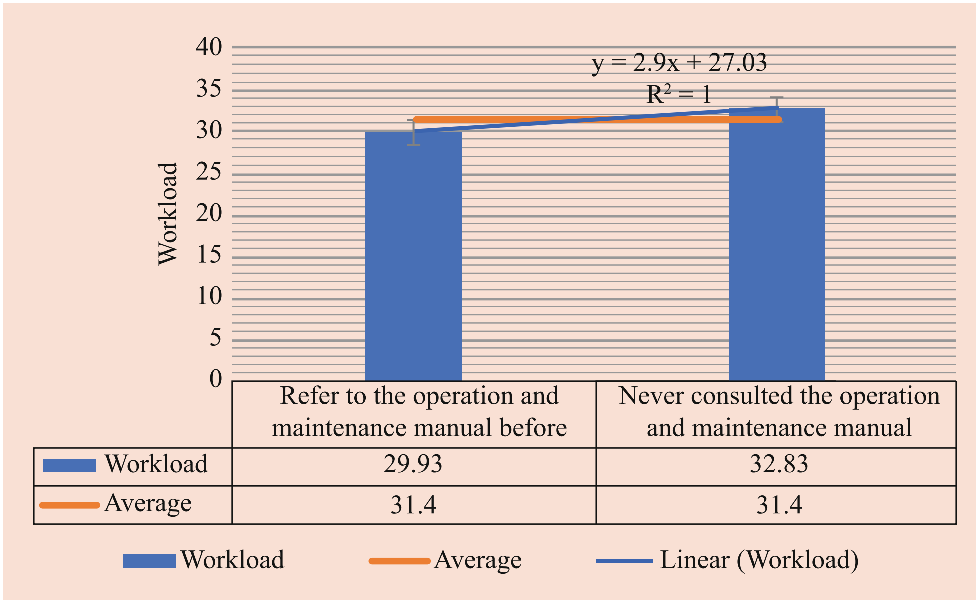


Fig. 21.34 Average workload according to BMIM consultation filter

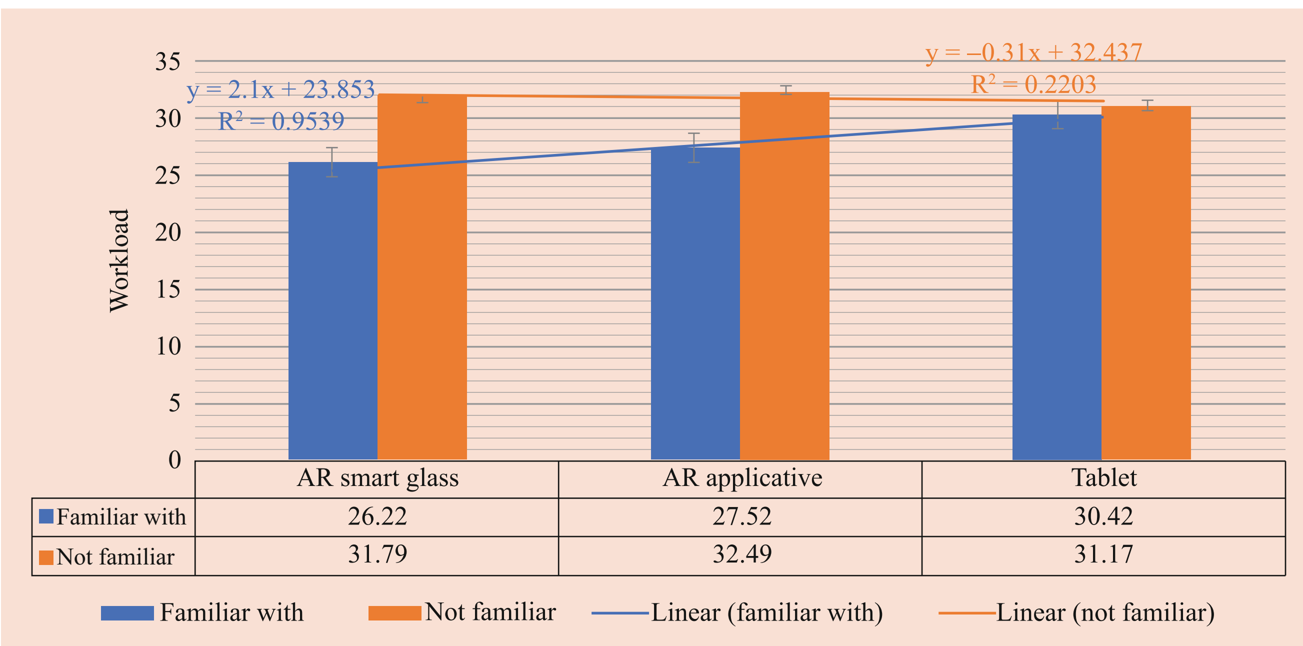


Fig. 21.35 Average workload according to device familiarity filter

never consulted the manual reached a lower workload (29.93) than those who had previously consulted the manual (32.83). However, we observe that the workload values are very close. Thus, it can be inferred that prior consultation with the BMIM does not significantly influence performance (Fig. 21.34).

Finally, considering the level of familiarity with the technology, we found that individuals who have experience with AR glasses and those familiar with AR had a lower workload (26.22 and 27.52, respectively) than those who had familiarity with tablets (30.42). Even though the values are very close to the average workload of 30.48, familiarity with the AR technology seems to reduce workload (Fig. 21.35).

21.7 Generalization

The proposed artifacts apply to a class of problems that aim to deploy AR specifically for assembly, maintenance, and instruction tasks. This study finds that the incorporation of AR technology can make a significant contribution to corrective and preventive maintenance, described in NBR 5674 [100]. The artifacts derived from the DSR method are presented as constructs, a model, a method, and prototypes. Constructs are the parts of the model that schematically represent the application. The model traces the relationships between the constructs, which are the maintenance component, the AR marker, and the AR visualization, as shown in Fig. 21.36. Methods describe the steps necessary to develop the AR application (Fig. 21.37). Figure 21.38 presents the two prototypes resulting from the model and method developed in this research.

Figure 21.39 presents the identification of possible points of incorporation of AR acting as a BMIM facilitator. Items marked with blue magnifying glass symbol indicate that AR can enhance the existing information by overlaying virtual objects. Items marked with gray magnifying glass symbol indicate that AR can enhance the existing information by overlaying textual information. Whenever the component demands an instructional task, we proved the benefit of accomplishing that task by incorporating AR visualization.

The proposed general model for incorporating AR into a task (Fig. 21.36) has been mapped throughout a complete BMIM considering the identified insertion points (Fig. 21.39). This scheme guides future implementations of AR in all parts of BMIM (Fig. 21.40). To achieve this goal, markers, sensors, or other types of tracking devices can be embedded in different building components allowing seamless presentation of information and instructions from BMIM throughout the building.

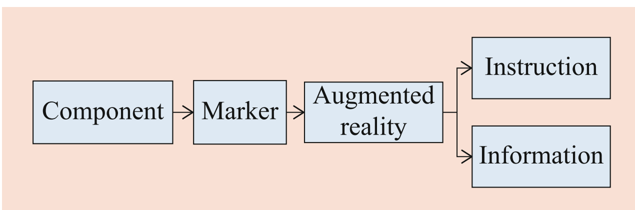


Fig. 21.36 General model by activity

21.8 Conclusion

Considering the three forms of evaluation performed, by measuring the workload presented, it is concluded that the proposed artifact met the desired requirements for its appli-

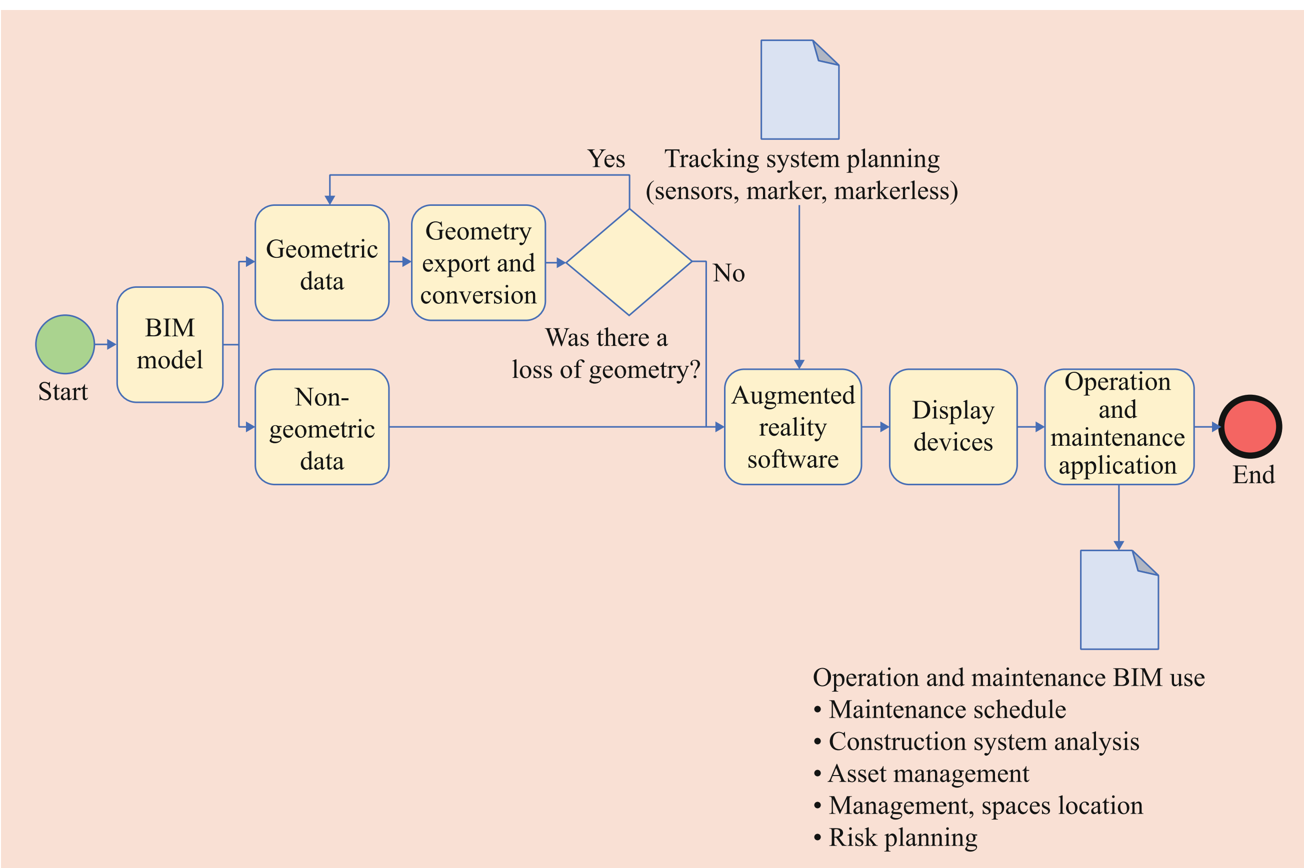


Fig. 21.37 BIM and AR model process applied to operation and maintenance

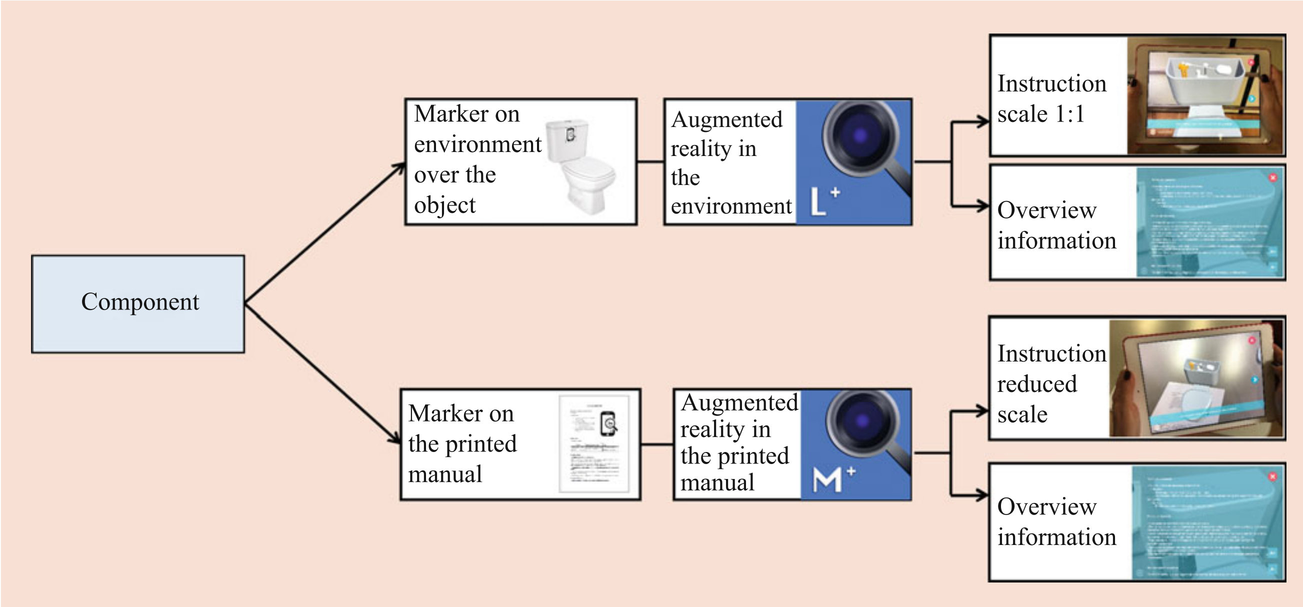


Fig. 21.38 AR application prototypes developed and evaluated in this study

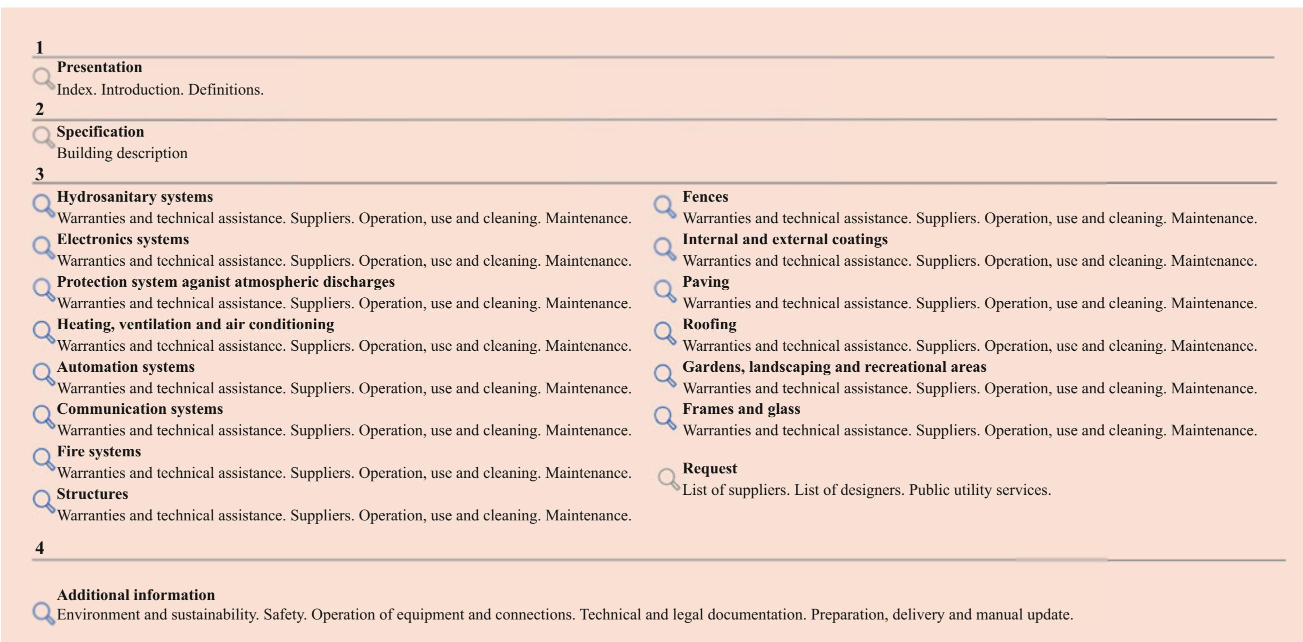


Fig. 21.39 AR incorporation points into to the complete BMIM

cation. In the analysis of workload and associated factors, the BMIM in paper plus AR viewed on a tablet device was identified as the best solution, while using the traditional BMIM led to the worst performance. Therefore, it was found that the insertion of AR visualization in BMIM can stimulate gains (i.e., better performance) regarding the use of the manual.

Regarding the analysis of the workload factors, the traditional BMIM plus AR viewed on tablet achieved the best ratings for performance and frustration. The traditional BMIM (with no AR) achieved the best performance in terms of physical demand, and BMIM incorporated in the environment with AR (LAR application) viewed on tablet achieved the best temporal and mental demand scores. As for the worst performances, the traditional BMIM assumed this rating in frustration level and performance, and BMIM incorporated in the environment with AR viewed on tablet was the worst in physical demand. The manual incorporated in the environment viewed with smart glasses reached the worst rating in

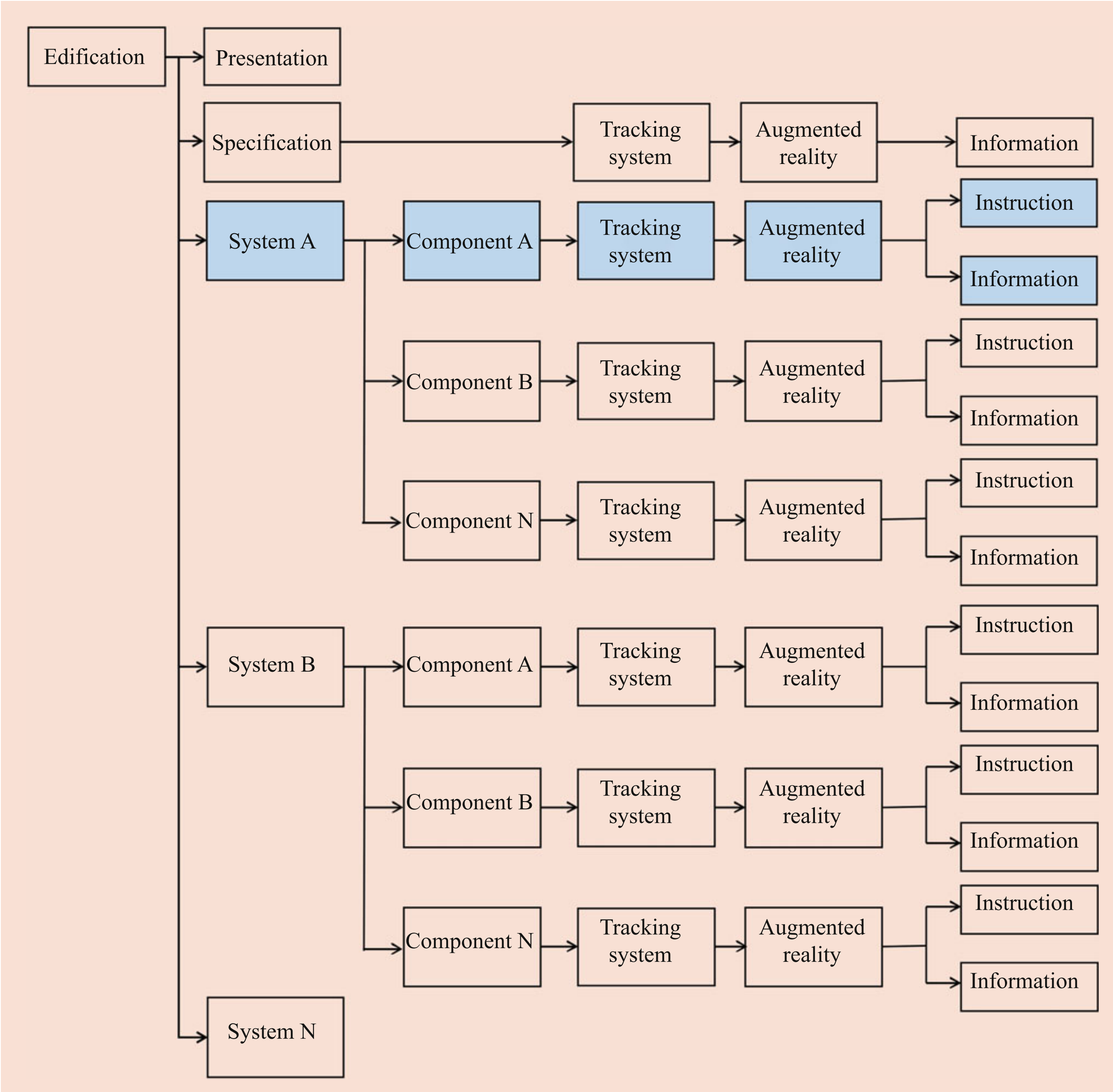


Fig. 21.40 AR incorporation framework throughout BMIM

mental demand, temporal demand, and effort. This analysis highlights the areas of improvement for future technology development to maximize user performance and stimulate gain.

In the context of workload analysis considering perception filters, we observed that, in general, age may influence workload when task is performed with the support of AR. It was also found that male participants found the task to be easier and individuals with lower education levels showed more positive workload influence. Besides, prior experience with task and familiarity with technology were found to favorably interfere with performance by decreasing workload. However, prior consultation with BMIM does not seem to influence user performance.

The analysis presented in this study points to potential improvement in BMIM through incorporating AR. Results highlight the influence of individual factors on workload for each type of implementation and point facility and building managers to where efforts should be invested for improvement. Considering the two types of information delivery devices, the tablet assumed a better performance. Further analysis demonstrates that the insertion of AR technology, regardless of deployment method (distributed in the environment or overlaid on paper manual) or delivery device (tablet or smart glasses), improves user experience with BMIM. In short, it is proved that the insertion of AR technology, regardless of the insertion (distributed in the environment or paper manual) or employed device (tablet or smart glasses), acts favorably integrated with the BMIM. Also, very close workload scores resulting from the execution of the task supported by the BMIM-enhanced environment indicate the potential of BMIM to be aligned with Industry 4.0 and the Internet of Things. This assertion is reinforced by Bock [101], who opines that construction automation technologies are rapidly merging with the built environment, becoming part of buildings, components, and furniture.

21.9 Data Availability

All data, models, and code that support the findings of this study are available from the authors upon reasonable request, with the exception of proprietary or confidential data which may only be provided with restrictions (e.g., anonymized data).

Supplementary Video Material *Examples of the applications described in this chapter can be seen in the following video(s):* https://www.youtube.com/watch?v=7rDDuKeEiwk and https://www.youtube.com/watch?v=T90HXxrV5yU

References

1. ABNT: Manutenção de Edificações – Procedimento (Building Maintenance – Procedure). Associação Brasileira de Normas Técnicas (Brazilian Association of Technical Standards), NBR 5674, Rio de Janeiro (1999). In Portuguese
2. Roy, A., Fisk, A.: Usable and Safe Operating Manuals for Consumer Goods: A Guideline. Institute of Scientific and Technical Communications, Peterborough (2004)
3. BSI: Guide to Facilities Maintenance Management. British Standards Institution, BS 8210, London (2012)
4. BSI: Preparation of Instructions for Use – Structuring, Content and Presentation – Part 1: General Principles and Detailed Requirements. British Standards Institution, BS EN 82079-1, London (2012)
5. BSI: Maintenance – Documentation for Maintenance. British Standards Institution, BS EN 13460:2009, London (2010)
6. Behzadan, A.H., Vassigh, S., Mostafavi, A.: Teaching millennials with augmented reality: cases from the U.S. education system. PARC Res. Archit. Build. Constr. **7**(4), 265–272 (2016)
7. Zevi, B.: Saber Ver a Arquitetura, 5th edn. Martins Fontes, São Paulo (1996). In Portuguese
8. Levy, P.: Inteligencia Colectiva, 5th edn. Rio de Janeiro, Brazil (1989). Edited by Loyola, in Portuguese
9. Cardoso, A., Júnior, E.L., Kirner, C., Kelner, J.: Tecnologias para o Desenvolvimento de Sistemas de Realidade Virtual e Aumentad. UFPE Press, Recife (2007). In Portuguese
10. Behzadan, A.H., Kamat, V.R.: Enabling discovery-based learning in construction using telepresent augmented reality. Autom. Constr. **33**, 3–10 (2013)
11. Lopes, R.D.: Realidade aumentada: conceitos, sistemas e aplicações. In: Criação, Representação e Visualização Digitais: Tecnologias Digitais de Criação, Representação e Visualização no Processo de Projeto, pp. 147–164. UNB, Brasília (2012). In Portuguese
12. Rankohi, S., Waugh, L.: Review and analysis of augmented reality literature for construction industry. Vis. Eng. **1**(9), 1–18 (2013)
13. Dib, H.N., Adamo, N.: An augmented reality environment for students' learning of steel connection behavior. In: Proc. International Workshop on Computing in Civil Engineering, ASCE, Seattle, pp. 51–58 (2017)
14. Hou, L., Wang, X., Truijens, M.: Using augmented reality to facilitate piping assembly: an experiment-based evaluation, computing. Civ. Eng. **29**(1), 05014007 (2014)
15. Hou, L., Wang, X., Bernold, L., Love, P.E.D.: Using animated augmented reality to cognitively guide assembly. J. Comput. Civ. Eng. **27**(5), 439–451 (2013)
16. Wang, X., Love, P.E.D., Kim, M.J., Park, C.S., Sing, C.P., Hou, L.: A conceptual framework for integrating building information modeling with augmented reality. Autom. Constr. **34**, 37–44 (2013)
17. Olbrich, M., Graf, H., Kahn, S., Engelke, T., Keil, J., Reiss, P., Webel, S., Bockholt, U., Picinbono, G.: Augmented reality supporting user-centric building information management. Vis. Comput. **29**(10), 1093–1105 (2013)
18. Irizarry, J., Gheisari, M., Williams, G., Roper, K.: Ambient intelligence environments for accessing building information: a healthcare facility management scenario. Facilities. **32**(3/4), 120–138 (2014)
19. Lee, S., Akin, Ö.: Augmented reality-based computational fieldwork support for equipment operations and maintenance. Autom. Constr. **20**(4), 338–352 (2011)
20. ABNT: Diretrizes para Elaboração de Manuais de Uso, Operação e Manutenção das Edificações – Requisitos Para Elaboração e Apresentação dos Conteúdos (Guidelines for The Preparation of Building Use, Operation and Maintenance Manuals – Requirements for the Preparation and Presentation of Contents). Associação Brasileira de Normas Técnicas (Brazilian Association of Technical Standards), NBR 14037, Rio de Janeiro (2014). In Portuguese
21. Lukka, K.: The constructive research approach. In: Ojala, L., Hilmola, O.P. (eds.) Case Study Research in Logistics, Series B1, pp. 83–101. Publications of the Turku School of Economics and Business Administration, Turku (2003)
22. Leite, F., Cho, Y., Behzadan, A.H., Lee, S.H., Choe, S., Fang, Y., Akhavian, R., Hwang, S.: Visualization, information modeling, and simulation: grand challenges in the construction industry. J. Comput. Civ. Eng. **30**(6), 04016035 (2016)
23. Pereira, R.E., Gheisari, M., Esmaeili, B.: Using panoramic augmented reality to develop a virtual safety training environment. In: Proc. Construction Research Congress, ASCE (2018)
24. Heydarian, A., Carneiro, J.P., Gerber, D.J., Becerik-Gerber, B.: Immersive virtual environments versus physical built environments: a benchmarking study for building design and user-built environment explorations. Autom. Constr. **54**, 116–126 (2015)
25. Williams, G., Gheisari, M., Chen, P.J., Irizarry, J.: BIM2MAR: an efficient BIM translation to mobile augmented reality applications, management. Engineering. **31**(1), A4014009 (2014)
26. Nee, A.Y.C., Ong, S.K., Chryssolouris, G., Mourtzis, D.: Augmented reality applications in design and manufacturing. CIRP Ann. Manuf. Technol. **61**(2), 657–679 (2012)

21

27. Rohani, M., Fan, M., Yu, C.: Advanced visualization and simulation techniques for modern construction management. Indoor Built Environ. **23**(5), 665–674 (2014)
28. Zhang, J., Ong, S.K., Nee, A.Y.C.: A multi-regional computation scheme in an AR-assisted in situ CNC simulation environment. Comput. Aided Des. **42**(12), 1167–1177 (2010)
29. Ng, L.X., Wang, Z.B., Ong, S.K., Nee, A.Y.C.: Integrated product design and assembly planning in an augmented reality environment. Assem. Autom. **33**(4), 345–359 (2013)
30. Li, X., Yi, W., Chi, H.L., Wang, X., Chan, A.P.C.: A critical review of virtual and augmented reality (VR/AR) applications in construction safety. Autom. Constr. **86**, 150–162 (2018)
31. Ayer, S.K., Chalhoub, J.: Mixed reality for electrical prefabrication tasks. In: Proc. International Workshop on Computing in Civil Engineering, ASCE, Seattle, pp. 76–83 (2017)
32. Graf, H., Soubra, S., Picinbono, G., Keough, I.: Lifecycle building card: toward paperless and visual lifecycle management tools. In: Proc. Symposium on Simulation for Architecture and Urban Design, San Diego, pp. 5–12 (2011)
33. Irizarry, J., Gheisari, M., Williams, G., Walker, B.N.: InfoSPOT: a mobile augmented reality method for accessing building information through a situation awareness approach. Autom. Constr. **33**, 11–23 (2013)
34. Felizardo, K.R., Martins, R.M.: Engenharia de Software Experimental: Revisão Sistemática, Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo, São Carlos (2009)
35. Kitchenham, B.: Procedures for Performing Systematic Reviews. Technical Report TR/SE-0401. Keele University, Keele (2004)
36. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. **6**(4), 355–385 (1997)
37. Müller, L., Aslan, I., Krüben, L.: GuideMe: a Mobile augmented reality system to display user manuals for home appliances. In: Proc. Advances in Computer Entertainment, Boekelo, pp. 152–167 (2013)
38. Shirazi, A., Behzadan, A.H.: Content delivery using augmented reality to enhance students' performance in a building design and assembly project. Adv. Eng. Educ. **4**(3), 1–24 (2015)
39. Dunston, P.S., Wang, X.: A hierarchical taxonomy of AEC operations for mixed reality applications. Inf. Technol. Constr. **16**, 433–444 (2011)
40. Regenbrecht, H., Baratoff, G., Wilke, W.: Augmented reality projects in the automotive and aerospace industries. IEEE Comput. Graph. Appl. **25**(6), 48–56 (2005)
41. Schall, G., Zollmann, S., Reitmayr, G.: Smart Vidente: advances in mobile augmented reality for interactive visualization of underground infrastructure. Pers. Ubiquit. Comput. **17**(7), 1533–1549 (2013)
42. Takano, K., Hata, N., Kansaku, K.: Towards intelligent environments: an augmented reality–brain–machine interface operated with a see-through head-mount display. Front. Neurosci. **5**, 1–5 (2011)
43. Chen, Y., Chi, H.L., Kang, S.C.J., Hsieh, S.H.: Attention-based user interface design for a tele-operated crane. J. Comput. Civ. Eng. **30**(3), 04015030 (2015)
44. Ceruti, A., Liverani, A., Bombardi, T.: Augmented vision and interactive monitoring in 3D printing process. Int. J. Interact. Des. Manuf. **11**(2), 385–395 (2016)
45. Lima, J.P.: Markerless tracking system for augmented reality in the automotive industry. Expert Syst. Appl. **82**, 100–114 (2017)
46. Marín, R., Sanz, P.J., Del Pobil, A.P.: The UJI online robot: an education and training experience. Auton. Robot. **15**(3), 283–297 (2003)
47. Arisandi, R., Otsuki, M., Kimura, A., Shibata, F., Tamura, H.: Virtual handcrafting: building virtual wood models using ToolDevice. Proc. IEEE. **102**(2), 185–195 (2014)
48. Zaher, M., Greenwood, D., Marzouk, M.: Mobile augmented reality applications for construction projects. Constr. Innov. **18**(2), 152–166 (2018)
49. Ahn, S., Han, S., Al-Hussein, M.: 2D drawing visualization framework for applying projection-based augmented reality in a panelized construction manufacturing facility: proof of concept. J. Comput. Civ. Eng. **33**(5), 040190321–0401903215 (2019)
50. Hara, H., Komatsu, R., Shiota, N.: Architecture for digital technology utilization to accelerate digital innovation. Fujitsu Sci. Tech. J. **54**(3), 39–46 (2018)
51. Patti, E.: Information modeling for virtual and augmented reality. IT Prof. **19**(3), 52–60 (2017)
52. Ferreira, J., Vale, A., Ribeiro, I.: Localization of cask and plug remote handling system in ITER using multiple video cameras. Fusion Eng. Des. **88**(9–10), 1992–1996 (2013)
53. Hou, L.: A framework of innovative learning for skill development in complex operational tasks. Autom. Constr. **83**, 29–40 (2017)
54. Crescenzio, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., Salti, S.: Augmented reality for aircraft maintenance training and operations support. IEEE Comput. Graph. Appl. **31**(1), 96–101 (2011)
55. Aleotti, J., Micconi, G., Caselli, S., Benassi, G., Zambelli, N., Bettelli, M., Zappettini, A.: Detection of nuclear sources by UAV teleoperation using a visuo-haptic augmented reality interface. Sensors. **17**(10), 1–22 (2017)
56. Pejoska-Laajola, J., Reponen, S., Virnes, M., Leinonen, T.: Mobile augmented communication for remote collaboration in a physical work context. Australas. J. Educ. Technol. **33**(6), 11–26 (2017)
57. Bademosi, F., Blinn, N., Issa, R.A.: Use of augmented reality technology to enhance comprehension of steel structure construction. In: Proc. Joint Conference on Computing in Construction, Edinburgh, pp. 58–79 (2017)
58. Yang, Z., Shi, J., Jiang, W., Sui, Y., Wu, Y., Ma, S., Kang, C., Li, H.: Influences of augmented reality assistance on performance and cognitive loads in different stages of assembly task. Front. Psychol. **10**, 1–17 (2019)
59. Chalhoub, J., Ayer, S.K.: Using mixed reality for electrical construction design communication. Autom. Constr. **86**, 1–10 (2018)
60. Blanco-Novoa, Ó., Fernández-Caramés, T.M., Fraga-Lamas, P., Vilar-Montesinos, M.A.: A practical evaluation of commercial industrial augmented reality systems in an Industry 4.0 shipyard. IEEE Access. **6**, 8201–8218 (2018)
61. Doshi, A., Smith, R.T., Thomas, B.H., Bouras, C.: Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. Int. J. Adv. Manuf. Technol. **89**(5–8), 1279–1293 (2016)
62. Bruns, W.F.: Complex construction kits for coupled real and virtual engineering workspaces. In: Proc. International Workshop on Cooperative Buildings, Pittsburgh, pp. 55–68 (1999)
63. Li, S.Q., Peng, T., Wang, J.F.: Mixed reality-based interactive technology for aircraft cabin assembly. Chin. J. Mech. Eng. **22**(3), 403–409 (2009)
64. Diao, P.H., Shih, N.J.: BIM-Based AR Maintenance System (BARMS) as an intelligent instruction platform for complex plumbing facilities. Appl. Sci. **9**(8), 1–12 (2019)
65. Bosché, F., Abdel-Wahab, M., Carozza, L.: Towards a mixed reality system for construction trade training. J. Comput. Civ. Eng. **30**(2), 040150161–0401501612 (2016)
66. Selonen, P.: Mixed reality web service platform. Multimedia Systems. **18**(3), 215–230 (2012)
67. Huang, J.M., Ong, S.K., Nee, A.Y.C.: Visualization and interaction of finite element analysis in augmented reality. Comput. Aided Des. **84**, 1–14 (2017)
68. Chen, H., Huang, P.: 3D AR-based modeling for discrete-event simulation of transport operations in construction. Autom. Constr. **33**, 123–136 (2013)

69. Tavares, P., Costa, C.M., Rocha, L., Malaca, P., Costa, P., Moreira, A.P., Sousa, A., Veiga, G.: Collaborative welding system using BIM for robotic reprogramming and spatial augmented reality. Autom. Constr. **106**, 1–12 (2019)
70. Kim, D., Hwang, D.: Non-marker based mobile augmented reality and its applications using object recognition. J. Univ. Comput. Sci. **18**(20), 2832–2850 (2012)
71. Alsafouri, S., Ayer, S.K.: Mobile augmented reality to influence design and constructability review sessions. Archit. Eng. **25**(3), 040190161–0401901611 (2019)
72. Ciuffini, A.F., Di Cecca, C., Ferrise, F., Mapelli, C., Barella, S.: Application of virtual/augmented reality in steelmaking plants layout planning and logistics. Metall. Ital. **2016**(7), 5–10 (2016)
73. Park, H., Moon, H., Lee, Y.J.: Tangible augmented prototyping of digital handheld products. Comput. Ind. **60**(2), 114–125 (2009)
74. Wu, W., Hartless, J., Tesei, A., Gunji, V.: Design assessment in virtual and mixed reality environments: comparison of novices and experts. J. Constr. Eng. Manag. **145**(9), 040190491–0401904914 (2019)
75. Park, C.S., Le, Q.T., Pedro, A., Lim, C.R.: Interactive building anatomy modeling for experiential building construction education. J. Prof. Issues Eng. Educ. Pract. **142**(3), 040150191–0401501912 (2016)
76. Fujinami, N., Inagawa, K., Nishijo, A.: Sokan: a middleware for a tabletop procedure-aware information display. Multimed. Tools Appl. **57**(2), 269–293 (2012)
77. Bifulco, P., Narducci, F., Vertucci, R., Ambruosi, P., Cesarelli, M., Romano, M.: Telemedicine supported by augmented reality: an interactive guide for untrained people in performing an ECG test. Biomed. Eng. Online. **13**(1), 1–16 (2014)
78. Khalek, I.A., Chalhoub, J.M., Ayer, S.K.: Augmented reality for identifying maintainability concerns during design. Adv. Civ. Eng. **2019**, 1–12 (2019)
79. García, E.S., García-Valldecabres, J., Blasco, M.J.V.: The use of HBIM models as a tool for dissemination and public use management of historical architecture: a review. Int. J. Sustain. Dev. Plan. **13**(1), 96–107 (2018)
80. Atherinis, D., Bakowski, B., Velcek, M., Moon, S.: Developing and laboratory testing a smart system for automated falsework inspection in construction. J. Constr. Eng. Manag. **144**(3), 040171191–0401711910 (2018)
81. Ge, L., Kuester, F.: Integrative simulation environment for conceptual structural analysis, computing. Civ. Eng. **29**(4), B4014004 (2015)
82. Chu, M., Matthews, J., Love, P.E.D.: Integrating mobile building information modelling and augmented reality systems: an experimental study. Autom. Constr. **85**, 305–316 (2018)
83. Liu, F., Seipel, S.: On the precision of third person perspective augmented reality for target designation tasks. Multimed. Tools Appl. **76**(14), 15279–15296 (2016)
84. Hernández, J., Lerones, P., Bonsma, P., Delft, A., Deighton, R., Braun, J.D.: Jan-Derrick: an IFC interoperability framework for self-inspection process in buildings. Buildings. **8**(2), 1–9 (2018)
85. Chen, K.: A BIM-based location aware AR collaborative framework for facility maintenance management. Electron. Inf. Technol. Constr. **24**, 360–380 (2019)
86. Reitmayr, G., Schmalstieg, D.: A platform for location based augmented reality applications. ÖGAI J. **21**(1), 1–5 (2002)
87. Fraga-Lamas, P., Fernández-Caramés, T.M., Blanco-Novoa, Ó., Vilar-Montesinos, M.A.: A review on industrial augmented reality systems for the Industry 4.0 shipyard. IEEE Access. **6**, 13358–13375 (2018)
88. Liu, F., Seipel, S.: Precision study on augmented reality-based visual guidance for facility management tasks. Autom. Constr. **90**, 79–90 (2018)
89. Neges, M., Koch, C., König, M., Abramovici, M.: Combining visual natural markers and IMU for improved AR based indoor navigation. Adv. Eng. Inform. **31**, 18–31 (2017)
90. Carreira, P., Castelo, T., Gomes, C.C., Ferreira, A.: Virtual reality as integration environments for facilities management. Eng. Constr. Archit. Manag. **25**(1), 90–112 (2018)
91. Chen, Y.C., Chi, H.L., Kang, S.C.J., Hsieh, S.H.: Attention-based user interface design for a tele-operated crane. J. Comput. Civ. Eng. **30**(3), 040150301–0401503012 (2016)
92. Pedro, A., Le, Q.T., Park, C.S.: Framework for integrating safety into construction methods education through interactive virtual reality. J. Prof. Issues Eng. Educ. Pract. **142**(2), 040150111 (2016)
93. Dresch, A., Lacerda, D.P., Antunes Júnior, J.A.V.: Design science research: método de pesquisa para avanço da ciência e tecnologia (Bookman, 2015). In Portuguese
94. Moreira, L.C., Tonoli, J.G., Ruschel, R.C.: A prática do Manual do Proprietário da Edificação: Uma Classificação Conforme a NBR 14037. Gest. Tecnol. Proj. **13**(3), 119–134 (2018). In Portuguese
95. Dunser, A., Billinghurst, M.: Evaluating augmented reality systems. In: Furht, B. (ed.) Handbook of Augmented Reality, pp. 289–307. Springer, London (2011)
96. Hart, S.G.: NASA task load index (TLX) (1986)
97. Cole, R., Purao, S., Rossi, M., Sein, M.K.: Being proactive: where action research meets design research. In: Proc. International Conference on Information Systems, Las Vegas (2005)
98. Moreira, L.C.: O Manual do Proprietário da Edificação Assistido Pela Realidade Aumentada. Ph.D. thesis, Universidade Estadual de Campinas, Campinas (2018). In Portuguese
99. Grupo Astra.: https://3dwarehouse.sketchup.com/by/grupoastra?tab=models
100. ABNT: Manutenção de Edificações – Requisitos para o Sistema de Gestão da Manutenção (Building Maintenance – Requirements for the Maintenance Management System). Associação Brasileira de Normas Técnicas (Brazilian Association of Technical Standards), NBR 5674, Rio de Janeiro (2012). In Portuguese
101. Bock, T.: The future of construction automation: technological disruption and the upcoming ubiquity of robotics. Autom. Constr. **59**, 113–121 (2015)

Lorena C. S. Moreira is a professor at the Federal University of Bahia (UFBA, Brazil). Ph.D. in Architecture, Technology, and City from the University of Campinas, and she has been a visiting researcher at Missouri State University and Karlsruhe Institute of Technology. Her area of research includes architecture with an emphasis on Information Technology and Communication (ICT), mixed reality, building information modeling (BIM), and facility management.

Regina C. Ruschel, Ph.D. in Electrical Engineering and Automation from the University of Campinas, Master of Science in Civil Engineering from the University of Arkansas, and Civil Engineer from the Municipal Foundation of Education of Piracicaba. Researcher and contributing professor at the Graduate Program of Architecture, Technology and Cities of the School of Civil Engineering, Architecture and Urban Design of the University of Campinas.

Amir H. Behzadan, Ph.D. in Civil Engineering (2008) and M.Eng. in Construction Engineering and Management (2005) both from the University of Michigan, Ann Arbor. Associate Professor at Texas A&M University, Department of Construction Science. His area of research includes artificial intelligence (AI), data-driven simulation, and sensing and visualization with applications in urban informatics, smart health and ergonomics, disaster resilience, and safety and productivity.

An Augmented Reality Platform for Interactive Finite Element Analysis

22

J. M. Huang, Soh Khim Ong, and Andrew Yeh Ching Nee

Contents

22.1 Introduction 533
22.1.1 Brief Overview of FEA 534
22.1.2 Augmented Reality 534
22.1.3 Research Motivations and Objectives 535

22.2 Research Background 535
22.2.1 Interactive FEA in VR 535
22.2.2 Numerical Simulation and Scientific Visualization in AR 536
22.2.3 Real-Time Finite Element Modeling 537
22.2.4 Discussion 538
22.2.5 Summary 538

22.3 FEA-AR Integrated System 538
22.3.1 System Design Considerations 538
22.3.2 System Architecture 539
22.3.3 AR Environment Setup 540
22.3.4 Summary 543

22.4 Visualization and Exploration of FEA Results 544
22.4.1 Scientific Visualization with VTK 544
22.4.2 Scientific Visualization in AR 545
22.4.3 Data Manipulation and Exploration 546
22.4.4 Summary 548

22.5 Real-Time FEA in AR 548
22.5.1 Real-Time FEA Solution 548
22.5.2 Computation of Inverse Stiffness Matrix 548
22.5.3 Load Acquisition 549
22.5.4 System Workflow and Time Synchronization 552
22.5.5 Summary 553

22.6 Interactive Model Modification 553
22.6.1 Adding Geometric Models 554
22.6.2 Local Mesh Refinement 554
22.6.3 Summary 555

22.7 System Implementation and Case Studies 555
22.7.1 Case Study of a Step Ladder 555
22.7.2 Case Study of a Proving Ring 563
22.7.3 A Prototype Application for Education 563
22.7.4 Summary 565

22.8 Conclusions and Future Readings 566
22.8.1 Some Open Issues 569

References 572

Abstract

Finite element analysis (FEA) is usually carried out off-site and using computer desktops, i.e., computer-generated graphics, which does not promote a user's perception and interaction and limits its applications. This chapter first gives an overview of related FEA and AR technologies and presents the feasibility of enhancing finite element structural analysis with AR technology. A novel system has been proposed which integrates sensor measurement, FEA simulation, and scientific visualization into an AR-based environment. This system can acquire input data using sensors and visualize FEA results directly on real-world objects. Several intuitive interaction methods have been developed for enhancing structural investigation and data exploration. A prototype system has been built and tested using several case studies to validate the proposed methods and evaluate the system performance.

Keywords

Finite element analysis (FEA) · Augmented reality (AR) · AR-based FEA · Structural health monitoring · Sensor network · Simulation · Visualization · Structural mechanics

22.1 Introduction

Numerical simulation is widely used in engineering practices to analyze and predict the behavior of real-world physical systems. As a versatile tool, finite element analysis (FEA) has been applied in many engineering fields, for example, structural mechanics, construction, thermal and fluid

J. M. Huang (✉) · S. K. Ong · A. Y. C. Nee
Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore
e-mail: mpeongsk@nus.edu.sg; mpeneeyc@nus.edu.sg

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_22

engineering, etc. However, conventional FEA is usually carried out in offline and virtual environments on computer desktops, i.e., computer-generated graphics, which does not promote a user's perception and interaction. Moreover, most standard FEA software has counterintuitive user interfaces. Such interfaces are not efficient and often discourage the novice users without technical background from implementing FEA. Augmented reality (AR) technology has seen rapid development in the last two decades with a myriad of applications. With the use of AR, FEA simulation systems can acquire input data directly from physical environments. Visualizing FEA results in the real scene facilitates result interpretation. The ease of input and manipulation in AR environments triggers innovative interactions for data exploration and various analysis tasks.

22.1.1 Brief Overview of FEA

Boundary value problems arise in different engineering fields. In FEA, these problems are solved with the finite element method (FEM). By discretizing the problem domain into a number of disjoint subdomains, i.e., elements, the method generates numerical approximations to the exact solutions using a piecewise polynomial interpolation scheme. The most typical application of FEA is structural analysis, such as determining the deformation and stresses of a structure, determining the natural frequencies and modes of vibrations, etc.

In practice, an FEA usually consists of three stages, i.e., pre-processing, solution, and post-processing. In the pre-processing stage, a geometric model is built for the structure to be analyzed and discretized to generate the mesh model. Various types of elements are available for modeling, such as truss elements, plate elements, tetrahedral elements, etc. Elements are selected according to the behavior of the structure, characteristics of the elements, etc. After the mesh model is built, material properties and boundary conditions are assigned. The boundary conditions may include a combination of loads and displacement constraints. The engineers need to determine the boundary conditions properly to obtain reliable results.

Equations are formed and solved in the solution stage. The stiffness matrices that describe the behavior of each element are computed and assembled to generate the system of equations. There are two main types of solution methods, namely, direct and iterative methods. Direct methods are usually based on Gaussian elimination. The equations are solved with a finite number of algebraic operations to produce exact solutions. Iterative methods are usually based on the conjugate gradient method. In an iterative process, a sequence of approximate solutions is generated, which converges to a solution of acceptable accuracy. Depending on the type of analysis, some additional solution techniques may be required. For examples, time integration schemes are needed for dynamic problems. Nonlinear systems are solved using iteration schemes.

In the post-processing stage, the results generated in the solution stage are sorted and displayed to the user. Graphical representations of the results are created to help with the understanding and interpretation of the data. Many display options are available, and the choice depends on the form of the data. For instance, displacement vectors are displayed as deformed structural shape. Scalar stresses are displayed on the model as colored contours. Dynamic animations are generated for time-dependent results.

Commercial FEA software systems such as ANSYS, Abaqus, COMSOL, etc. have greatly facilitated FEA studies. These standard FEA packages are generally composed of three separate modules corresponding to the three analysis stages, namely, pre-processor, solver, and post-processor. The data generated from one module is used in the next module. These general-purpose FEA packages provide a wide range of element and material options and are capable of handling different types of analysis, including linear and nonlinear analysis, static and dynamic analysis, etc.

22.1.2 Augmented Reality

AR is a technology that supplements the real world with computer-generated information such as graphics, video, and text registered onto the real scene [1]. It is often discussed together with virtual reality (VR) technology [2]. VR, also called virtual environment, immerses users in a completely synthetic world without experiencing the real world, while AR superimposes the real world with virtual objects. An AR system is expected to have the following three characteristics [3]: (1) real and virtual objects are combined in a real environment, (2) it allows user interaction in real time, and (3) real and virtual objects are registered with each other.

AR enhances a user's perception of and interaction with the surrounding real-world environment. Virtual and real objects complement each other in AR. The digital data, such as texts, graphics, and videos, are superimposed onto the real world to display the information that the user cannot detect directly with human senses. The real world, in the meanwhile, provides physical contexts for a better understanding of the digital information. Furthermore, AR bridges the real and virtual worlds, enabling the user to interact with digital information using natural interfaces. The physical objects are useful to serve as interface elements in an AR environment. By linking their physical properties with digital data, the real objects can be manipulated directly to control the generation and display of the digital data. The AR interfaces can provide the user with an intuitive experience of digital information. AR has already been demonstrated to be a novel human-computer interaction (HCI) tool. The applications of AR include navigation, manufacturing, entertainment, education, etc.

22.1.3 Research Motivations and Objectives

FEA processes are limited by the traditional computer setups as the entire process is conducted in a totally virtual and offline environment. The virtual environment does not offer the human senses of the physical entities, such as texture and material, as well as the surrounding physical context. Moreover, FE models are often simplified by discarding non-essential geometric features or by idealizing 3D geometries with low-dimensional elements, e.g., 2D plate elements. The simplification saves computation power but results in differences between the FE models and the corresponding real objects.

Another problem lies in the interactivity. Most FEA software has counterintuitive user interfaces which are designed for expert users. Such interfaces require laborious data entry and command browsing, such that users are often required to spend a considerable amount of time learning the software. This undoubtedly discourages the implementation of FEA. Besides, traditional FEA usually consists of three steps, i.e., pre-processing, solution, and post-processing. After the user builds the model and specifies the relevant parameters, the FEA system solves the equations and displays the results. The number of nodes of the model and the processor performance determine the computation time. Traditional FEA systems have low interactivity as users may need to repeat steps due to parameter changes. In many situations, FEA systems should be able to update results quickly in response to variation of parameters, e.g., changes to a structure due to changes in loading conditions, and modification of product design. Such demands necessitate the development of interactive FEA technologies.

Motivated by the drawbacks of the operating environment and interactivity of traditional FEA systems, this research proposes to build an intuitive and interactive FEA environment using AR technology. The objectives of this research are listed as follows:

- An AR-based system for FEA will be constructed. The system can acquire simulation parameters directly from the physical world, so as to reduce the user's efforts to prepare, input, and edit the data. The FEA results will be generated and superimposed on the corresponding real objects, facilitating understanding and interpretation of the data.
- The system will provide easy access to FEA results. The use can explore the data intuitively in real-world environments. Besides observing the results on the model surfaces, the user can explore the interior results and extract the data in the regions of interest easily.
- An interactive simulation environment will be achieved. The user will be able to interact with the FE model intuitively. They can make changes of the variables, such as loads, geometric model, mesh resolution, etc. The FEA results will be updated with a short lag time or near real time.

Specifically, the research issues to achieve these objectives are as follows:

- Developing an integrated system for performing FEA tasks in an AR environment. Enabling automatic data acquisition by using sensors to acquire input data directly from the real world and converting the data to relevant simulation or visualization parameters for FEA
- Visualizing and superimposing FEA results on the corresponding real objects. Developing interaction methods for enhanced exploration of the volumetric result data
- Enabling intuitive and efficient interaction methods for performing FEA simulations in the AR environment. Building AR interfaces for specific FEA tasks. Investigating the methodology and applications of real-time FEA in AR

22.2 Research Background

This section gives a review of the studies related to this research topic.

22.2.1 Interactive FEA in VR

To overcome the shortcomings of conventional FEA environments, i.e., desktop-based environments, VR technology has been employed by many researchers to achieve immersive environments. The ease of navigation and manipulation in the virtual environments facilitates exploration of FEA results. Various visualization and interaction approaches have been developed. Stereoscope visualization was used by Scherer and Wabner [4] to study FEA results. Glyph-based representation of multidimensional stress data is created to enhance understanding and interpretation of complex datasets. Hafner et al. [5] used a method to post-process electromagnetic solutions for representation in VR. With Visualization Toolkit (VTK) [6], interactive operation of 3D solution data has been achieved. The user can operate a virtual plane to slice through the solution data and operate the data objects, such as grabbing a data object to access the information. Lee and El-Tawil [7] adopted the Virtual Reality Markup Language (VRML) in their FEMvrml system to visualize the time-dependent results of structural analysis. Java-based interactions were used to control the animation prepared a priori. In these studies, VR technology is used solely for post-processing FEA solutions.

Ingrassia and Cappello [8] developed VirDe to facilitate the designers in performing design, CAD modeling, and FEA simulation in a VR environment. To facilitate the understanding of structural behavior of buildings, Setareh et al. [9] developed an application for the users to construct structures and simulate the effects of environmental loading conditions. On the other hand, interactive FEA applications are always more immersive. Such systems can update FEA results in response to the parameter variations. VR-based FEA is useful for structural investigation, as interactive simulation allows users to concentrate on structural design rather than paying much attention to operate the simulation tools. An efficient system is expected to generate results of acceptable accuracy in near real time.

Liverani et al. [10] used a VR system for FEA for simulating shell structures. Load changing results can be obtained in near real time on a scaled down model. For more complex models, real-time solutions cannot be achieved using standard solvers. Several works have studied artificial neural networks (ANN) in achieving near real-time interaction for specific applications. In sports engineering, Hambli et al. [11] used an ANN to simulate real-time deformation of a tennis ball hitting a racket. The ANN was trained with input/output pairs that were calculated using FEA. The fast computation of ANN allows the users to play a VR tennis game, and the impact of the ball can be felt with a haptic glove. Connell and Tullberg [12] presented a two-stage framework involving an approximation module and a precise module for the integration of FEA with VR. Cheng and Tu [13] used ANN for simulating real-time deformation of mechanical components subject to external forces. Component geometry can be altered by the user to study the resulting effects.

The common practice is to alter a FE model using a CAD system. Direct mesh manipulation without going through a CAD modeler will be ideal for real-time interactive simulation. Yeh and Vance [14] developed an interactive stress analysis method for the geometric design of components. The designers are able to alter the FEA mesh directly, and the resulting stresses can be updated using linear Taylor series approximations. Ryken and Vance [15] used this approach to analyze the stresses of a tractor lift arm in a CAVE-like VR system. Rose et al. [16] developed a mesh editing method by the direct manipulation of the nodes. Graf and Stork [17] were able to obtain real-time results for moving loads, as they used the pre-computed inverse stiffness matrix, and the results are calculated only for elements which are visible. With a mesh modification mechanism, the system allows the user to drag geometric features directly to change the positions. In addition, the user can create cross sections to access the interior FEA results.

22.2.2 Numerical Simulation and Scientific Visualization in AR

Scientific visualization techniques are usually employed to illustrate complex data graphically, such as scalar, vector, and tensor fields, helping users to understand and gain insights from the data. Compared with VR, AR operates in real-world environments. Scientific visualization of simulated or measured data in the real environment is highly intuitive for various applications. Mixed and augmented reality has found good applications in the medical field [18] and in neurosurgery [19]. However, the integration of VR and AR in the medical field could have legal implications and would need to be exercised carefully with full compliance of healthcare requirements [20]. There are ample examples of engineering applications, such as manufacturing [21, 22] and in architectural and civil engineering [23]. A survey of industrial AR applications can be found in [24] and in human-robot collaboration [25]. AR in education and training has taken a major step as many of the physical phenomenon can be demonstrated using AR simulation [26, 27, 28].

Mobile AR platforms have good applications outdoor. However, it is not possible for mobile devices to have sufficient computational power; hence, powerful servers are required. Weidlich et al. [29] created a mobile AR system using a client-server architecture with bidirectional communication capabilities. FEA computation is performed on the server, and the output is sent to the client which renders and provides the interactive functions. Heuveline et al. [30] presented a mobile system for the simulation of airflow in an urban environment. A CFD simulation based on an airflow model was carried out using a server. Visualization of the flow field was displayed on a smartphone using a hybrid tracking technique. Interactive functions can be supported by utilizing the sensors and the touchscreen of the mobile phone.

The elements of sensing and measurement play pivotal roles in AR technologies and applications [31]. Underkoffler et al. [32] presented the "Luminous Room" which is an interior architectural space where its surfaces can capture and display visual information. An interactive simulation allows physical objects placed in the space to behave like obstacles in a computational flow field which can be updated. Niebling et al. [33] developed an AR application of a water turbine. A user can adjust the rotating angle of the real turbine blades, and the changes will be input to the CFD software, and updated results will be superimposed for visualization on the turbine. An AR sandbox [34] was developed at UC Davis for teaching earth science. In this system, users can create different topographies by shaping real sands, and a simulated water flow will be displayed on the sandbox.

Time delay is inevitable in large-scale numerical simulation. Real-time simulation will greatly promote the interactivity and offer attractive features as users can receive precise time and space feedbacks. Kaufmann and Meyer [35] developed PhysicsPlayground for teaching of mechanics. Using this system, students can create and manipulate virtual objects interactively for performing physics experiments in 3D space. Valentini and Pezzuti [36, 37] used AR methodology to create interactive kinematic and dynamic simulation of rigid and deformable bodies. Mannus et al. [38] developed a magnetic field simulation system for teaching magnetism. The user can explore the effect of magnetic fields using real magnets with visual and haptic feedbacks from sensors.

In an AR environment, real-time numerical simulators can acquire boundary conditions directly from the real world using sensors. Malkawi and Srinivasan [39] developed an interactive AR environment of a room using sensors to monitor air temperature and air flow changes to generate updated boundary conditions required in CFD simulation. In the medical application, Haouchine et al. [40] developed a visualization method to observe human tissue deformation during minimally invasive liver surgery. A real-time model is constructed based on the co-rotational FE method. The co-rotational method decomposes the element configuration into rotational and deformational components. This method can handle large deformation while using a linear stress-strain relationship. This study estimated the deformation of soft objects, and it has achieved visually plausible results. The method, however, may not be feasible for engineering analysis. It is not always necessary to solve equations online for real-time FEA applications. In some situations, FEA can be performed offline to investigate the relationship between the simulation objectives and the relevant variables, so as to simplify online computation. Bernasconi et al. [41] studied crack growth monitoring in single-lap joints. With data from strain gauges, the crack position can be computed and displayed on the joints in real time using AR.

Real-time FEA simulation with AR interfaces has been used in several educational applications. Fiorentino et al. [42] developed an interactive simulation system for learning structural mechanics. A camera was used to detect the displacement constraints altered by the user for updating FEA results. This was implemented on a cantilever which was deformed manually. Matsutomo et al. [43, 44] developed two real-time AR visualization systems for the teaching of magnetics using simulation accelerated with a pre-computation and a 2D mesh deformation method.

22.2.3 Real-Time Finite Element Modeling

Real-time finite element (FE) modeling is necessary in various applications [45]. In surgical simulation, the ability to compute the deformation of human tissues in real time is a promising technology for medical training. VR technology is usually employed to create an immersive environment with realistic visual and haptic feedbacks. To allow real-time interactions, the FE models require update rates for visual and haptic feedbacks higher than 24 Hz and 300 Hz, respectively [46], while exhibiting correct physical behavior corresponding to the behavior of real human tissues. Standard FEA solvers, such as conjugate gradient method, cannot satisfy the speed needed. Researchers have managed to reduce the solution time with pre-computation or model simplification.

Linear elastic models with pre-computation techniques have been adopted in many studies. Bro-Nielsen and Cotin [47], in their pioneering work, developed tetrahedral elements of linear elasticity for monitoring real-time deformation. Cotin et al. [46] developed an approach to evaluate real-time deformation due to displacement constraints instead of forces. Nikitin et al. [48] developed large elastic object simulation in an interactive VR environment using an offline inversion of the stiffness matrix. In another medical application, Berkley et al. [49] developed a real-time simulation of suturing using FEM. These studies pre-compute the models offline, such that the online solution time is reduced. However, when the mesh models are modified, e.g., tissue cutting, the pre-computed data needs to be updated accordingly. Lee et al. [50] proposed a topology modification method for tissue cutting, which updates the inverse stiffness matrix using the Sherman-Morrison-Woodbury formula, rather than re-computing the inverse matrix completely.

Linear elastic models are used in the abovementioned studies. Loading is assumed to be quasi-static, and dynamic effects are neglected to reduce computational power and time requirements. Dynamic analysis can be performed using explicit or implicit time integration schemes [51]. Explicit integration schemes are adopted in many studies for developing real-time nonlinear models [52, 53, 54]. The main advantage of explicit schemes is the involvement of only the mass and damping matrices which are diagonal if mass lumping is used . However, explicit schemes are conditionally stable, i.e., stable only if the time step is smaller than a threshold. Hence, they are not feasible for the real-time simulation of stiff objects which usually demand extremely small time steps. In contrast, implicit schemes can be unconditionally stable, e.g., the backward Euler method, which implies that large time steps may be used. However, solving an algebraic system of equations at each time step is necessary for the implicit scheme. Geometry nonlinearity can be handled using some real-time deformation models by extracting local material rotations from co-rotational models [55], and implementing parallel computation can dramatically accelerate FEA solvers in certain situations [56]. A few open-source libraries are available for implementing real-time deformation models, such as SOFA [57] and Vega [58]. In addition, a study for prosthetic fitting shows the possibility of combining load measurement and real-time FE models to support the evaluation of internal stresses in tissues [59].

These real-time FE models can achieve visually plausible deformation for soft objects, but most models are not suitable for engineering analysis due to the computational cost [55]. However, the real-time methods for linear models are practical for structural analysis.

22.2.4 Discussion

Literature review reveals the advantages of AR-based environment over VR-based environment and indicates the benefits of applying AR technology to FEA, i.e., visualization of simulation results in the real context, automatic data acquisition from the physical world, and intuitive interfaces for performing simulation and data exploration. However, most researchers used AR technology to facilitate the visualization and navigation of FEA results, but did not consider integrating FEA processes into AR environments. The potential interaction methods for performing FEA and exploring the result data have not been investigated thoroughly. Some AR-based methods allow the user to control simulation variables, and cameras are normally used to capture user inputs. For real-world objects and environments, many parameters cannot be controlled by the users, but have to depend on events which occur during the operation. To record these parameters, appropriate sensors are required to perform the measurements. In addition, simplified models can be used to achieve faster simulation in some real-time applications. These methods have limited capabilities for engineering analysis. This research will investigate the methods for interactive FEA in AR environments. Real-time visualization of FEA results will be achieved with loads measured using sensors. Besides, user interaction methods will be developed to enhance FEA simulation and result exploration.

22.2.5 Summary

The section gives a review of the studies related to the application of VR and AR technologies to numerical simulation. The advantages of AR-based environment over VR-based environment for engineering analysis are discussed, and the research issues are identified. Real-time performance is important for interactive FEA in an AR environment. The reported studies on real-time FE modeling techniques are examined. The development of real-time FE methods provides technical support for achieving fast FEA solutions.

22.3 FEA-AR Integrated System

In this section, an integrated system is proposed to enhance FEA of structures with AR. The architecture of this FEA-AR integrated system and the environment setup are presented.

22.3.1 System Design Considerations

Robust system construction relies on the use of appropriate sensing and tracking technologies to facilitate visualization and interaction of FEA processes. As shown in Fig. 22.1, a typical FEA simulation starts with a mesh model. Next, material properties and boundary conditions will be assigned to the model, and a system of equations is formed and solved. The solutions are transferred to the visualization module subsequently. In the visualization module, FEA results are processed using filters of specific functions, so as to derive the data to be visualized according to the user's requirement. The output data of the filters is mapped to graphic primitives and rendered to displays with viewing transformations finally.

In an AR environment, graphically visualized FEA results can be registered onto a real scene by tracking the user's viewpoint. Depending on the purposes, AR interfaces can be established to allow the control of simulation and visualization parameters. For instance, the user can modify the mesh model or boundary conditions to investigate the effects on the FEA results and control the parameters of data filters to extract a portion of the result data for observation. Sensors can be employed to capture human body motions to support natural user input and measure varying boundary conditions, e.g., displacement constraints and loads. However, it is usually difficult to measure the small displacement con-

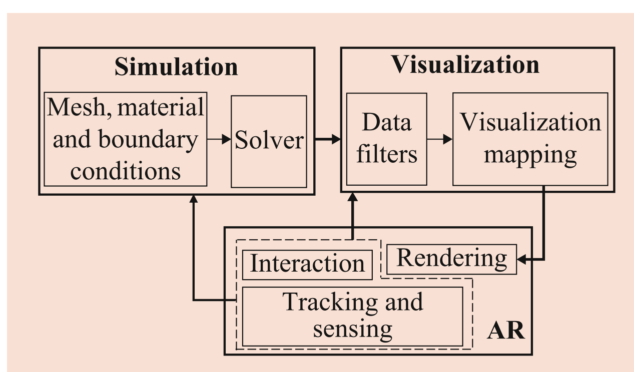


Fig. 22.1 Schematics for applying AR to FEA

straints that are applied to engineering structures built with stiff materials. Measuring loads is more practical. Engineers usually investigate the behavior of structures under different loading conditions. Real-time FEM has been developed to simulate structures under varying loads, but the method has limitations.

22.3.2 System Architecture

With the considerations mentioned in section above, a framework is proposed as shown in Fig. 22.2. There are seven modules in the system, namely, the sensor, load acquisition, FE model, solution, post-processing, rendering, and user interaction.

The sensor module comprises of different sensors such as force, position, and orientation, in the form of force sensors and vision-based trackers. Vision-based trackers are used in this research due to their stability and accuracy. Wireless sensor networks (WSN) are useful for monitoring spatially distributed loads. The user can configure a WSN online through the bidirectional communication between the WSN and load acquisition module. The load acquisition module manages the information of the sensors, such as sensor parameters and communication addresses, and the load data collected from sensors or user input. With the nodal forces and FE model, the solution module is established and used to solve the equations to generate displacements and stresses. Besides having a standard preconditioned conjugate gradient (PCG) solver, the solution module uses a real-time solver to speed up the simulation for varying loading conditions. The real-time solver has been implemented in two phases, i.e., offline pre-computation and online solution. In the offline pre-computation phase, user specifies material properties and boundary conditions after the creation of the mesh models. Next, relevant matrices are computed and variables are initialized. In the online solution phase, nodal loads are obtained from the load acquisition module, and deformation and stresses are next computed. However, should there be any modifications made to the model, updating of the relevant matrices is required.

The post-processing module is developed using scientific visualization techniques, like color coding and data slicing. In this module, simulation results from the solution module are processed using data filters, which allow results to be visualized according to the user's preference. By using the

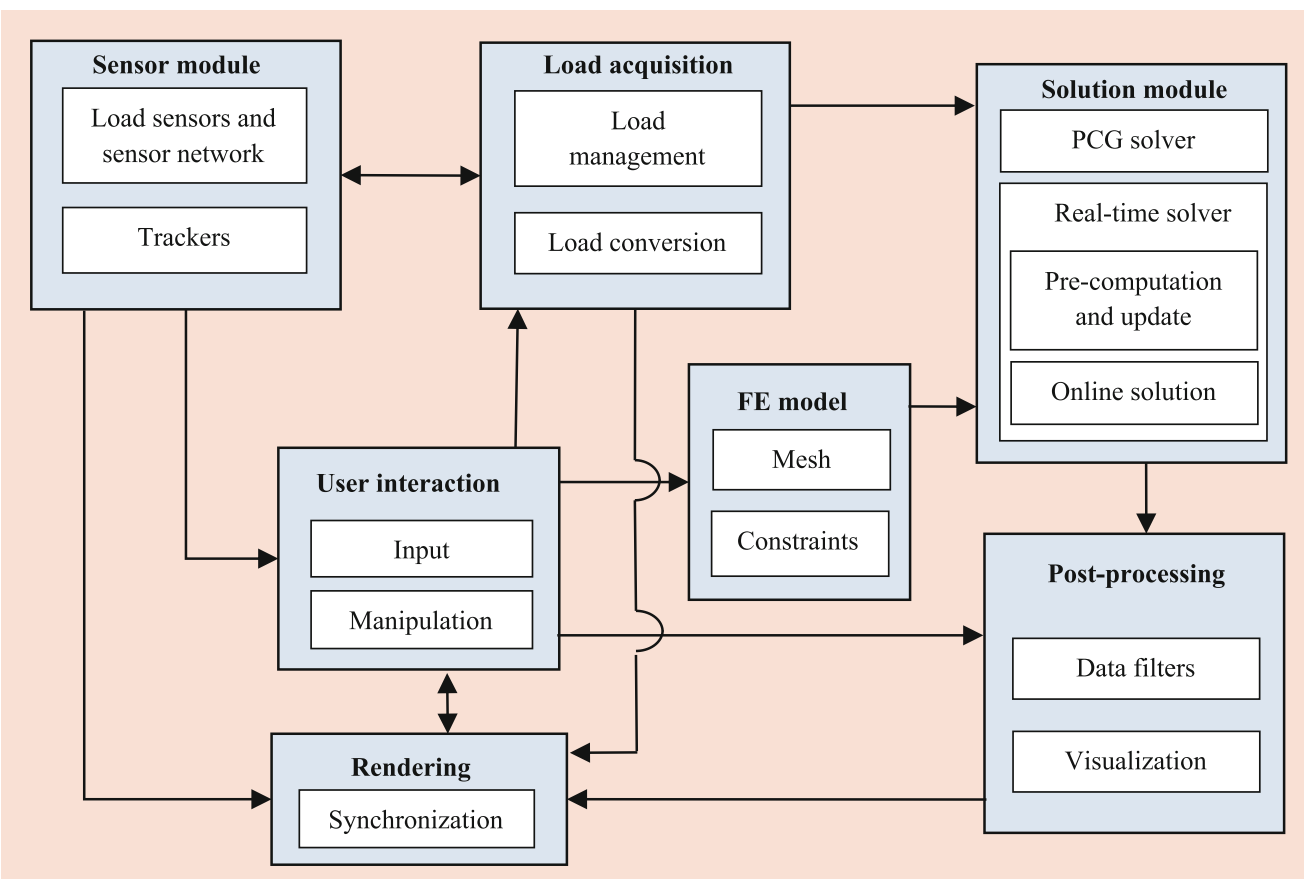


Fig. 22.2 System architecture

tracking data from the sensor module, virtual objects such as the FEA outputs, loads, slicing planes, etc. are registered into the real-world environment and rendered on display devices, such as HMDs, monitors, etc. Different computation threads are used to perform FEA simulation and AR rendering. During real-time simulation, it is necessary to ensure that results are updated in every frame using a synchronization process. When real-time solutions are not achievable, e.g., a model with geometric modifications is solved using the PCG solver, results are updated asynchronously. The user interaction module is established for meeting the various user operation requirements, e.g., applying loads to a structure to investigate the behavior, exploring the volumetric FEA results through data slicing and clipping, manipulating the results for observation, etc. A customized user interface consisting of a 3D input device and a virtual interaction panel has been developed.

22.3.3 AR Environment Setup

Hardware Configuration

Depending on the analyzed objects and user purposes, two types of hardware configurations can be adopted to implement the system as shown in Figs. 22.3 and 22.4. A 3D input device and a virtual panel are created for user interaction, which will be described in the next subsection. Besides, load sensors can be employed to measure load input, and the configuration depends on the specific applications. The configuration of a WSN will be elaborated in a later section.

In Configuration A as shown in Fig. 22.3, the AR-based system is run on a desktop with a monitor for display. A webcam that is supported on a tripod is used for live video capture and marker tracking. The user can move the camera or the analyzed structure to observe the FE model from different perspectives. This setup can provide a stable view for the user to interact with virtual objects, and the display can be shared by multiple users. However, it is not convenient to move the tripod for different perspectives, especially when frequent changes of viewpoint are needed. The immobile hardware configuration is suitable for a tabletop workspace, which does not allow structures of a large size.

In Configuration B as shown in Fig. 22.4, a HMD is used to capture live video and display. The system is run on a laptop that is carried by the user. This configuration is wearable, thus providing greater freedom for the user to move around the analyzed structures for different views. Larger models can be handled using this configuration instead of Configuration A. However, the display cannot be shared by multiple users. It is not easy to keep the camera still to obtain static views when the user is performing operations, which may hinder the user from positioning the 3D input device properly during tasks.

Interaction Tools

Interaction tools are created for data input and system control, which consist of a 3D input device and a virtual panel (Fig. 22.5). The 3D input device is built by attaching a marker cube to a wireless mouse. The marker cube is used for pose tracking. The buttons and wheel of the mouse are used for triggering and parameter adjustment. Using the input device, virtual objects can be manipulated in the AR environment, such as adding virtual loads, slicing planes, activating virtual buttons, etc.

A virtual interaction panel is built for the user to control the system and input necessary data. There are two types of virtual panels [60]. The first type of virtual panels is built in the screen coordinate system, while the other type of virtual

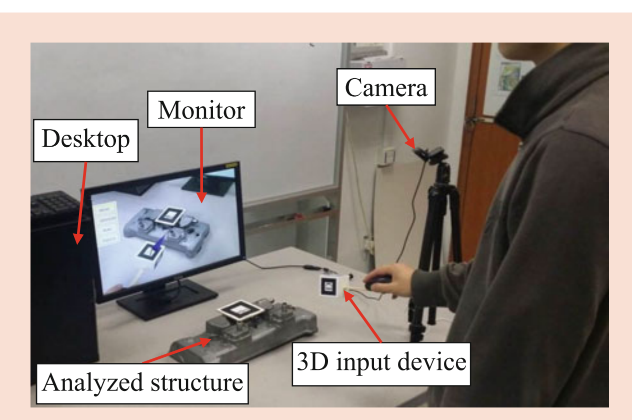


Fig. 22.3 Hardware Configuration A for a tabletop workspace

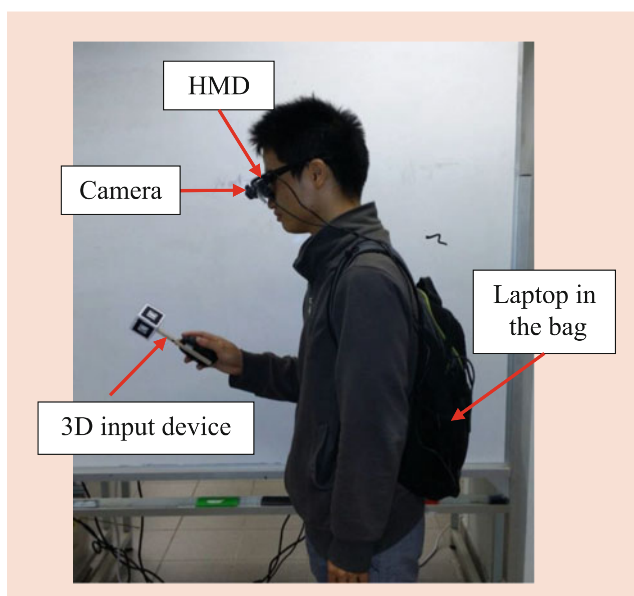


Fig. 22.4 Hardware Configuration B for a mobile system

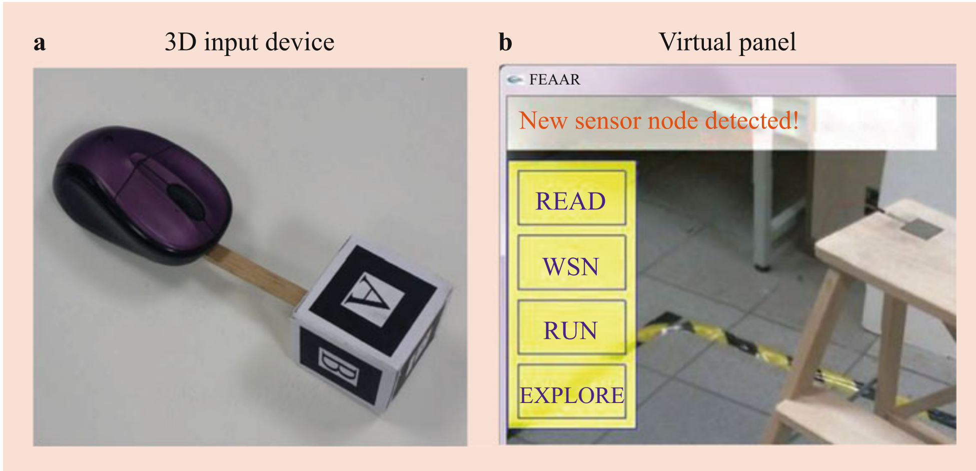


Fig. 22.5 The 3D input device and virtual panel: (**a**) 3D input device and (**b**) virtual panel

panels is in the world coordinate system. The former type is adopted to ensure that the information rendered on the panel can always be observed. The virtual panel has both buttons and text displays. Clicking the buttons with the input device will trigger user events. Users are assisted by text messages during the operations. The panel is arranged at the upper left corner of the screen, such that it will not block the user from observing the structure. Moreover, the user can click the buttons on the panel with a natural hand pose.

Coordinate Systems and Transformation

There are mainly four coordinate systems (CS) in the AR-based system, namely, screen CS, camera CS, structure CS, and device CS. As shown in Fig. 22.6, the screen CS is a 2D coordinate system in the image plane for displaying virtual objects. The origin of this coordinate system is located at the upper left corner of the image. The camera CS, structure CS, and device CS are the local coordinate systems of the camera, physical structure, and input device, respectively. Markers are attached to the physical structure for tracking the pose of the camera CS with respect to the structure CS. The transformation is represented by a 4×4 homogeneous matrix T_S^C. The pose of the device CS with respect to the camera CS, represented by the matrix T_C^D, is tracked by using the marker cube. Based on these transformation relationships, the pose of the device CS with respect to the structure CS can be computed using Eq. (22.1).

$$T_S^D = T_S^C T_C^D \tag{22.1}$$

This configuration facilitates the user to use the input device to perform operations directly on the physical structure, such as applying virtual loads by placing the loads at the specific locations, exploring spatial FEA results with a hand-held slicing plane, grasping virtual models, and placing them at certain locations. All the virtual objects that are attached to the device CS or structure CS will be transformed to the camera CS and projected to the image plane for rendering.

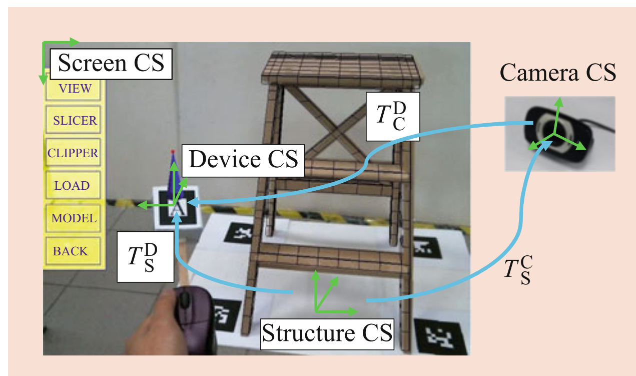


Fig. 22.6 Coordinate systems and transformation

Object Selection Techniques

Object selection is fundamental for interaction. The user operates the input device to select various virtual objects for interaction, e.g., buttons on the virtual panel, elements of the mesh model, etc. 2D selection is achieved using the OpenGL selection mode, which utilizes graphic hardware to support object selection. In this selection mode, a viewing volume is created to pick the objects in this volume. The following steps are performed to implement the selection.

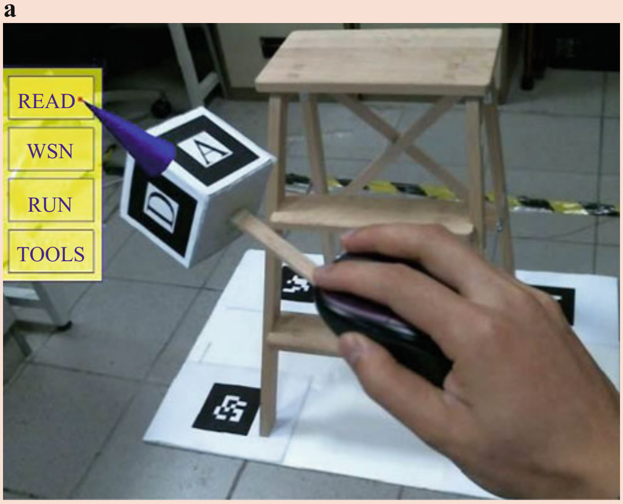
1. Specify the selection buffer using *glSelectBuffer()*, enter the selection mode using *glRenderMode(GL_SELECT)*, and initialize the name stack using *glInitNames()* and *glPushNames()*.

2. Define the viewing volume that is used for drawing the objects to be selected. The viewing volume is defined through setting the model-view matrix, projection matrix, and clipping planes.
3. Draw all the selectable objects and assign names to them in the name stack. OpenGL supports geometric primitives including points, lines, and polygons; hence, some simple geometric objects of a FE model, such as nodes and element faces, can be drawn for selection directly. More complex objects are assembled using these primitives, e.g., a hexahedral element is assembled using quadrilaterals.
4. Exit the selection mode with *glRenderMode(GL_RENDER)*, and interpret the selection buffer which contains the information of the selected objects. The selection buffer contains a list of geometric primitives which intersect the viewing volume, and each primitive corresponds to a hit record in the selection buffer. A hit record consists of four items, i.e., the list of names on the name stack when a hit occurs, the minimum and maximum z values of all the vertices of the primitives that intersect the viewing volume since the last recorded hit, and the contents of the name stack when the hit occurs.

Figure 22.7 illustrates the selection of virtual objects using this 2D selection method. A virtual cone is created on the input device, and the tip of the cone is mapped to a point on the screen, which is drawn as a red dot. This point is used as a 2D cursor. To select objects with this 2D cursor, a projection matrix is created using *gluPickMatrix()* to define the viewing volume. This projection matrix restricts the drawing to a small region in the screen coordinate system. The center of this region is the location of the 2D cursor. The objects that are drawn in this region will be selected and recorded in the selection buffer. By processing the selection buffer, the selected objects are identified. Various objects drawn in the environment can be selected using the 2D cursor, such as the virtual buttons, nodes, element faces, etc.

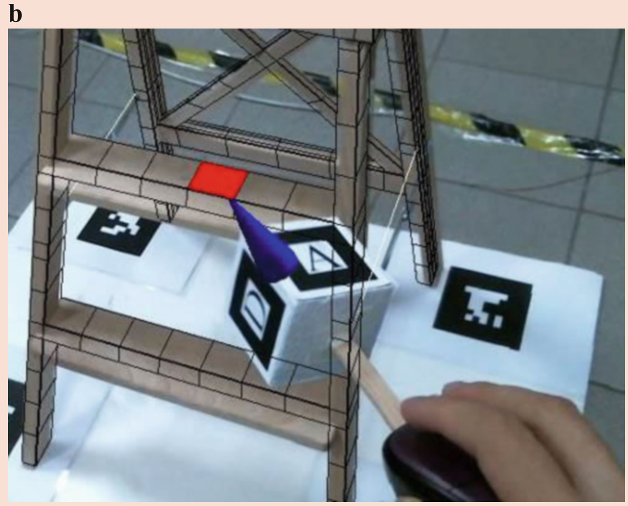
It is straightforward to select visible objects using this 2D selection method. However, the method is not useful for selecting objects in a 3D space, especially when there is occlusion of the object by other objects. Many techniques for 3D selection have been proposed for VR and AR applications [61, 62]. Two classical techniques are virtual hands and rays. However, these techniques are mostly applicable for selecting single objects. Multiple objects are usually selected in FEA, e.g., element refinement, picking element faces to apply force loading. Not many VR-based methods have been reported for multiple object selection [63]. Such methods are not easy to implement as collision detection is needed, and some of them are designed for specific input devices. In this work, OpenGL was used to develop the 3D selection approach. As shown in Fig. 22.8, a virtual cube representing a viewing volume is created using the function *glOrtho()*. The pose of the viewing volume is controlled through setting the model-view matrix. By setting this matrix using the transformation matrix between the device CS and structure CS, the viewing volume is attached to the input device. Therefore, the user can manipulate the input device and select the objects that

a



Select a virtual button

b



Select an element face

Fig. 22.7 Selecting virtual objects using the 2D selection method: (**a**) select a virtual button and (**b**) select an element face

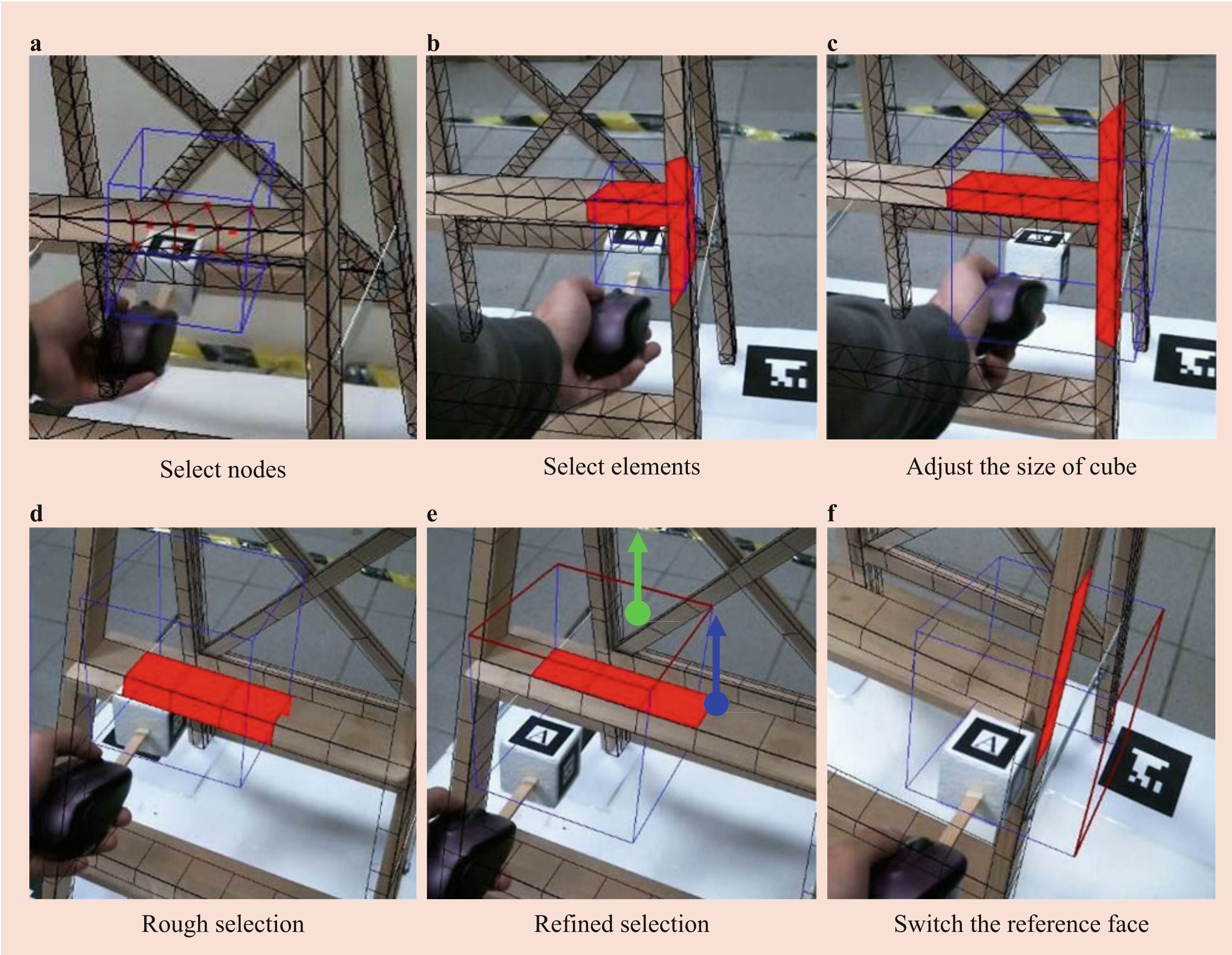


Fig. 22.8 3D multiple object selection using a virtual cube: (**a**) select nodes, (**b**) select elements, (**c**) adjust the size of cube, (**d**) rough selection, (**e**) refined selection, and (**f**) switch the reference face

intersect the virtual cube. The size of the viewing volume is controlled through setting the clipping planes using *glOrtho()*. The user can adjust the size by rolling the mouse wheel (Fig. 22.8c).

In this 3D selection approach, a user can select a multiple number of objects in the AR environment. Selection of element faces is achieved as depicted in Fig. 22.8d. This selection method is not very practical as all the faces intersecting with the virtual cube will be selected while the user is only interested in faces with a specific orientation such as those which are horizontal. In order to circumvent this, the selection method is fine-tuned by using one of the virtual cube faces as a reference face (Fig. 22.8e). The normals of the reference face and the element faces which intersect the virtual cube are computed. Element faces are selected on condition when the angles between the element faces and the reference face are smaller than a certain threshold. The user can select the faces more specifically by either rotating the virtual cube or changing the reference face (Fig. 22.8f).

22.3.4 Summary

To enhance the visualization and interaction of FEA using the AR technology, a system has been designed to integrate FEA simulation, AR visualization, and sensor measurement. This section describes the architecture of this system, as well as the hardware configuration and interaction techniques. A 3D multiple object selection method has been developed to enable efficient selection of virtual objects for FEA tasks. The prototype system can acquire load data directly from the load cells and provide AR interfaces for user interaction. The simulation results can be visualized on the corresponding real objects. With these capabilities, this system can support intuitive and efficient structural analysis.

22.4 Visualization and Exploration of FEA Results

Scientific visualization is an important tool for data investigation. By creating graphic representations of 3D or higher-dimensional data, it provides a visual approach for understanding and investigation of data. The visualization process, i.e., the process of converting raw data to a displayable image, can be described using a pipeline [64, 65] as shown in Fig. 22.9. Three main phases are involved in the pipeline, i.e., data enrichment and filtering, visualization mapping, and rendering. Data enrichment and filtering involve the processes of modifying or selecting the data, in order to enrich the information content of the data or extract the data to be visualized. The common methods used in this phase include interpolation, sampling, domain transformation, noise filtering, etc. Next, visualization objects are constructed, which contain attributes that are mapped to the data, such as geometry, color, transparency, etc. The visualization objects are transformed into images in the rendering phase, and the typical operations include viewing transformation, lighting calculations, etc.

A number of visualization systems are constructed based on this visualization pipeline, such as VTK, Prefuse, and OpenDX . This section presents the method to integrate VTK into the AR-based system, so as to visualize FEA results on the analyzed structures. Furthermore, intuitive interaction methods are developed for enhanced data exploration based on data filtering techniques.

22.4.1 Scientific Visualization with VTK

VTK is used for interactive visualization as well as post-processing of FEA results. It is an object-oriented library built on top of OpenGL. Figure 22.10 depicts the process to visualize FEA results using VTK.

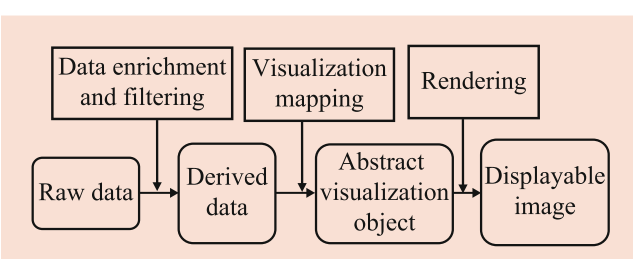


Fig. 22.9 A general visualization pipeline

Each dataset processed consists of an organizing structure and associated data attributes. Cells and points are used to specify the topology and geometry of the organizing structure. Various types of data attributes can be associated with the cells or points, including scalars, vectors, normals, tensors, etc. VTK can handle both structured and unstructured FE meshes. A vtkUnstructuredGrid dataset is used to establish the FE mesh by assigning the elements to the cells and nodes to the points, with elemental or nodal solutions as the associated attributes. vtkFilters are optional for data manipulation and transformation. The data is mapped to the graphic primitives using vtkDatasetMappers, i.e., visualization mapping. A vtkLookupTable is constructed to code the scalar solutions in different colors in the hue-saturation-value (HSV) color space. The range of the scalar solutions is specified using SetScalarRange(). The number of colors in this lookup table is set using SetNumberofTableValues(). A linear interpolation is performed to the hue attribute within a range that is set to [0, 0.667] using SetHueRange(). The other attributes, i.e., saturation and value, are set to 1. After the mapping, the visualization entity is formed using a vtkActor, and the visual properties can be adjusted, such as opacity and color. The vtkActor is finally rendered using a vtkRenderer, and the viewing transformation is set using a vtkCamera. A vtkActor can be rendered with vtkCameras of different configurations. Figure 22.11a shows an example in which the von Mises stresses of a bracket are visualized by using this color coding technique. The displacement values can also be visualized with this approach. To visualize the deformation, i.e., the 3D displacement vectors, a vtkWarpVector filter can be employed (Fig. 22.11b). The exaggerating scale of the deformation can be adjusted using SetScaleFactor().

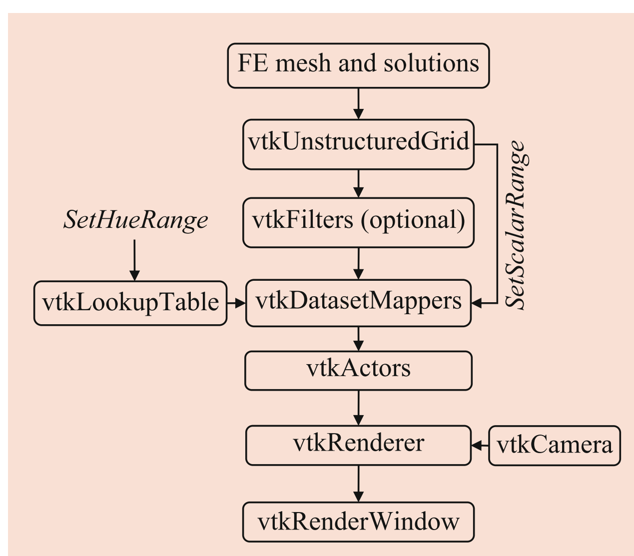


Fig. 22.10 Visualizing FEA results with VTK

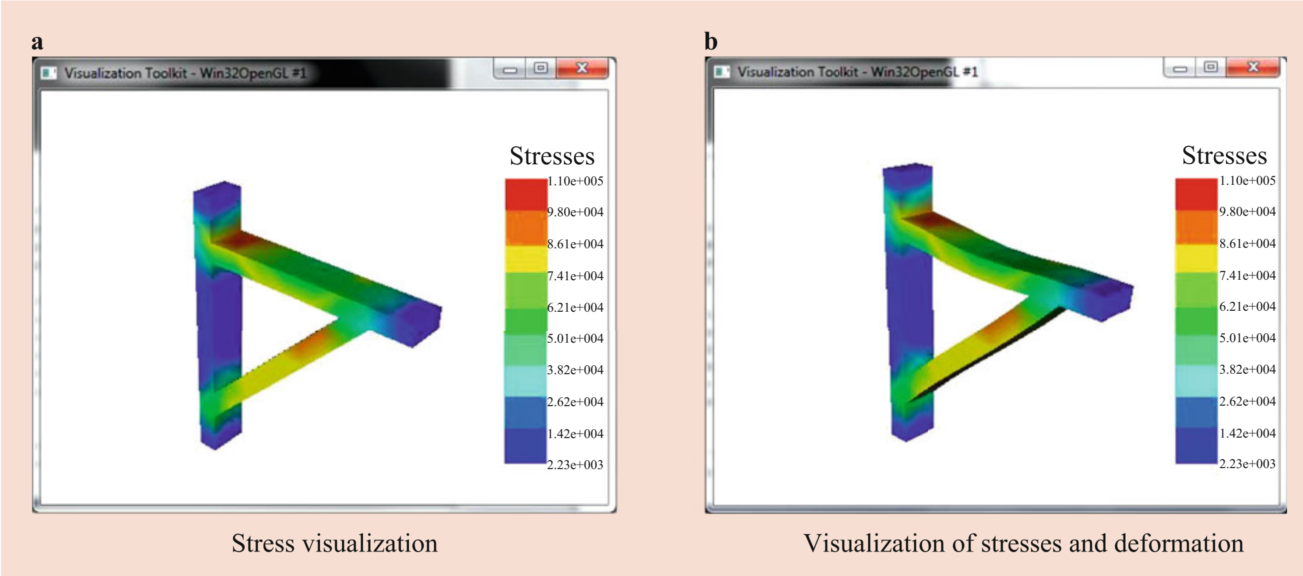


Fig. 22.11 Visualizing the FEA results of a bracket: (**a**) stress visualization and (**b**) visualization of stresses and deformation

22.4.2 Scientific Visualization in AR

Integration of VTK and AR

To visualize FEA results in real-world environments, the visualization entities, i.e., vtkActors, are registered in the relevant coordinate systems, such as the structure CS and device CS. As shown in Fig. 22.12, the viewing transformation for a vtkRenderer is controlled through a vtkCamera. By configuring the vtkCamera with the intrinsic and extrinsic matrices from camera calibration and tracking, respectively, vtkActors can be registered into the coordinate systems that are tracked using the camera and finally rendered with a background of real-world environments. The user can control the visualization parameters through AR interfaces, such as the parameters of vtkFilters, the visual properties of vtkActors, etc. With the use of various vtkFilters and established AR interfaces, the user is able to intuitively manipulate and investigate FEA results that are visualized on the physical structure.

In this research, ARToolKit is employed to build the AR-based environment, in which rendering is performed using the OpenGL Utility Toolkit (GLUT). In VTK, however, rendering is performed using vtkRenderer. The integration of ARToolKit and VTK in rendering is achieved in two steps, i.e., merging the image frames and configuring the viewing transformation. There are two approaches to merge the image frames from these two systems. One approach is to display the vtkActors in the render window that is created using ARToolKit. The other approach is to transfer the video frames and camera matrices from ARToolKit to VTK pipeline. The former approach has been implemented because the process is straightforward and less data transformation is needed. In this approach, the handle of the render window created in ARToolKit is acquired and used to initialize the vtkRenderWindow in VTK. The erasing function of the vtkRenderWindow is switched off to prevent it from erasing the video frames acquired by ARToolKit. After merging the image frames, the viewing transformation in VTK should be configured using the camera matrix from ARToolKit. A camera matrix can be decomposed into extrinsic and intrinsic matrices. The extrinsic matrix transforms objects from world coordinate system to camera coordinate system, while the intrinsic matrix transforms objects from camera coordinate system onto the image plane. In ARToolKit, the intrinsic and extrinsic matrices of the camera are computed and loaded into the projection matrix and model-view matrix, respectively. In VTK, the viewing transformation is configured by setting the projection and model-view transformations of the vtkCamera. The VTK source code is properly modified, such that the viewing transformation of the vtkCamera can be configured directly with the camera matrices computed with ARToolKit. Figure 22.13 shows a flowchart for the integration of ARToolKit and VTK.

Occlusion Handling

Occlusion occurs when real objects are overlaid with virtual objects. As can be seen in Fig. 22.14a, incorrect views are generated when a bracket is overlaid with the mesh graphics obtained using the *vtkExtractEdges* filter. The physical bracket is supposed to block some parts of the virtual mesh. The overlaying without occlusion handling leads to misunderstanding of the FE model.

To address this problem, a mask is created to remove the graphics which are supposed to be invisible. The mesh model

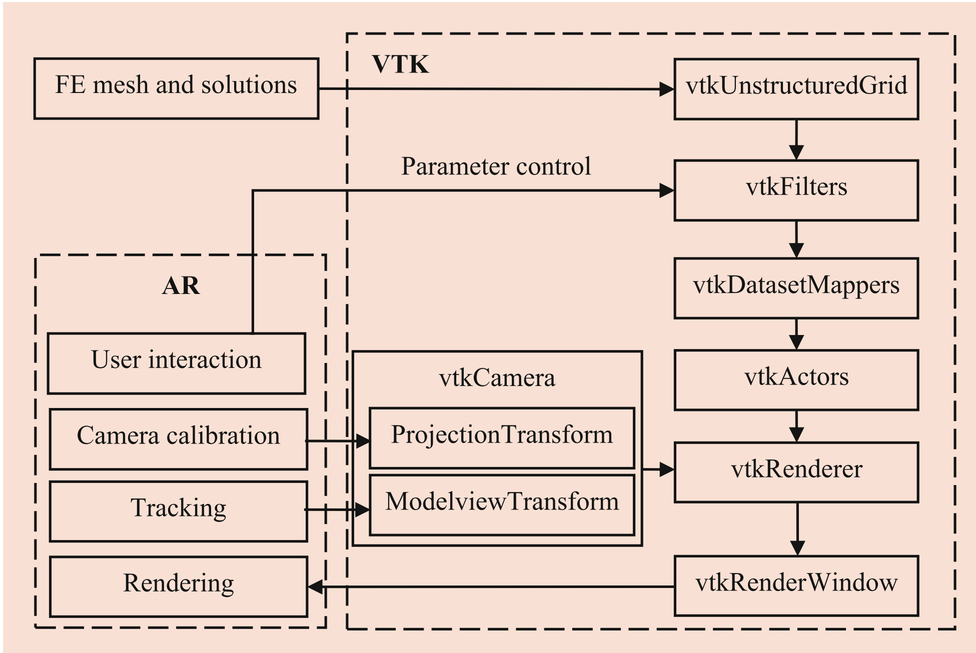


Fig. 22.12 Integration of VTK and AR

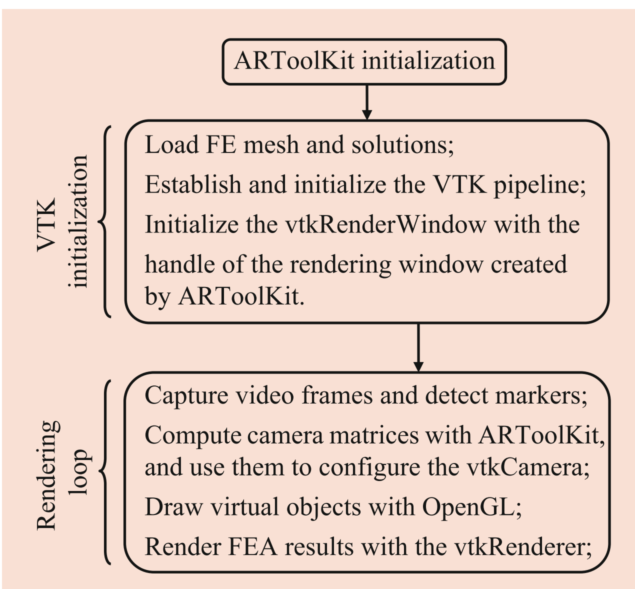


Fig. 22.13 Flowchart for the integration of ARToolKit and VTK

is utilized to draw the mask before rendering the virtual objects. Next, the color buffer of the mask is erased but the depth buffer remains. When virtual objects are rendered, depth tests are performed using the depth buffer of the mask, such that only the visible parts of the virtual objects will be drawn. With this approach, correct images are achieved as shown in Fig. 22.14b.

22.4.3 Data Manipulation and Exploration

Based on the integration of VTK and AR, FEA results can be visualized on the corresponding physical structure. Moreover, the user can manipulate the data directly using a 3D input device. Interactive data slicing and clipping methods are developed to enable detailed exploration of FEA results in an intuitive manner. Figure 22.15 depicts the approach to achieve data manipulation and exploration in the AR-based environment.

By tracking the pose of the viewpoint with respect to the physical structure, FEA results can be superimposed on the structure. By moving the viewpoint of the camera, the user can observe FEA results from different perspectives. However, this approach may not be feasible due to physical restrictions and the limitations of the tracking method. The trackers may not work well if the markers are beyond the field of view from the camera. To circumvent this, direct manipulation of data is used. The FEA results can be attached to the 3D input device by transforming the vtkActor from the structure CS to the device CS and configuring the vtkCamera with the transformation matrices from device tracking. As a result, the user can manipulate the visualized data directly by hand. Zooming can be achieved by moving the data closer to the viewpoint or triggering the mouse to send a scale factor to the vtkActor.

Data slicing and clipping are well-established scientific visualization techniques for exploring volumetric datasets. Intuitive and efficient data exploration can be achieved using AR. The filters vtkCutter and vtkClipDataset are utilized to

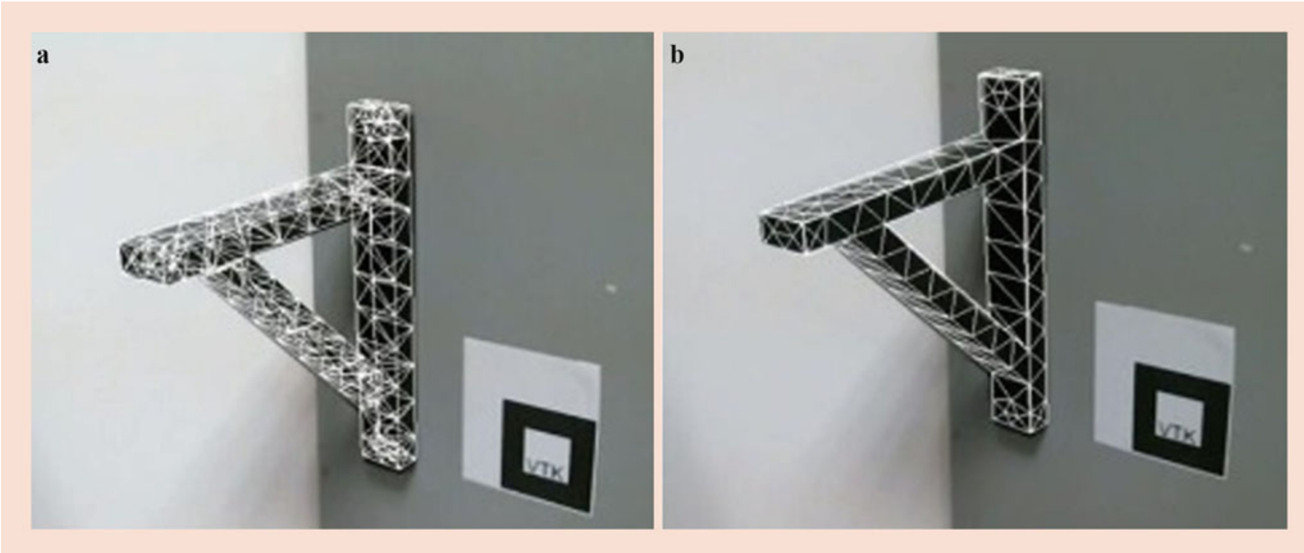


Fig. 22.14 Occlusion handling for overlaying: (**a**) mesh without occlusion handling and (**b**) mesh with occlusion handling

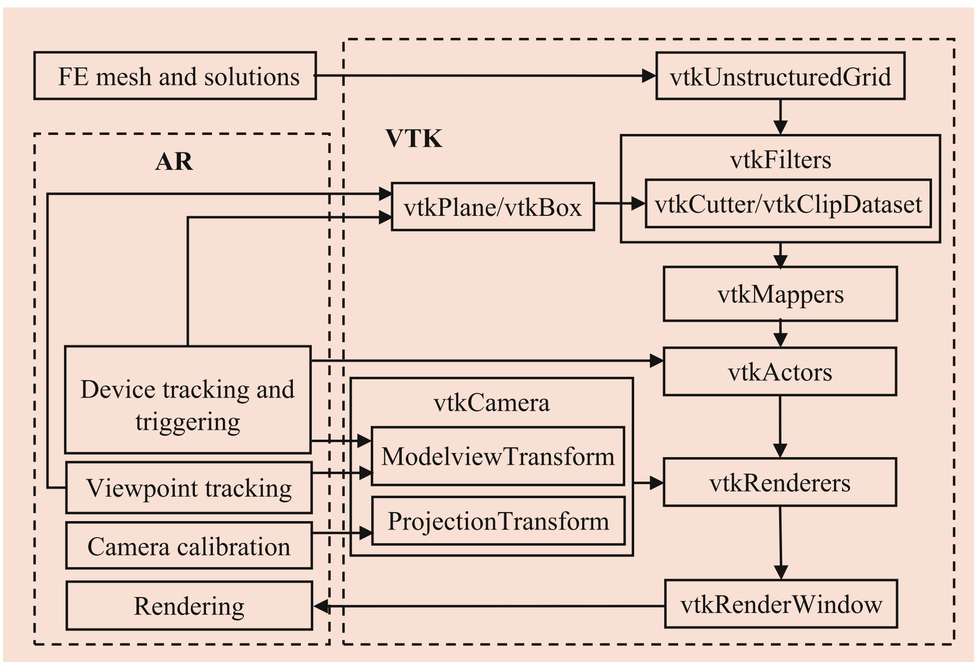


Fig. 22.15 Schematics for data manipulation and exploration

develop the interaction methods. To examine the interior of a volumetric dataset, a slicing plane can be controlled to cut through the dataset, such that the data on the slicing plane is obtained. Using the data clipping method, a user can clip a dataset using a cube or a plane, in order to isolate the portions of data of interest. The planes and cubes for slicing and clipping are defined using vtkPlane and vtkBox, respectively. Their sizes can be controlled by the user. When attaching a virtual plane or cube to the device CS, the user can manipulate it manually. When attached to the camera CS, the plane or cube can be manipulated by moving the viewpoint. By manipulating the virtual plane or cube, the user can slice or clip the data interactively on the physical structure. Multiple data slices or clips can be created in different locations on the structure. The vtkActors of these data slices or clips are registered in the structure CS. A data slice or clip can also be attached to the device CS for user manipulation.

22.4.4 Summary

This section presents the method to enhance the visualization and exploration of FEA results with AR. As a scientific visualization tool, VTK is integrated with the AR-based system for visualization and post-processing of FEA results. With such integration, the system can superimpose FE meshes and results on the corresponding physical structure, which facilitates interpretation of the FEA results. Furthermore, intuitive data exploration is achievable by establishing AR interfaces based on various data filters that are provided by VTK. This section presents several AR-assisted data exploration methods, including data manipulation, slicing, and clipping. Data slicing and clipping can be conducted by manipulating the handheld 3D input device or moving the viewpoint.

22.5 Real-Time FEA in AR

Aided by AR, the FEA simulator can acquire boundary conditions directly from the actual operating environment of the structures. Real-time FEA solvers are applicable for generating solutions at a fast speed. The combination of real-time FEA simulation with automatic load acquisition results in an innovative way to investigate structural behavior. This section presents the method to achieve real-time FEA simulation in the AR-based environment with loads acquired using sensors or via a natural interface.

22.5.1 Real-Time FEA Solution

Time integration strategies are used for dynamic loading simulation, and these are the explicit and implicit methods. In deciding the time step, it is necessary to consider the material properties, mesh size, and natural frequencies. Real-time simulation can be achieved if the computation time required for a time step is smaller than the time step duration [66]. Since engineering structures are usually built with materials of high stiffness, this would limit the time step to very small value. Real-time dynamic response simulation will be difficult, but this situation is not always necessary for varying loads. Loading can be assumed to be static if the time taken is at least three times higher than the natural period of the structure [67]. Many real-life situations adopt this quasi-static approximation. Since linear elastic models are mostly used in practical structural analysis, a real-time solver using the inverse of the stiffness matrix computed a priori is often developed [47, 48]. Such a system can be adapted quite easily to suit other types of real-time solvers as all of them have the same function, i.e., calculating time-dependent displacements $u(t_i)$ with corresponding load inputs $f(t_i)$, $i = 1, 2, 3 \ldots, n$.

An elastic FE model can be expressed using a sparse linear system in Eq. (22.2), where K is the stiffness matrix and u and f are the vectors representing nodal displacements and loads, respectively.

$$Ku = f \tag{22.2}$$

The solution of this linear system is carried out in two phases, i.e., offline pre-computation and online solution. In the offline pre-computation phase, the stiffness matrix K is established after the mesh models, material properties, and constraints have been specified. After that, the inverse of the stiffness matrix K^{-1} is computed. During the online solution phase, the load vector f is updated with sensor data, and the displacement vector u is calculated using a matrix-vector multiplication as in Eq. (22.3).

$$u = K^{-1}f \tag{22.3}$$

The online solution phase is a linear superposition of the inverse matrix columns. Using the displacement solution, stresses anywhere inside an element can be determined using Eq. (22.4), where E is the elasticity and B is the strain-displacement matrix, respectively. Nodal stresses can be evaluated using direct evaluation at nodes, extrapolation from Gauss points, etc. [51].

$$\sigma = EBu \tag{22.4}$$

By skipping the zero entries in the load vectors, the online solution process can be accelerated. Furthermore, if one is only interested in stresses in certain regions, the online solution time can be further shortened by computing results only for the elements involved. However, additional effort is required to extract the relevant elements from the mesh data.

22.5.2 Computation of Inverse Stiffness Matrix

Based on a mesh model, element stiffness matrices are computed and assembled to form the global stiffness matrix. A full global stiffness matrix is singular and thus invertible. After proper constraints have been applied, this matrix is reduced to a non-singular, symmetric, and positive definite matrix. To compute the inverse matrix, a unit load is applied to each degree of freedom (DOF) of the FE model, resulting in a number of linear systems. These linear systems are solved using standard solvers, e.g., conjugate gradient methods, Gaussian elimination, etc. For the linear system in which a unit load is applied to the i-th DOF as shown in Eq. (22.5), the displacement solution u corresponds to the i-th column in the inverse stiffness matrix.

$$\begin{bmatrix} k_{1,1} \cdots\cdots\cdots & \cdots \\ \cdots\cdots\cdots\cdots & \cdots \\ \cdots\cdots k_{i,i} \cdots & \cdots \\ \cdots\cdots\cdots\cdots & \cdots \\ \cdots\cdots\cdots\cdots & k_{N,N} \end{bmatrix} \begin{bmatrix} u_1 \\ \cdots \\ u_i \\ \cdots \\ u_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (22.5)$$

If N is the number of DOFs of the FE model, N linear systems should be solved to obtain the inverse matrix. This offline computation may take minutes to hours depending on the matrix size and system performance. When iterative solvers are used, the computation time depends on the threshold error and stop criteria set for the required precision.

Matrix Inversion with PCG Method

The linear systems in Eq. (22.5) are solved using the PCG method, and the solutions are derived as the inverse of stiffness matrix. As the stiffness matrix is sparse, i.e., a high percentage of the entries of the matrix are zero, it is not computationally economical to store and solve a sparse linear system using general methods of linear algebra, which would waste storage space and computation power on the large number of zero entries. Instead, the storage and solution are usually carried out by taking advantage of the special structure of the stiffness matrix. However, the inverse matrix is normally a dense matrix; thus, it is stored using a regular storage method.

There are several methods for storing a sparse matrix, such as compressed row/column storage, coordinate format, skyline storage, etc. [68]. The compressed row/column storage is widely used in FEA. The ANSYS software adopts the compressed column storage, which is also called the Harwell-Boeing sparse matrix format [69]. In this storage scheme, a sparse matrix is stored in three arrays, named COLPTR, ROWIND, and VALUES. COLPTR array stores the subscripts of the first nonzero entries in each column. ROWIND array stores the row indices of the nonzero entries. VALUES array stores the nonzero entries in each column. The stiffness matrix is symmetric, and hence only the lower triangular part of the matrix is filled.

A best known iterative method for solving sparse linear systems in structural analysis is the conjugate gradient algorithm. A significant advantage of this method is that only matrix-vector multiplication is performed on the sparse matrix. The multiplication can be very efficient for the sparse matrix stored in compressed column storage format. The conjugate gradient algorithm is based on the premise of minimizing the function (22.6).

$$f(x) = \frac{1}{2}x \cdot A \cdot x - b \cdot x \quad (22.6)$$

The linear system $A \cdot x = b$ is solved with an iterative process to find an x that minimizes $f(x)$. The iterative process starts with an initial guess of the solution. The gradient of $f(x)$ for this initial solution is taken as the initial search direction. In each iteration, the optimal step size is determined for the current search direction to update the estimation for x, and the search direction is updated based on the previous search direction and the gradient of $f(x)$ at the current step. A preconditioner is usually employed to accelerate the convergence of this iterative process. The accuracy of the solution is controlled using the error threshold and stop criterion, e.g., the relative residual $\|b - Ax\|/\|b\|$ is less than a threshold. A solution of higher accuracy requires longer computation time. Therefore, the error threshold and stop criteria can be adjusted to reduce the computation time while the accuracy is acceptable.

Matrix Inversion Using External FEA Program

The computation of the inverse stiffness matrix is straightforward if the AR-based system has a built-in FEA program, because the mesh data and stiffness matrix can be accessed directly. When external FEA programs are employed to perform FEA tasks, such as mesh generation and matrix formation, the stiffness matrix can be extracted for computing the inverse matrix. However, the data structure of external FEA programs may not be accessible or readily available. To obtain the inverse stiffness matrix, an approach is to utilize the external FEA program to compute it directly. This approach is derived from the method called the sensitivity response method that is used by Doyle [70] for partially defined structural problems.

Most FEA programs provide scripting tools for the users to control FEA tasks and data output. By using the scripts, a matrix inversion program can be created as illustrated in Fig. 22.16. In this program, the nodes that may be subjected to loads are specified, such that the columns of inverse matrix will be computed for these nodes only. Next, script files are generated to perform analysis for the load cases in which a unit load is applied to each DOF associated with the nodes. These scripts are then executed by the FEA program. With the FE model built beforehand, the stiffness matrix is formed, and the equations are solved in the FEA program. The displacement solutions are output to the result files, which correspond to the columns of the inverse matrix. These files are read, and the columns are combined by the matrix inversion program.

22.5.3 Load Acquisition

Load data is acquired online for FEA solution. The load data will be converted to nodal loads, and the inverse stiffness matrix is multiplied to generate the displacement solution at a fast speed. Loading can be manually applied using a 3D input device or input from sensors. Only point loads will be

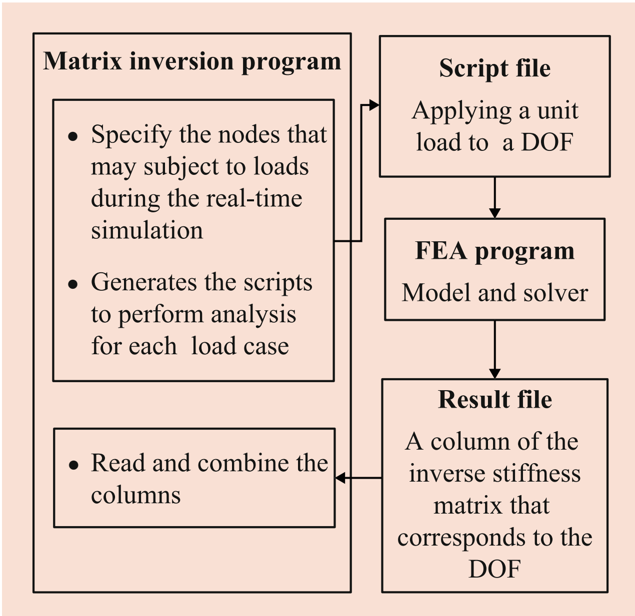


Fig. 22.16 Schematics for computing inverse stiffness matrix with an external FEA program

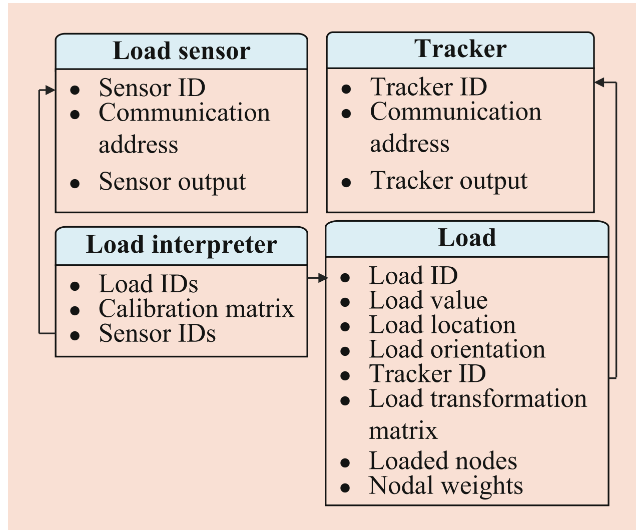


Fig. 22.17 Data structure for load acquisition

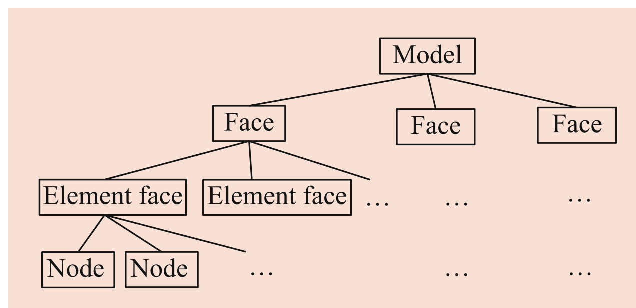


Fig. 22.18 Hierarchy data structure of the surface mesh

considered while the adaptability for distributed loads will be demonstrated.

Load Management and Conversion

All the loading data should be converted to nodal forces for FEA computation. A data structure is used to manage the loads as depicted in Fig. 22.17. Each load sensor or tracker is indexed with a unique ID number and a communication address. Load sensor outputs are interpreted into loading values using a calibration matrix. The interpreter is used to calibrate force sensors and where loads have been identified from local strains or deflections [70]. Each load has a unique ID number and relevant parameters for load conversion. In the case of virtual loads, the value, direction, and location are input using the 3D input device.

Two operations are performed for converting the acquired loads into nodal loads, namely, coordinate transformation and load allocation. According to the directions of the loads, a load transformation matrix is computed for each load. Using these matrices, the loads acquired from user input or load sensors are transformed into the structure CS. Next, the loads are allocated to the nodes subjected to these loads, i.e., converted to nodal loads. The allocation is achieved by assigning weights to the loaded nodes, which can be derived by computing the equivalent nodal forces with applied unit loads. The loaded nodes and nodal weights for load allocation are determined according to load locations. In this study, loads are applied on the surfaces of the model. The surface mesh data contains the information of the model faces, element faces, and nodes, which is organized using a hierarchy data structure as shown in Fig. 22.18. With the surface mesh data, the load locations on the model faces and the loaded nodes can be determined using polygonal collision detection methods, e.g., ray casting.

A data table is established to store the information of all the loads, i.e., ID number, value, location, orientation, transformation matrix, etc. An entry in the data table corresponds to a load. In each rendering loop during the real-time simulation, the load vector f in Eq. (22.3) is updated with the nodal loads computed throughout the table using Eq. (22.7), for updating the FEA results for every frame.

$$f_i = \sum_{j=1}^{n} w_{ij} \cdot T_j \cdot F_j \quad (22.7)$$

In Eq. (22.7), w_{ij} is the weight of node i subjected to load j. F_j and T_j are the corresponding load value and transformation matrix [75]. Using this data structure and computation approach, adding, removing, and adjusting of loads can be done quite easily.

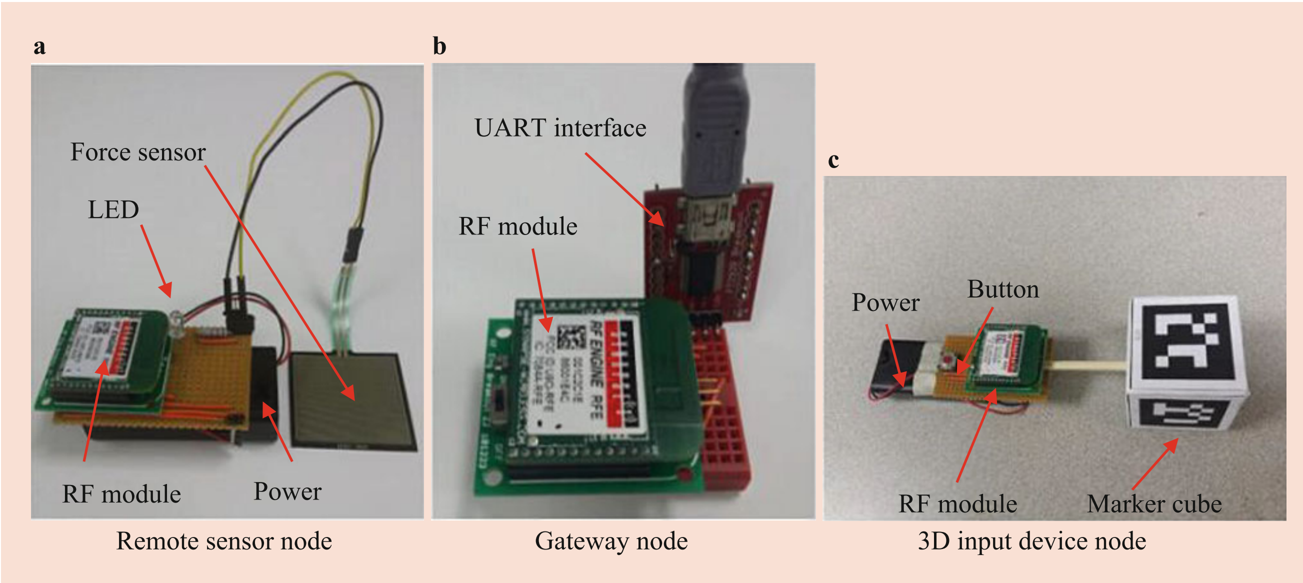


Fig. 22.19 Three types of nodes in the WSN: (**a**) remote sensor node, (**b**) gateway node, and (**c**) 3D input device node

WSN Configuration

Environmental conditions are commonly monitored using WSN, which normally consists of sensor nodes, gateways, and client application software. A WSN was built using the Synapse's SNAP system (Synapse's SNAP, n.d.) for acquiring data for real-time analysis. A Synapse RF module consists of a RF transceiver and a microcontroller that runs Python codes. Figure 22.19 shows the remote sensor node and the gateway node built for the network using the RF modules. The 3D input device can also be involved in the network by using a remote node to detect the button presses. The load data is acquired by the remote sensor node and transferred to the client application, while the gateway node provides the connection between the sensor nodes and the client application. Connection of a client to multiple sensor nodes only requires one gateway node, and the communication uses standard XML-RPC protocol [71], which is a remote procedure call (RPC) protocol which uses XML to encode the calls and HTTP to transfer the information. The nodes form a network automatically when they are within range of each other.

Since real-world loads are measured in the sensor coordinate systems, they would need to be transformed to the structure CS for load conversion. A 3×3 rotation transformation matrix is sufficient for the rotation transformation of the tri-axial force sensors. The transformation matrices can be generalized for uniaxial and biaxial load sensors by modifying some entries to zero. After a load sensor is registered in the AR environment, the loaded nodes and their weights can be determined according to the load location, such that the corresponding load conversion relationship is established and will not change during real-time simulation.

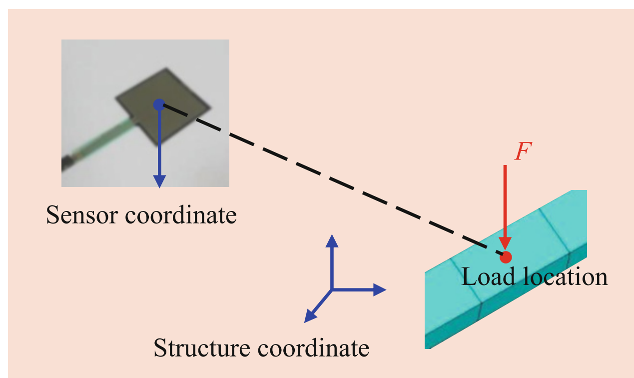


Fig. 22.20 Load conversion for the force sensor

As an example, Fig. 22.20 illustrates the load conversion for the force sensor.

A sensor node corresponds to an entry in the load data table. Using the data table, multiple sensor nodes are registered in the structure CS based on the following procedures:

(a) A sensor node is first switched on by the user. The node will continue to send requests for registration via RPCs with its network address.
(b) When the system receives the requests, it informs the user to register the sensor node.
(c) The user will proceed to specify the sensor coordinate system and load location using the input device. Input methods can be developed according to sensor types and input devices, e.g., casting a ray normal to the loaded area to register a uniaxial sensor attached on it and aligning the coordinate system of a 3D input device to that of a tri-

axial sensor. After that, the load transformation matrix is computed. The loaded nodes and nodal weights are determined with the specified load location. An entry for this sensor node is added to the data table subsequently.

(d) An ID number will be assigned via a RPC after the successful registration of the sensor node. Upon receipt of the RPC, the LED will be turned on by the sensor node to inform the user, and the force values and the ID number will be sent instead of registration requests. The force values are generated by multiplying the outputs of the force sensor with a calibration coefficient.

(e) Once the force values and ID number have been received, the system will index the entry according to the sensor node, and the load value is updated.

An important feature of the established WSN is that the same code is used by all the sensor nodes, and it is very efficient to add new sensor nodes online to meet the changing loading conditions.

Application of Virtual Loads

Virtual loads can be attached easily to the input device and manipulated by the user. A solid cone is used to represent a virtual point load as shown in Fig. 22.21. The load direction is the pointing direction of the cone, and the load value can be adjusted by rotating the mouse wheel. This virtual load is tracked by the camera. When the cone tip is near to the face of the model, a ray will radiate from the tip toward the face. The vertices of the intersected polygon will be the loaded nodes, and the intersection point will be the load location. A virtual load can be placed at a specific location on the structure by attaching the load to the structure CS. If adjustments are required, the user can select the load and attach this to the device to change the load parameters. The altered loading will be computed and reflected in the FEA results.

The loading approach can be modified for applying distributed loads, which is akin to applying pressure to the faces of a structure. Using the refined 3D selection method, the user can use a virtual cube to select a model face by selecting its element faces. With known pressure and areas of the selected faces, the force on each element face can be computed and allocated to the relevant nodes for FEA computation. Figure 22.22 shows the graphic representation of pressure. The loaded element faces are marked with crosses to indicate the loaded area, and the load direction is indicated by the cones.

22.5.4 System Workflow and Time Synchronization

Figure 22.23 shows the workflow of the real-time simulation, which starts with FE model pre-computation and initialization. Next, three computation threads are used to run the rendering loop, real-time solution, and load acquisition. When a WSN is registered, load acquisition and real-time solution will be activated. Load acquisition continues to receive sensor data and refreshes the load data table. Real-time solution converts the current load data into nodal loads and generates the solutions for the post-processing module to update the VTK datasets.

Load acquisition, FEA simulation, and rendering loops are performed at the same time but in different computation threads. A time synchronization scheme is developed as shown in Fig. 22.24 for different computation times and process dependencies. Load acquisition can have the highest

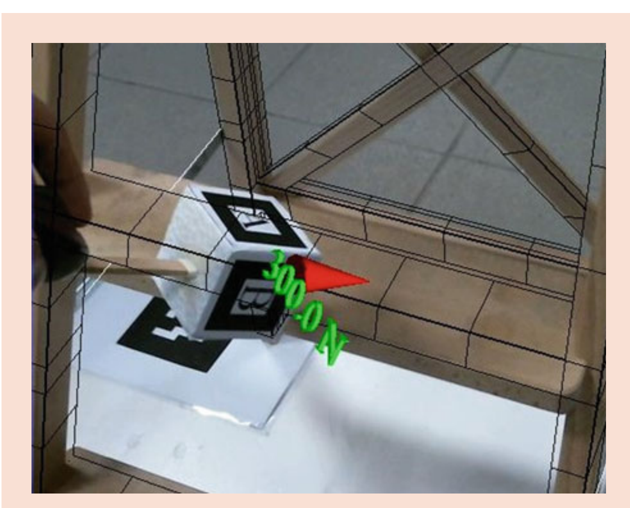


Fig. 22.21 Graphic representation of a point load

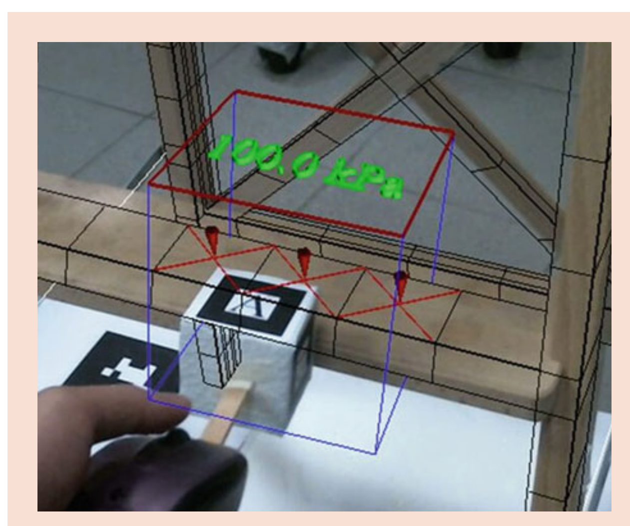


Fig. 22.22 Graphic representation of pressure

updating rate using appropriate WSN equipment or by manipulating the virtual loads. The simulation rate would need to be higher than the frame rendering rate for the simulation results to be updated and rendered in each frame. For performing dynamic analysis, multiple simulation steps are needed for a rendering frame, but for quasi-static approximation, one step is sufficient. The updated load data is input to the simulator for a simulation step at the start of each rending loop. The load data and the simulation step synchronization is not performed as the time offset is usually smaller than the load sampling interval. After completing AR tracking, simulation results are input to the post-processing module. If the tracking results arrive sooner than the simulation results, the rendering loop will need to wait for the simulation results. When both the tracking and simulation results are obtained, they are sent to the post-processing module for filtering and visualization.

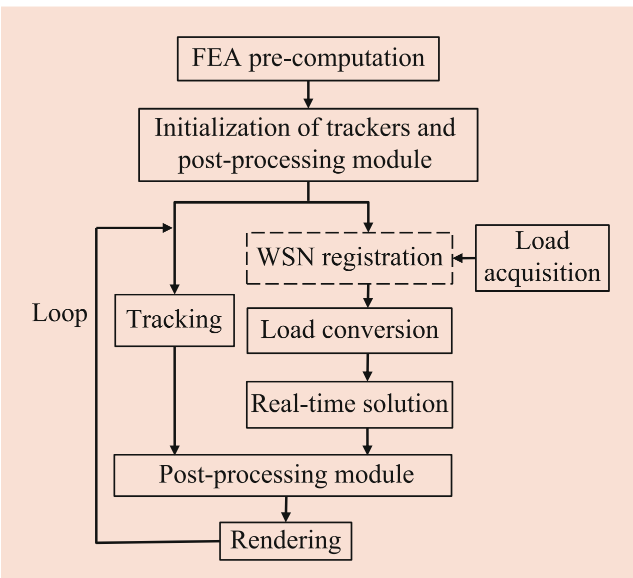


Fig. 22.23 Workflow diagram for the real-time simulation

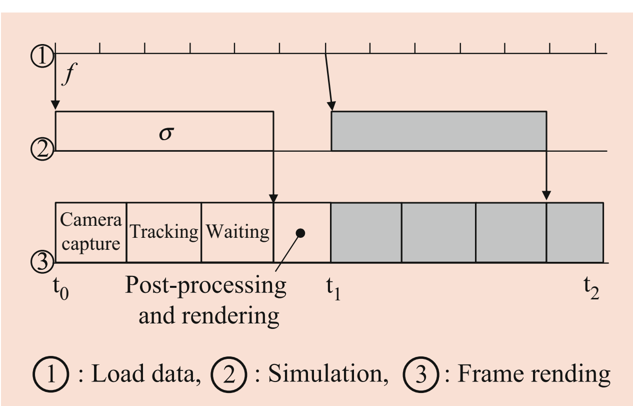


Fig. 22.24 Time synchronization scheme for the real-time simulation

22.5.5 Summary

This section presents the methodology for performing real-time FEA on structures under varying loads. Linear elastic FE model is adopted, and the solution is accelerated by pre-computing the inverse stiffness matrix offline. With the use of AR, the coupling of load acquisition and real-time FEA simulation becomes possible and promising. A load acquisition module is established to manage and convert the acquired load data into nodal loads for solution. A WSN is integrated into the system to acquire spatial distributed loads, and an AR-assisted sensor registration method is developed, which enables interactive configuration of the WSN on-site. Besides, an intuitive interface is developed for the user to apply virtual loads directly on the physical structure. These achievements contribute toward a novel method to investigate the structural behavior under different loading conditions.

22.6 Interactive Model Modification

In standard FEA software, the user needs to conduct lengthy operations to modify the model in the pre-processing stage and repeat the standardized steps to re-analyze the modified model as shown in Fig. 22.25a. This traditional approach is inefficient for performing and investigating modifications. In the current AR-based development, customized interfaces can be established to allow intuitive and efficient modifications. The FEA results can be updated automatically for modifying the model and conducting re-analysis (Fig. 22.25b).

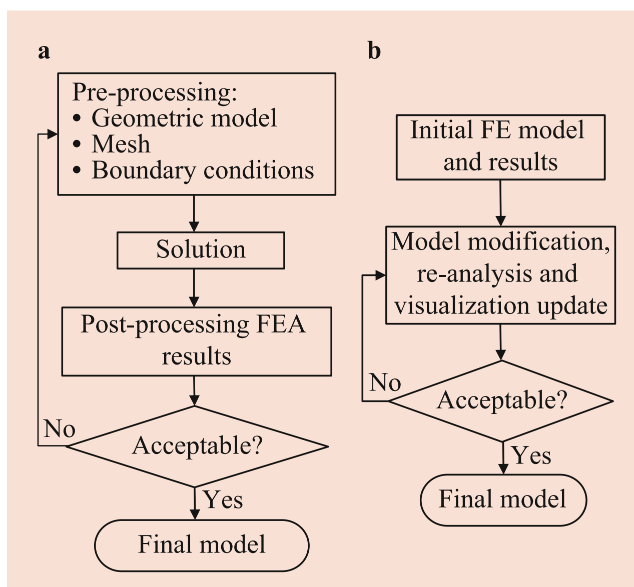


Fig. 22.25 Model modification in standard FEA software and the AR-based system: (**a**) FEA software and (**b**) AR-based system

22

Modifications can therefore be conducted interactively with direct FEA result update. This interactive modification approach enables the user to focus on investigating the effects of the modification on the analysis results, rather than being distracted by the system operations. However, the interaction approach and interface design depend largely on the operations needed and FEA tools available for the specific task. Two specific types of model modifications are presented in this section, i.e., adding geometric models and local mesh refinement.

22.6.1 Adding Geometric Models

The method enables the user to add new geometric entities to the current FE model, such as adding structural members to stiffen a structure. This method is described as follows.

The user can manipulate a new geometric model with an initial FE model using the 3D input device. As the geometric model can intersect the FE model, it is necessary to make sure the models can be joined (Fig. 22.26a). To join the new geometric model, a common method is to merge this with the original geometric model and regenerate the global mesh, which is time-consuming and inefficient. A more efficient way is to trim the new geometric model and mesh individually (Fig. 22.26b and c). The trimming is achieved by dividing the geometric model with the intersected faces of the FE model and then removing the unwanted portions which are determined according to the normal directions of the intersected faces. The dissimilar meshes need to be connected, by joining the newly generated mesh with the original mesh. Smooth connection requires mesh adaptation to generate common nodes shared by the contacting meshes. For simplifying the connection issue, the current research uses a method of connecting dissimilar meshes, which is to impose linear constraint equations to the nodes located in the contacting areas.

Figure 22.26d shows an example, in which the connection at node E is achieved by applying constraints to node E and the four surrounding nodes using Eq. (22.8). u_A, u_B, u_C, u_D, and u_E represent the displacements of nodes A, B, C, D, and E, respectively. w_A, w_B, w_C, and w_D are the weights which are equal to the values of the shape functions that are evaluated at the location of E inside the element face $ABCD$.

$$u_E = w_A u_A + w_B u_B + w_C u_C + w_D u_D \tag{22.8}$$

Multiple geometric models can be added by repeating these operations. The FE model is next re-analyzed keeping the same boundary conditions. The visualization of mesh and solutions is updated accordingly. To implement the modification, the user is only required to position the geometric models and specify the intersected faces of the FE model. The subsequent tasks can be performed automatically for updating the model and generating the results.

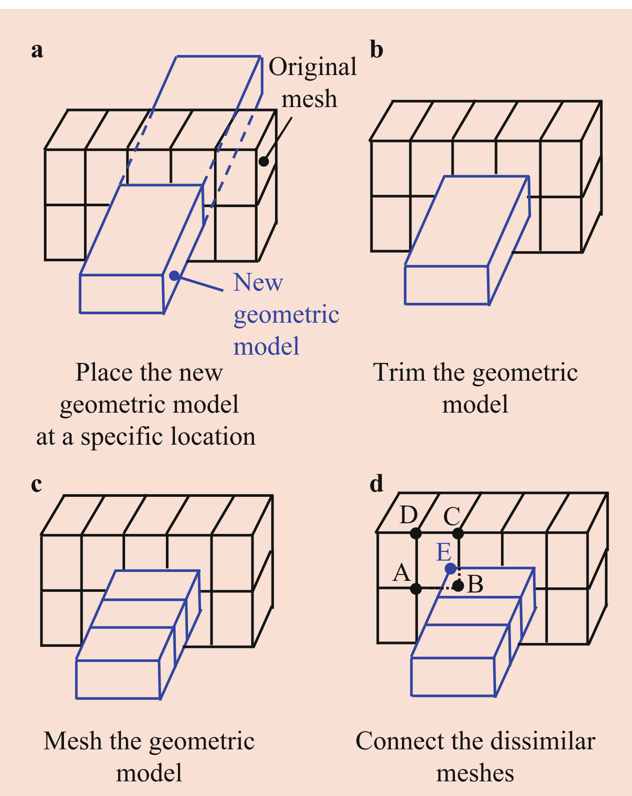


Fig. 22.26 The process of adding geometry model: (**a**) place the new geometric model at a specific location, (**b**) trim the geometric model, (**c**) mesh the geometric model, and (**d**) connect the dissimilar meshes

22.6.2 Local Mesh Refinement

When the local mesh around a point load has been changed due to mesh refinement, the loaded nodes and nodal weights of this point load should be changed accordingly for the computation of nodal loads. To re-apply a point load on a modified mesh, a ray is created which originates from the load location and has the same direction as the load. After the intersection of the ray and the surface mesh is obtained, the loaded nodes and nodal weights are determined accordingly to compute the equivalent nodal forces.

After the model modifications, such as adding new geometric models and local mesh refinement, the inverse stiffness matrix can be updated automatically if real-time simulations are required. In addition, the interactive modification methods require automatic processes to modify the geometric model or mesh, such as trimming the new geometric model, mesh generation, mesh refinement, etc. The feasibility of automation depends on the geometric modeling and meshing tools available.

22.6.3 Summary

The developed AR environment promotes interactive model modification processes in FEA. This section presents the interactive methods for adding geometric models and local mesh refinement. These methods are achieved using customized AR interfaces and automatic FEA procedures. The AR interfaces support intuitive and efficient control of the variables for model modification, while automatic FEA procedures enable direct update of FE models and results. These methods require automatic tools for certain processes, such as trimming of geometric models, mesh generation, mesh refinement, etc.

22.7 System Implementation and Case Studies

C++ in Microsoft Visual Studio 2010 was used to develop the prototype. A few open-source libraries are employed, namely, OpenGL for rendering and selection of virtual objects, VTK for visualization of FEA results, and ARToolKit for tracking. VTK is integrated with ARToolKit to allow implementation of scientific visualization algorithms in the AR-based environment as elaborated in Sec. 22.4. A standard webcam with a capture rate of 30 fps was used. A virtual panel was created with a customized menu, and this can be activated using a 3D input device. The commercial FEA software ANSYS is selected to perform some FEA tasks. ANSYS was selected as it has excellent solver capabilities and provides a powerful scripting language, i.e., ANSYS Parametric Design Language (APDL), that enables users to parameterize the models and automate the processes for FEA and data output. Figure 22.27 illustrates the bidirectional communication between ANSYS and the AR-based system. The AR-based system controls the generation of APDL codes written in text files and executed by ANSYS in the batch mode. ANSYS performs FEA tasks and outputs data accordingly. Using this approach, data exchange between ANSYS and the AR-based system can be achieved easily.

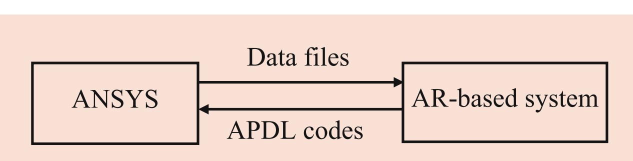


Fig. 22.27 Communication between ANSYS and the AR-based system

22.7.1 Case Study of a Step Ladder

A case study has been conducted on an off-the-shelf step ladder (Fig. 22.28), using a PC with a 3.2 GHz Intel processor and 8 GB RAM. It is noted that the intention is not to conduct a rigorous structural analysis but to validate the performance of the AR system.

Model Preparation

Figure 22.28 shows the step ladder used for the case study. A CAD model of the ladder is built using SolidWorks and imported into the ANSYS software. Non-essential features are omitted to simplify the model. ANSYS performs the tasks of mesh generation and matrix computation. The wooden

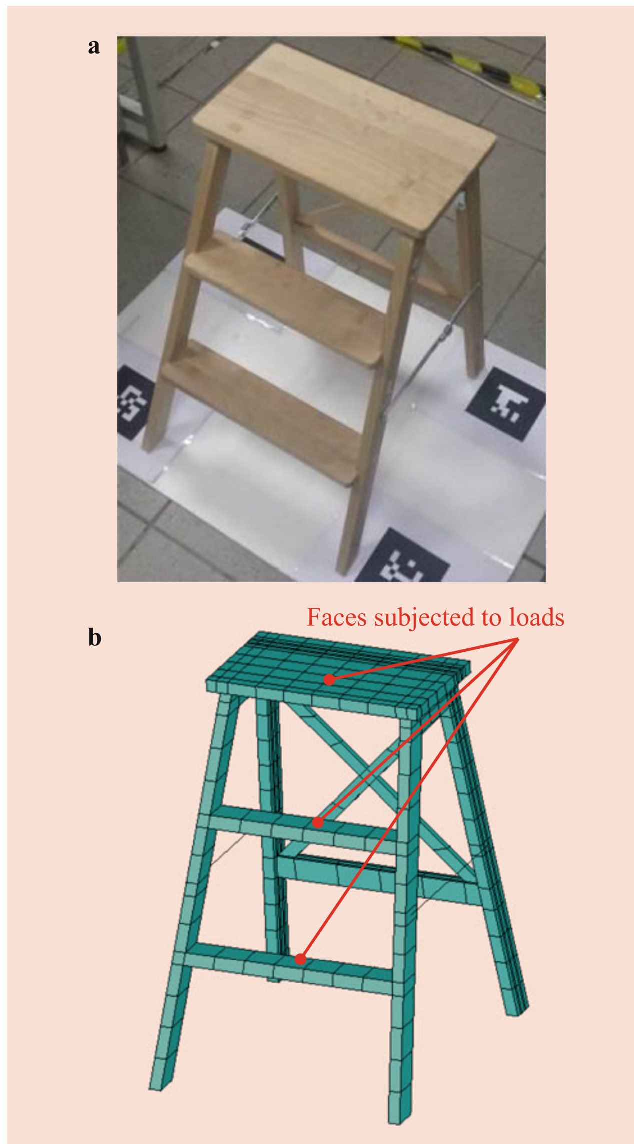


Fig. 22.28 The step ladder and finite element model: (**a**) step ladder and (**b**) finite element model

parts of the step ladder are modeled using linear eight-node hexahedral elements, and the two metal linkages are modeled using truss elements which are connected to the hexahedral elements using linear constraints. The material properties are assumed to be elastic isotropic. After the linear system has been established, the stiffness matrix is extracted from the ANSYS database, and the inverse matrix is computed using an external PCG solver. Only three of the faces shown in Fig. 22.28b can be loaded by the user; therefore, the inverse matrix is only computed for the nodes on these faces to reduce computation time. The surface mesh data is extracted using the APDL commands as shown in Table 22.1. All these data are input into the system during the initialization stage.

Data Visualization and Exploration

Visualization and Manipulation

By integrating VTK, FEA results can be superimposed on the real step ladder. Several visualization styles have been examined. As shown in Fig. 22.29, exaggerated displacements embody the shape of the deformation, but would lead to the misalignment of stress distributions in the real structure. Direct overlaying of the deformed model on the real structure provides clear views of the results (Fig. 22.29a), but may misinterpret the deformation without dealing with occlusion issues. However, when occlusion is addressed, part of the results becomes invisible (Fig. 22.29b). A user can see both the FEA results and the structure by making objects translucent (Fig. 22.29c).

User viewpoint can be changed to observe FEA results from different perspectives. The user can also move the input device to any position and attach the model with it (Fig. 22.30a) to translate, rotate, and zoom the model (Fig. 22.30b). This allows the user to access the bottom views of the structure and zoom into specific regions easily. It is better to position the input device at a location close to the model for attaching the model to the device, because it would be awkward to manipulate a model that is far away from the input device. The user can place the model at a appropriate location if some regions away from the marker cube cannot be observed (Fig. 22.30c).

Table 22.1 APDL commands for extracting surface mesh

Specify the faces of the model to extract the mesh data
For each face {
Select this face (*ASEL*)
Select all the nodes associated with the selected face (*NSLA*)
Select all the elements associated with the selected nodes (*ESLN*)
For each element {
Get the face number of the element that contains the select nodes (*NMFACE*)
Get node numbers of the element face (*NELEM*)
Write the node information in the data file (**VWRITE*)
}}

Slicing and Clipping FEA Results

The volumetric FEA results can be explored through data slicing and clipping. Two operating modes, hand-operated and view-based, have been implemented. Figure 22.31 illustrates the hand-operated mode. In this mode, the stresses of the step ladder can be examined using a handheld slicing plane or clipping cube (Fig. 22.31a and d). The data slices or clips are updated as the user moves the 3D input device. Translucent visualization is adopted for the global FEA re-

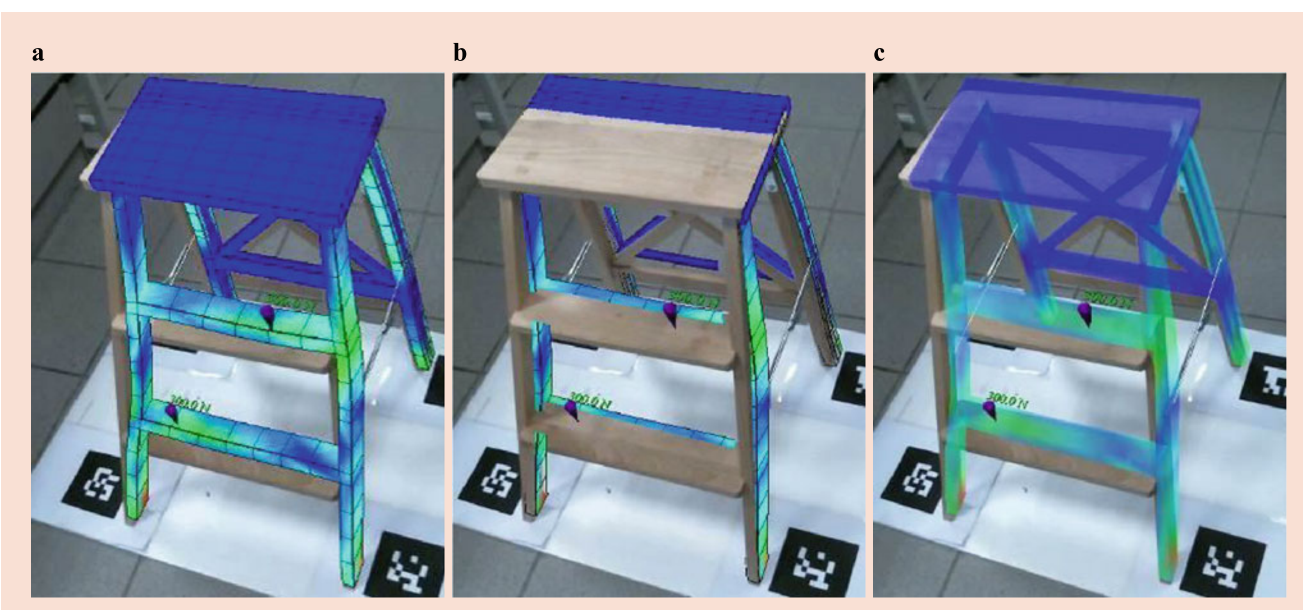


Fig. 22.29 Different visualization styles: (**a**) direct overlay, (**b**) overlay with occlusion handling, and (**c**) translucent visualization

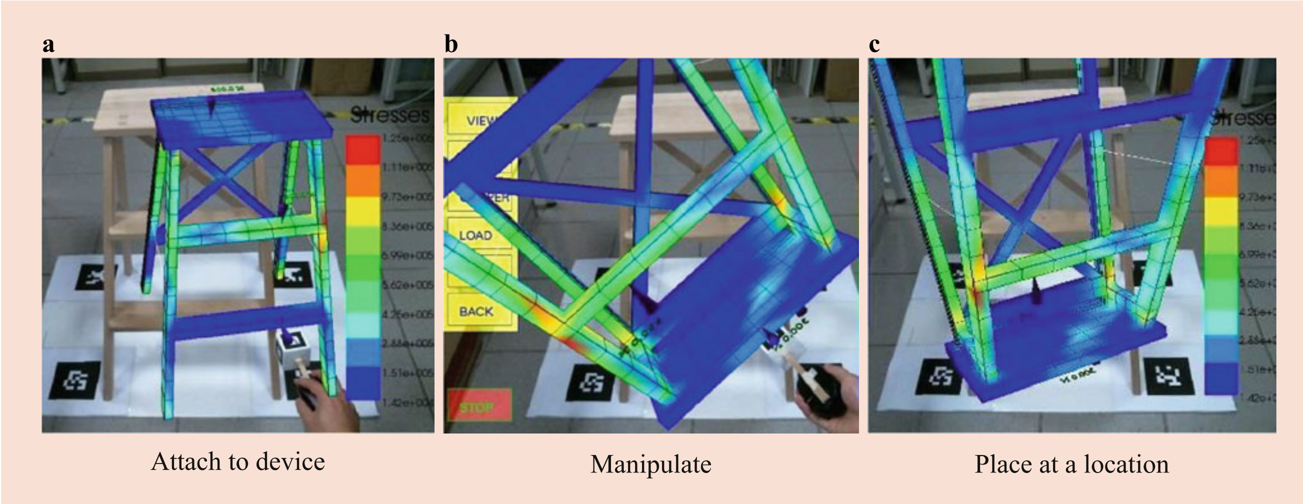


Fig. 22.30 Direct manipulation of FEA results: (**a**) attach to device, (**b**) manipulate, and (**c**) place at a location

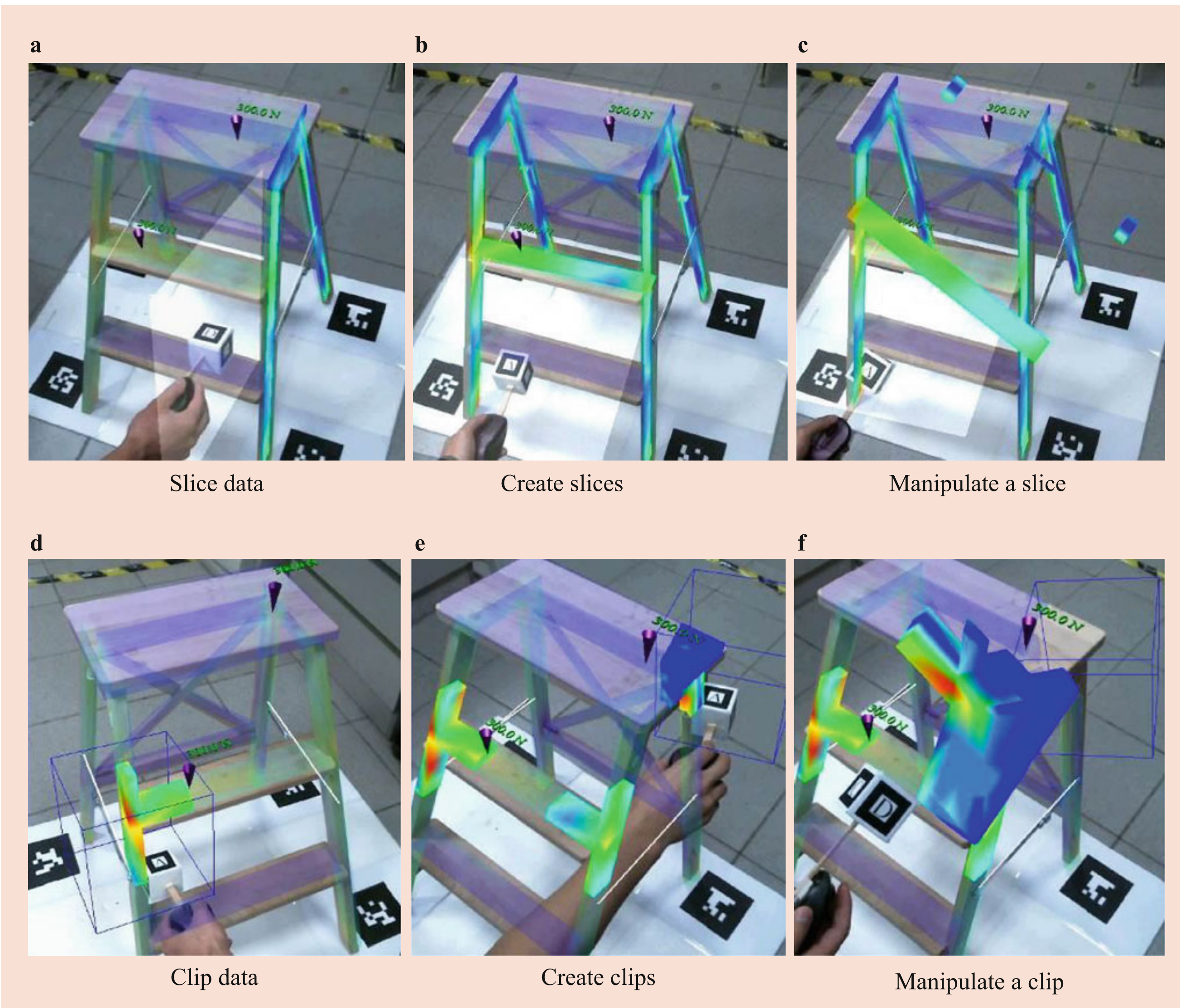


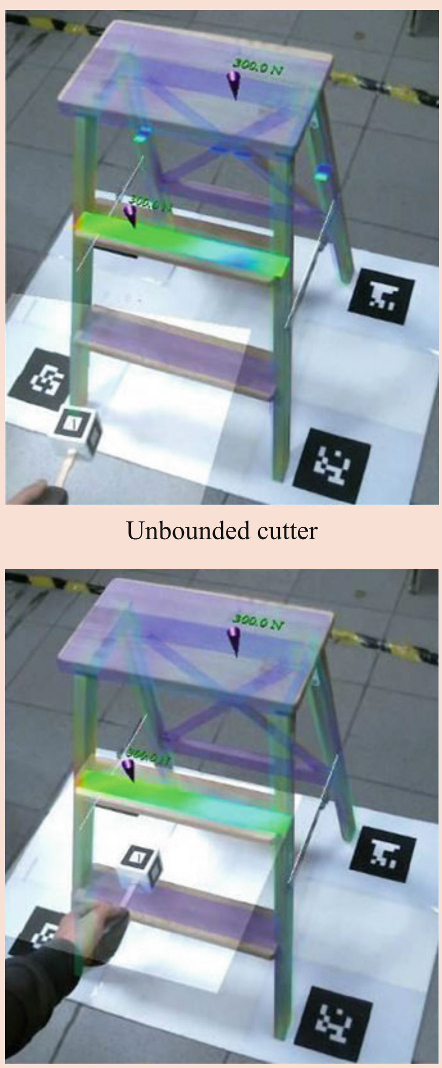
Fig. 22.31 Slicing and clipping FEA results in the hand-operated mode: (**a**) slice data, (**b**) create slices, (**c**) manipulate a slice, (**d**) clip data, (**e**) create clips, and (**f**) manipulate a clip

sults, such that the data slices or clips are visible. Using the buttons on the input device, the user can create data slices or clips at different locations on the structure for observation and comparison (Figs. 22.31b and e) or attach one slice or clip to the input device for manipulation (Fig. 22.31c and f).

For data slicing, an unbounded slicing plane allows the user to operate when the model is out of reach, but it may generate redundant slices if there are multiple intersection areas. This can be overcome with a bounded plane of adjustable size, but the model must be reachable. A comparison of bounded and unbounded slicing planes is shown in Fig. 22.32.

In the view-based mode, an unbounded slicing or clipping plane is placed in parallel with the view plane at an adjustable distance. The user can manipulate the slicing or clipping plane by moving the viewpoint or rolling the mouse wheel to adjust the position of the slicing or clipping plane relative to the view plane (Fig. 22.33a and b). Using a mouse button, the slicing or clipping plane becomes static on the structure, and data can be observed from different perspectives (Fig. 22.33c). However, the user cannot have a direct visual sense of the orientation of the slicing or clipping plane, since the slicing or clipping plane is always parallel to the view plane in this operating mode. Moreover, it would not be easy to locate the slicing or clipping plane when there is no intersection between the plane and the model for referencing.

a



Unbounded cutter

b

Bounded cutter

Fig. 22.32 Slicing with unbounded and bounded cutters: (**a**) unbounded cutter and (**b**) bounded cutter

Real-Time Simulation

Virtual loads can be applied using the 3D input device, or sensors can be installed to measure the loads in the actual loading environment. During real-time simulation, the FEA results can be updated at a fast speed in response to the load variations.

Real-Time Simulation with WSN

As shown in Fig. 22.34, the loading regions of the ladder are created by the user stepping on it, and four wireless force sensors are used for measuring the load input. After the preparation steps described in Subsection 22.7.1, the system is operated as follows to perform real-time simulation. The FE mesh data and inverse matrix are sent to the system when the "READ" button is triggered (Fig. 22.35a). Sensor nodes can be added or removed (Fig. 22.35b). To add a sensor after powering up, the user can observe a text message "New sensor node detected!". The sensor coordinate system is specified by casting a ray onto where the sensor is attached (Fig. 22.35c). The normal of the surface, which coincides with the sensor axis, is computed to establish the coordinate transformation. Point loads are applied to the nodes of the loaded elements. Successful registration will be reflected by a LED bulb. Real-time simulation is launched after using the "RUN" button (Fig. 22.35d).

Figure 22.36 shows the variations of stresses when a user steps on the ladder. In Fig. 22.36a, the color scale is updated with the maximum and minimum stress values for each frame. Color scale shown in Fig. 22.36b facilitates the comparison of different frames.

A test was conducted by inviting three subjects of different weights to step on the ladder. From Fig. 22.37, it can be observed that heavier subjects created higher stresses. During simulation, slicing or clipping is allowed to explore the results, which will be updated in every frame. Figure 22.38 shows data slicing during the real-time simulation.

Finer mesh can produce more accurate results but requires longer computation time; this is illustrated using four different mesh resolutions as shown in Fig. 22.39. Models A, B, and C have fewer nodes and are used for real-time

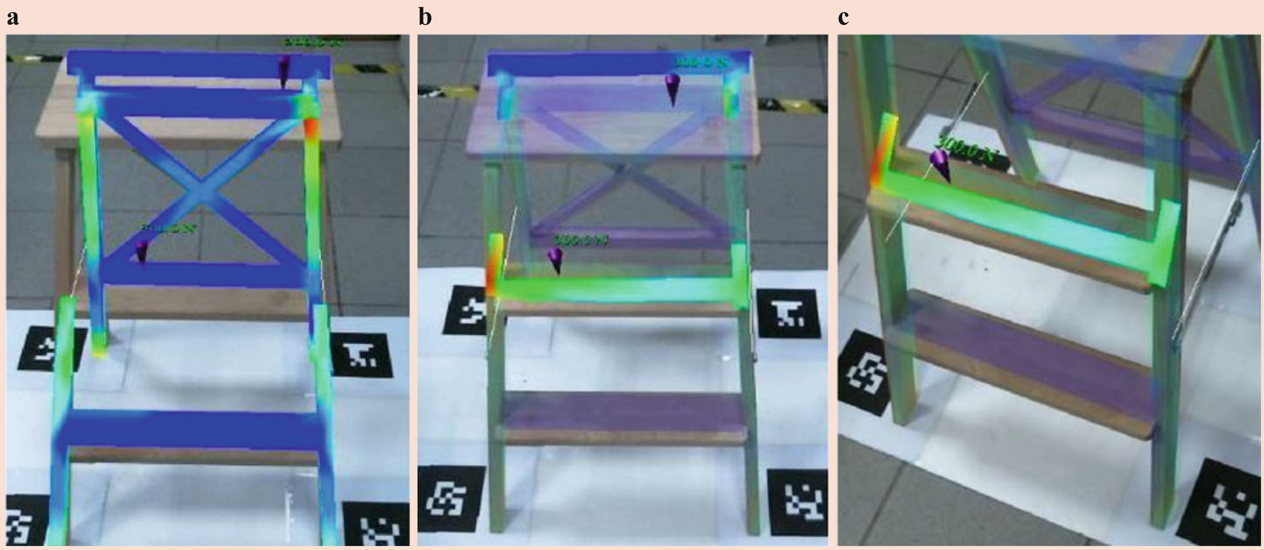


Fig. 22.33 Slicing and clipping FEA results in the view-based mode: (**a**) clip FEA results, (**b**) slice FEA results, and (**c**) static slicing plane

Fig. 22.34 Installation of wireless force sensors

simulations. Model D has the highest number of nodes and is used as a reference for static stress analysis. Figure 22.40 shows stress comparison under the same loading conditions. As can be seen, the stress distribution is more detailed when the mesh resolution increases, but this affects the frame rate as it drops drastically toward zero (Fig. 22.41).

Loading values are refreshed at 48 Hz using established WSN. Figure 22.42 shows the computation time for the other processes.

Real-Time Simulation with Virtual Loads

Figure 22.43 shows a scenario in which virtual point loads are applied to the step ladder. Stresses and deformations can be visualized on the step ladder when a subject moves up (Fig. 22.43a). More loads can be applied after finalizing the load location (Fig. 22.43b). The user can manipulate the load by adjusting both value and orientation (Fig. 22.43c) or remove it. Similarly, pressure can be applied. Figure 22.43d shows a real-time simulation of pressure exertion on selected areas.

In this scenario, the model has 944 nodes. This model requires a pre-computation time of around 88 s. Frame rate decreases to 12 fps when there are ten dynamically changing loads; it also decreases when the mesh becomes finer. The performance can be improved by optimizing the codes, especially for the matrix-vector multiplication.

Discussion

Real-time performance depends largely on the online solution time and hardware performance. Even with pre-computed inverse stiffness matrix, the solution time can still be significant when many nodes are used. Table 22.2 shows the results achieved in the scenario where WSN is used and the results achieved in the literature [27]. Although different hardware affects computation time, the benefits of pre-computation are significant. Previous tests using matrix-

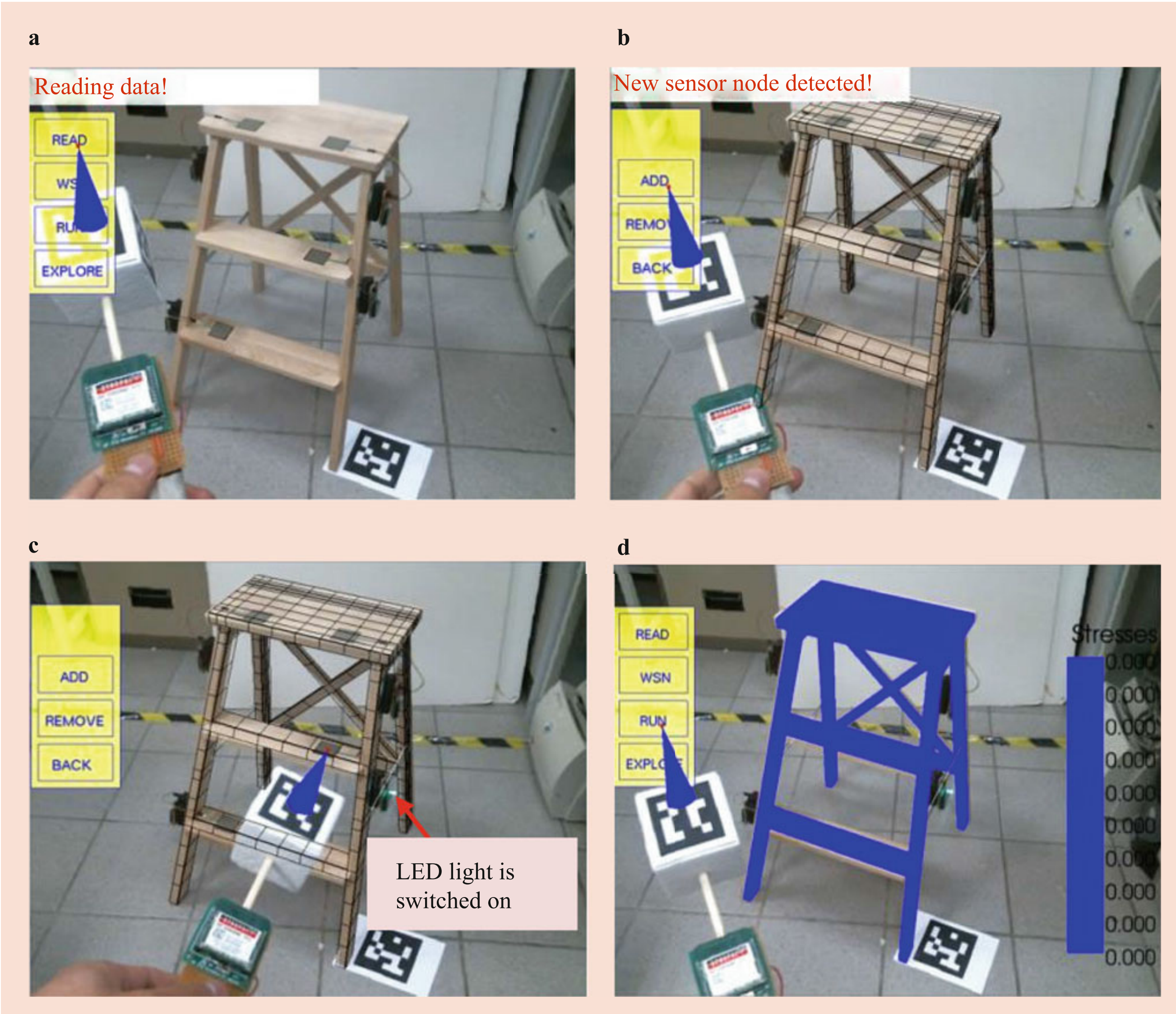


Fig. 22.35 System operation for real-time simulation: (**a**) read data, (**b**) add sensor nodes, (**c**) specify sensor coordinate systems, and (**d**) launch real-time simulation

vector multiplication can be at least ten times faster than other solution methods, such as conjugate gradient, Gaussian elimination, and several factorization techniques [47].

Table 22.2 shows the pre-computation time when the PCG solver has a tolerance of 1.0E-5. Once the inverse stiffness matrix has been computed, it can be stored for reuse. If the stiffness matrix changes, the inverse matrix should be updated. This happens when FEA parameters such as the geometry and constraints have been changed. The Sherman-Morrison-Woodbury formula can be employed to update the inverse matrix quickly under certain conditions [72].

Load input is assisted by using sensors. Once the force sensors are registered at certain locations, loads can be detected at these locations only. In contrast, virtual loads can be applied at any locations on the structure, but a collision computation is needed. The precision of applying virtual loads is determined by the precision of hand manipulation and vision-based tracking. Moreover, the loads measured using force sensors in this case study are simplified to point loads. The case study demonstrates the treatment of force sensors, but the approach can be adapted to the practical situations where loads are indirectly measured using other types of sensors, such as strain gauges and displacement sensors.

Re-Analysis with Model Modification

Structural Stiffening

To implement the method of adding geometric models, an interface has been developed for adding beams to stiffen

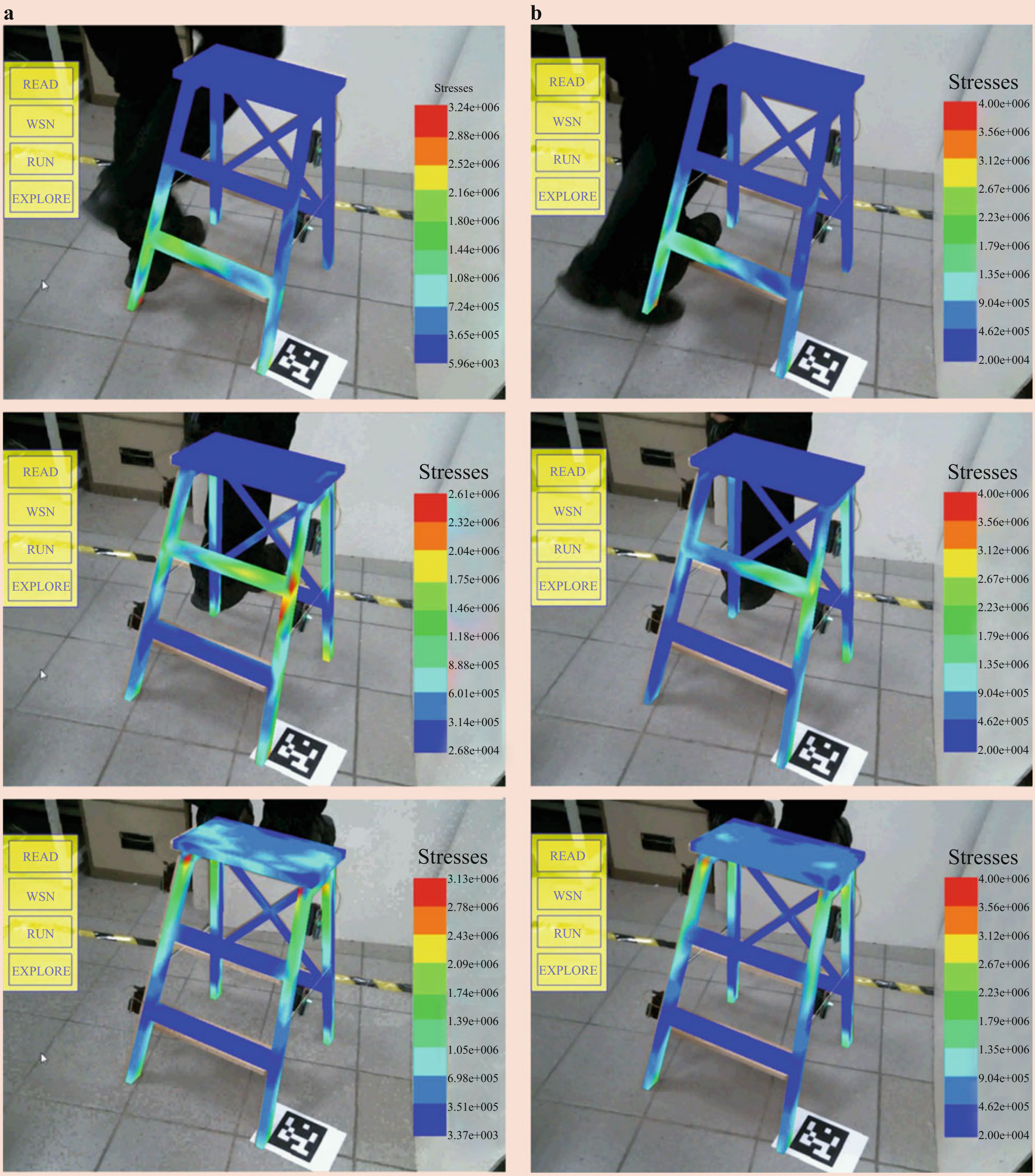


Fig. 22.36 Visualization of stresses during stepping: (**a**) with a dynamic color scale and (**b**) with a consistent color scale preset

the structure. The original model and results are shown in Fig. 22.44a and after manipulation and placing the beam at a specific location (Fig. 22.44b). The faces of the step ladder where the beam intersects need to be specified (Fig. 22.44c). Using the "CONFIRM" button, all the data will be transferred to ANSYS via APDL codes. The mesh connection is achieved using the CEINTF command in ANSYS (ANSYS, 2013). The new mesh and results are rendered after the completion of those tasks (Fig. 22.44d). Real-time simulation can be achieved after the computation is completed (Fig. 22.44e).

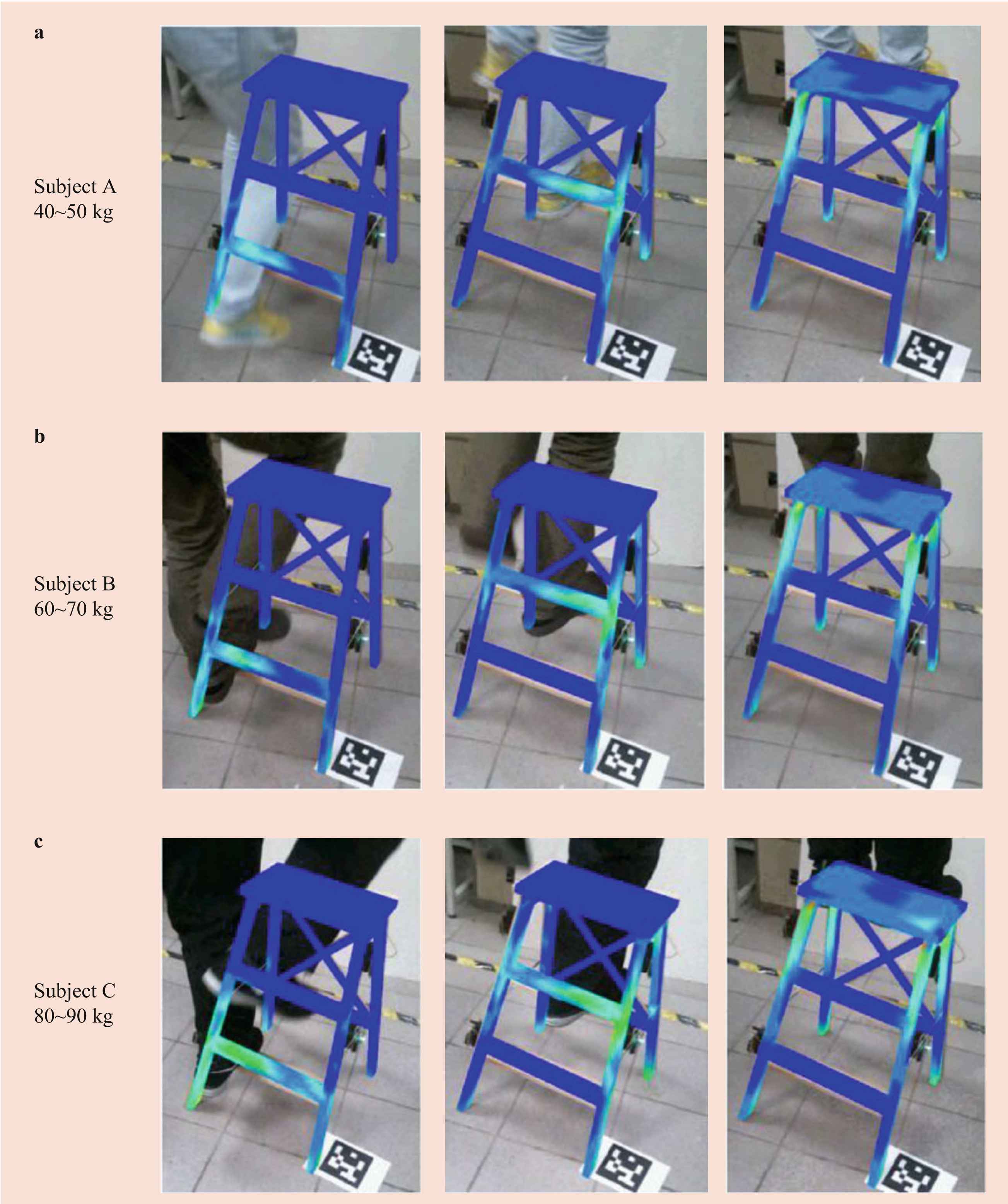


Fig. 22.37 Stress visualization for subjects of different weights: (**a**) Subject A weighing 40 ~50 kg, (**b**) Subject B weighing 60 ~70 kg, and (**c**) Subject C weighing 80 ~90 kg

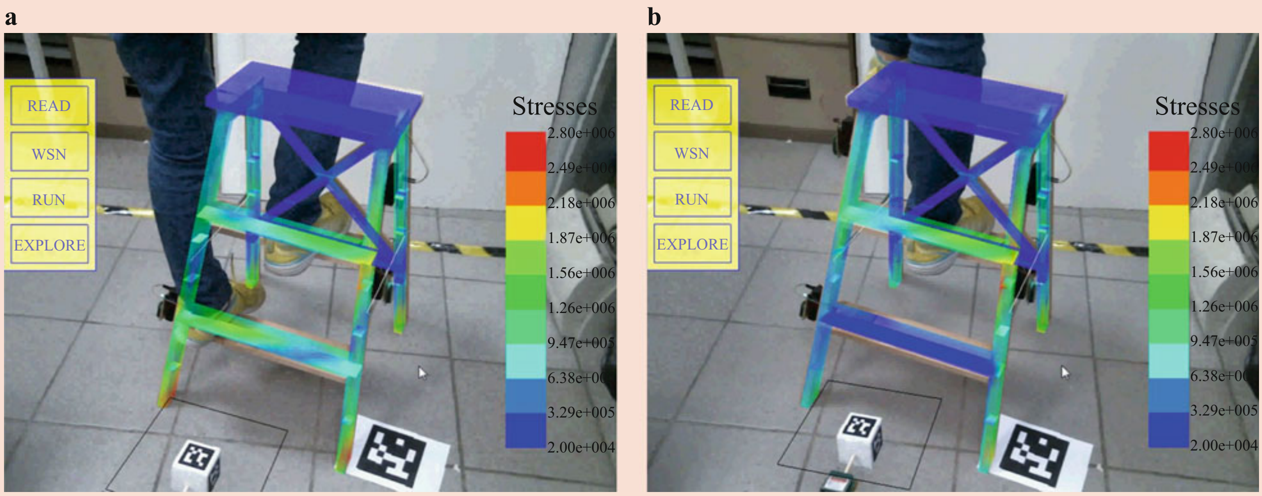


Fig. 22.38 Slicing real-time results: (**a**) slice FEA results and (**b**) sliced results updated under varying loading conditions

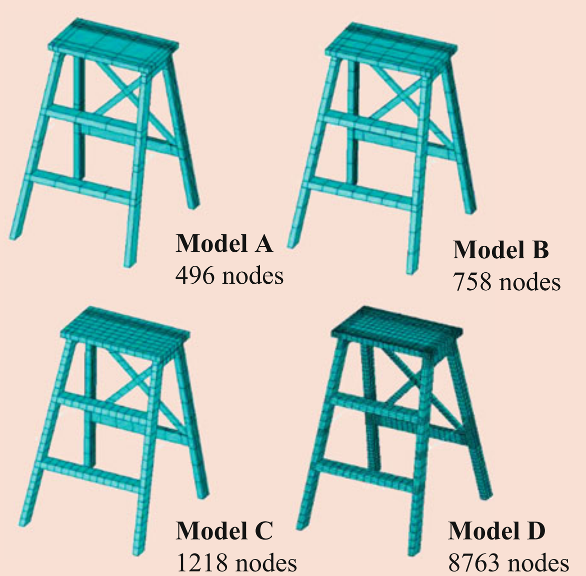


Fig. 22.39 Meshes of different resolution: (**a**) Model A of 496 nodes, (**b**) Model B of 758 nodes, (**c**) Model C of 1218 nodes, and (**d**) Model D of 8763 nodes

The FE model displays an improved structure after adding the beams, as the deformation and stresses on the step have been reduced (Fig. 22.44a and d). Discontinuous stress fields in the connecting areas have been observed as a result of connecting dissimilar meshes (Fig. 22.44f).

Local Mesh Refinement

Mesh refinement is implemented using a tetrahedral mesh of 1022 nodes. Figure 22.45a shows the original model; the user then selects the elements in the region (Fig. 22.45b). Figure 22.45c shows the refinement level set by the user. The equivalent nodal forces of the point loads are recalculated after mesh modification. The updated FE model is then rendered on the step ladder (Fig. 22.45d).

System Response Time

In model modification, the system response time is crucial for the user interactions, i.e., the time lag between the user input and the FEA result update. Table 22.3 shows the variation of system response time with different mesh resolutions.

22.7.2 Case Study of a Proving Ring

A second case study has been implemented to demonstrate the applicability of the real-time simulation method to other types of structures. As shown in Fig. 22.46, a proving ring that has a digital indicator is studied. The loads acting on the proving ring are measured using the dial indicator. As the measured values are encoded in the digital output signals of the indicator, a microprocessor is used to interpret the output signals according to the communication protocol. The system simulates the stresses on the proving ring when it is loaded on a testing machine (Fig. 22.46c and d). The model has 2394 nodes, and the frame rate is 19 fps on a laptop with a 2.3 GHz processor and 4 GB RAM.

22.7.3 A Prototype Application for Education

Due to the versatility of FEA in solving complex engineering and science problems, FEA courses can be found in almost all

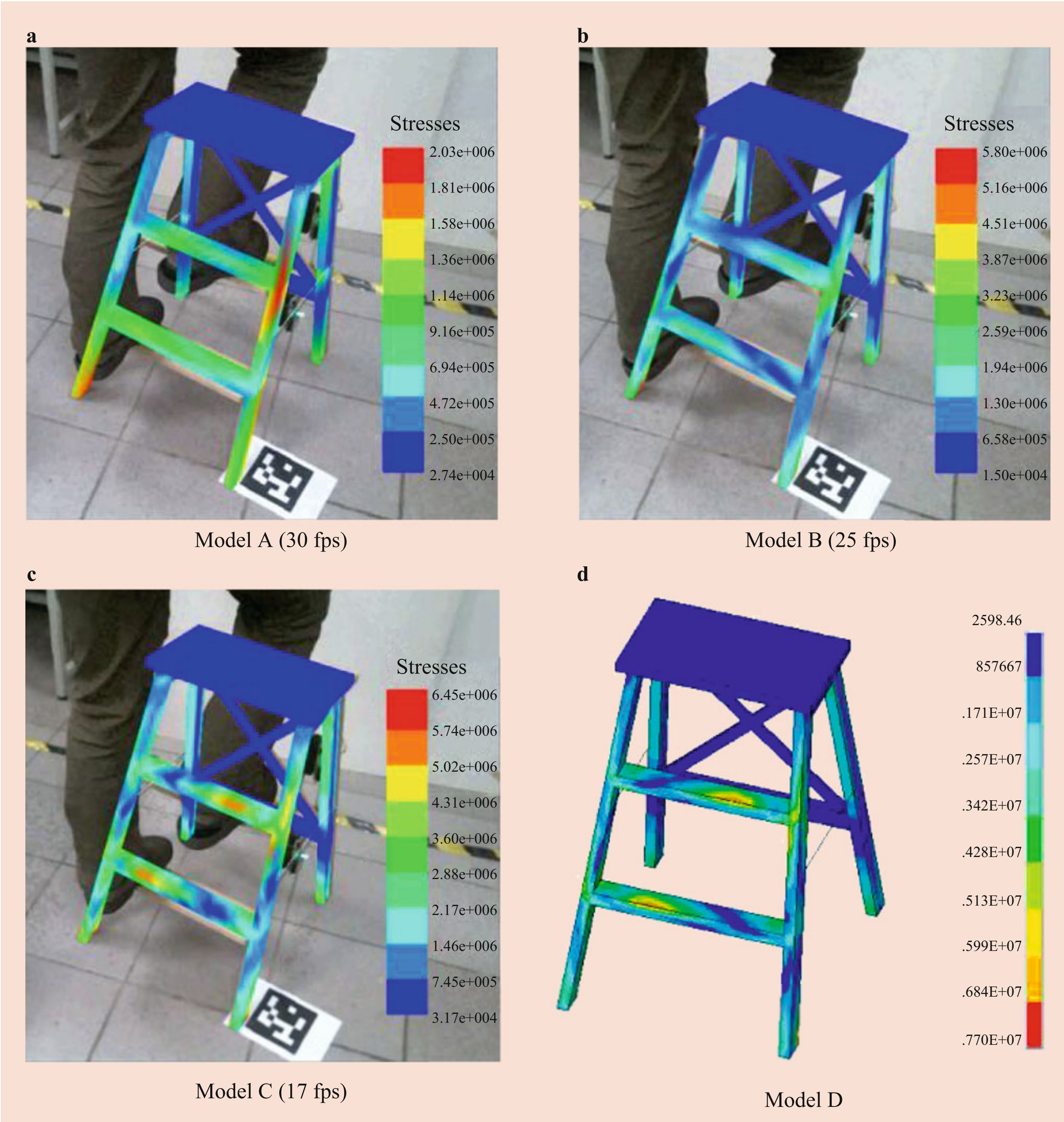


Fig. 22.40 Comparison of results for meshes of different resolution: (**a**) Model A, (**b**) Model B, (**c**) Model C, and (**d**) Model D

the engineering faculties. In traditional learning of FEA, the students need to go through rather abstract and tedious theories, and the commercial FEA software often has counter-intuitive user interfaces. The virtual environment of the FEA software does not facilitate easy understanding. In order to enhance the learning of FEA, a prototype application system has been developed as shown in Fig. 22.47. A truss structure, which is a common model used for the teaching of structural mechanics, is studied. The FE display can be observed from all angles by moving the camera around the structure. Loads are applied by pressing the Flexiforce pressure sensors, and the measured forces are sent to an Arduino microcontroller connected to the PC.

Using a hexahedral mesh, 776 nodes are generated. Constraints are applied to the faces in contact with the ground. The inverse stiffness matrix is computed a priori for all the

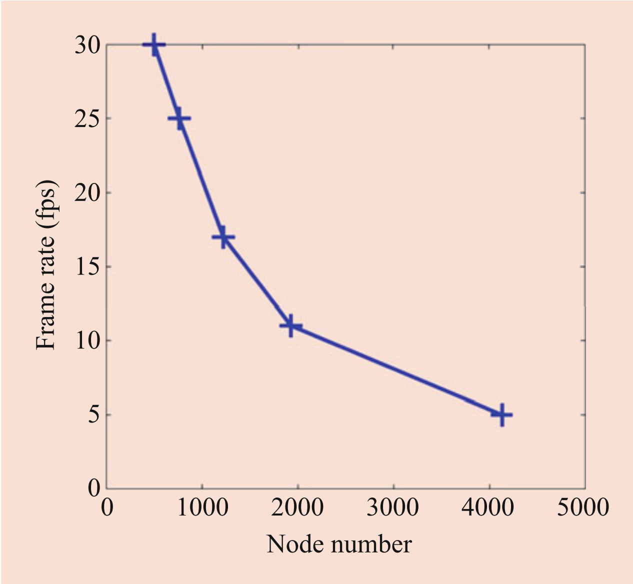


Fig. 22.41 Relationship between frame rate and node number

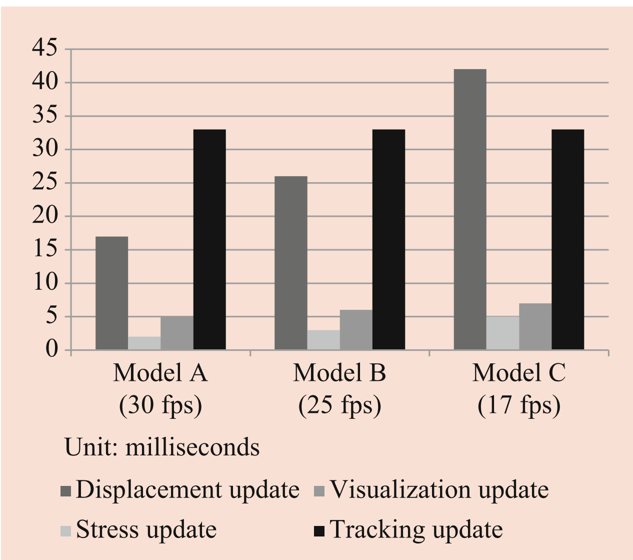


Fig. 22.42 Computation time for the relevant processes

loadable nodes. The time taken was about 4 min and 20 s using a PCG solver with a tolerance of 1.0E-5. The mesh data and inverse stiffness matrix are stored in text files and input into the system.

A user study was conducted to evaluate the advantages and effectiveness of the pedagogy of the proposed system and obtain feedbacks on the prototype system. Twenty students from mechanical and civil engineering were invited to participate. They were divided into two groups: beginner and expert. The beginner group consists of eight participants with basic knowledge of FEA but have not used FEA software. The expert group consists of 12 participants with extensive experience in structural analysis using FEA software. The user study has been designed with three sessions, namely, training, practicing, and post-test. A training session was first given to all the participants, and they were then divided into smaller teams of two to three students with detailed instruction of the tasks to be undertaken. After the training session, each participant was required to perform three types of tasks as illustrated in Fig. 22.48, including load application, result exploration, and structural stiffening.

In the post-test session, each participant was asked to complete a questionnaire which includes a five-point Likert scale to evaluate the system qualitatively in learning, intuitiveness, interaction, and satisfaction (5 = most positive, 1 = most negative) and additional questions to gather participant's comments and suggestions on the interfaces, functions, and comparison between standard FEA software and the developed AR-based system. To evaluate the effect on learning, a matching question is designed based on the task of load application in the practicing session. The question has two columns of captions with different loading conditions and FEA results. The students were required to find the correct deformation and stresses for the loading conditions within 15 min.

All participants were impressed by the capability of performing FEA tasks directly on a real structure. Figure 22.49 shows the results of the Likert-based survey. The average score for "ease of learning" is 4.2, which is the highest among all the five categories. The result has strongly indicated that this system is "beginner-friendly." The expert group gives higher ratings on intuitiveness, ease of use, and learning. Eight participants who have had FEA experience all agreed that the developed system is more intuitive and easier to use than commercial FEA software. Due to the AR visualization benefits, cognitive loads of the FEA operating environment are reduced, and the students can focus more on the structural investigation.

For the matching question, three quarters of the students had answered at least seven out of nine questions correctly as shown in Fig. 22.50. Although more students from the expert group had answered all the questions correctly, there is little difference between the two groups, indicating that the AR-based learning environment is very valuable for the beginners.

22.7.4 Summary

A prototype system has been built based on the proposed methods. Case studies have been carried out to validate the concept and evaluate the performance. The results have demonstrated the feasibility and effectiveness of the proposed methods for enhanced structural analysis. Some issues of visualization and interaction have been discussed, such as

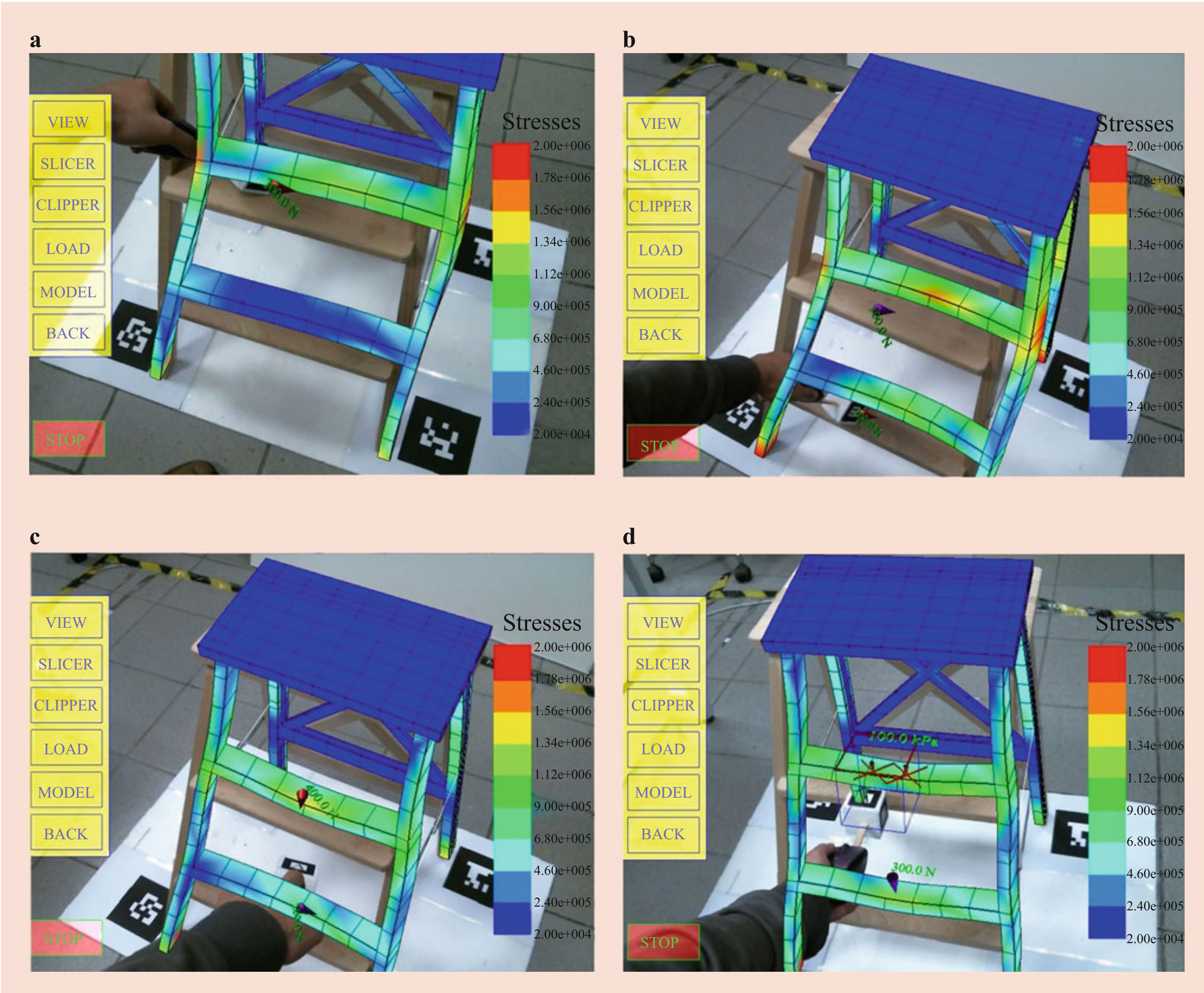


Fig. 22.43 Application of virtual loads to a step ladder: (**a**) apply a load, (**b**) apply more loads, (**c**) adjust a load, and (**d**) apply pressure

Table 22.2 Comparison of the results achieved in the two studies

Model	Element type	Number of nodes	Solution time (ms)	FEA result update rate (Hz)	Pre-computation time (s)
The step ladder	Solid hexahedron	1218 758 496	42 26 17	17 25 30	149 58 41
A cantilever (Uva et al.)	Shell	926 314 132	822 323 153	1.2 3.1 6.5	

visualization styles for result overlay and operating modes for data exploration. Besides, the response time performance of the AR-based system has been evaluated. Based on the AR-based system, an educational prototype application has been developed. The results of the user study illuminate the advantages of the application for enhanced learning of FEA. The results also reveal the limitations of the system in precision and stability of hand operation, tracking accuracy, etc.

22.8 Conclusions and Future Readings

The main objective of this research is to explore the feasibility of applying AR technology to enhance FEA operations. It has been achieved using an FEA-AR integrated platform which has been designed and implemented. The platform can provide visualization of FEA results in real-world environments and supports intuitive interaction methods for structure analysis. The research outcome demonstrates the significance of AR for enhanced visualization and interaction of FEA.

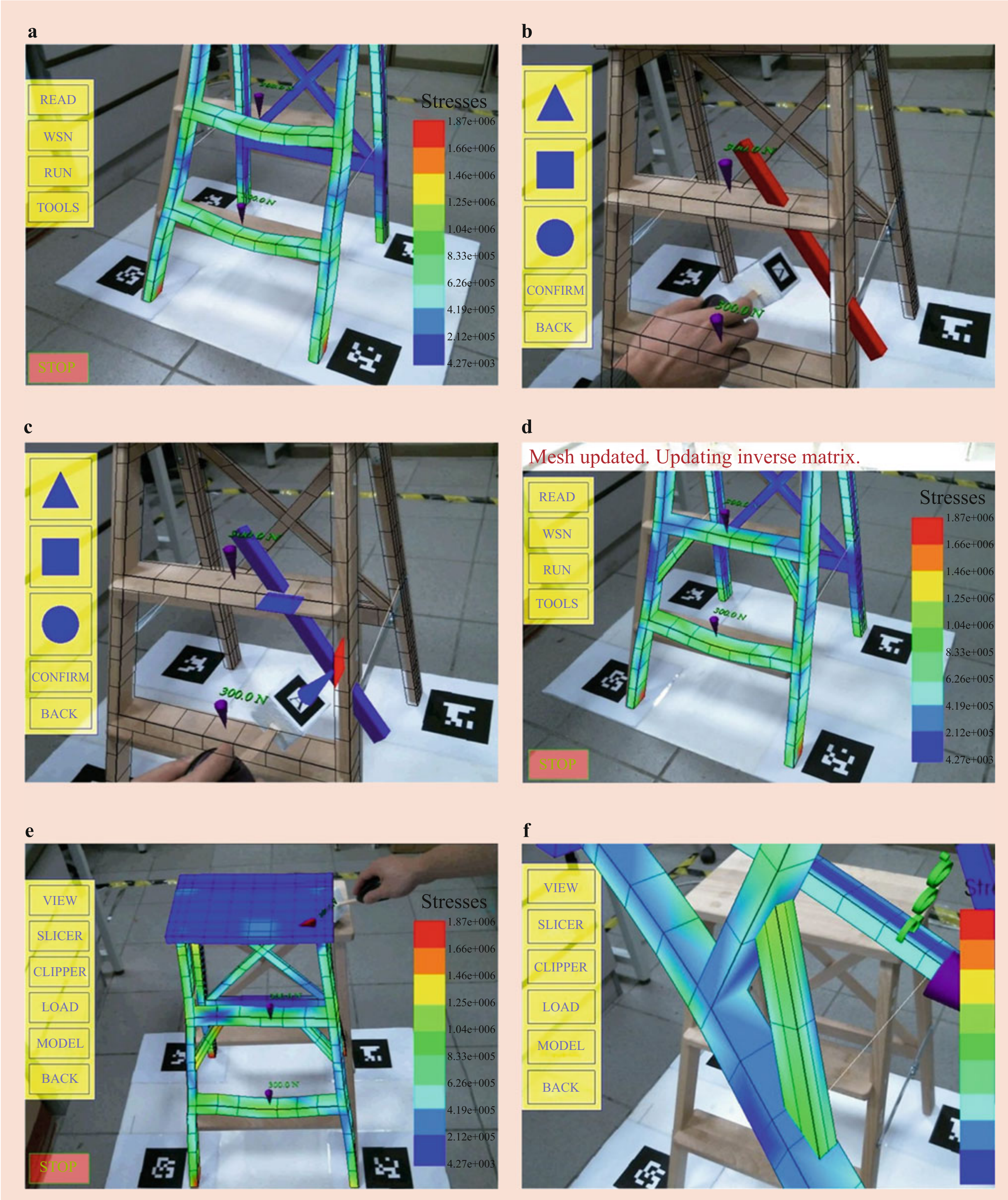


Fig. 22.44 Adding beams to stiffen the structure: (**a**) initial model, (**b**) place the beam, (**c**) specify intersected faces, (**d**) updated model, (**e**) apply virtual loads, and (**f**) discontinuous stress

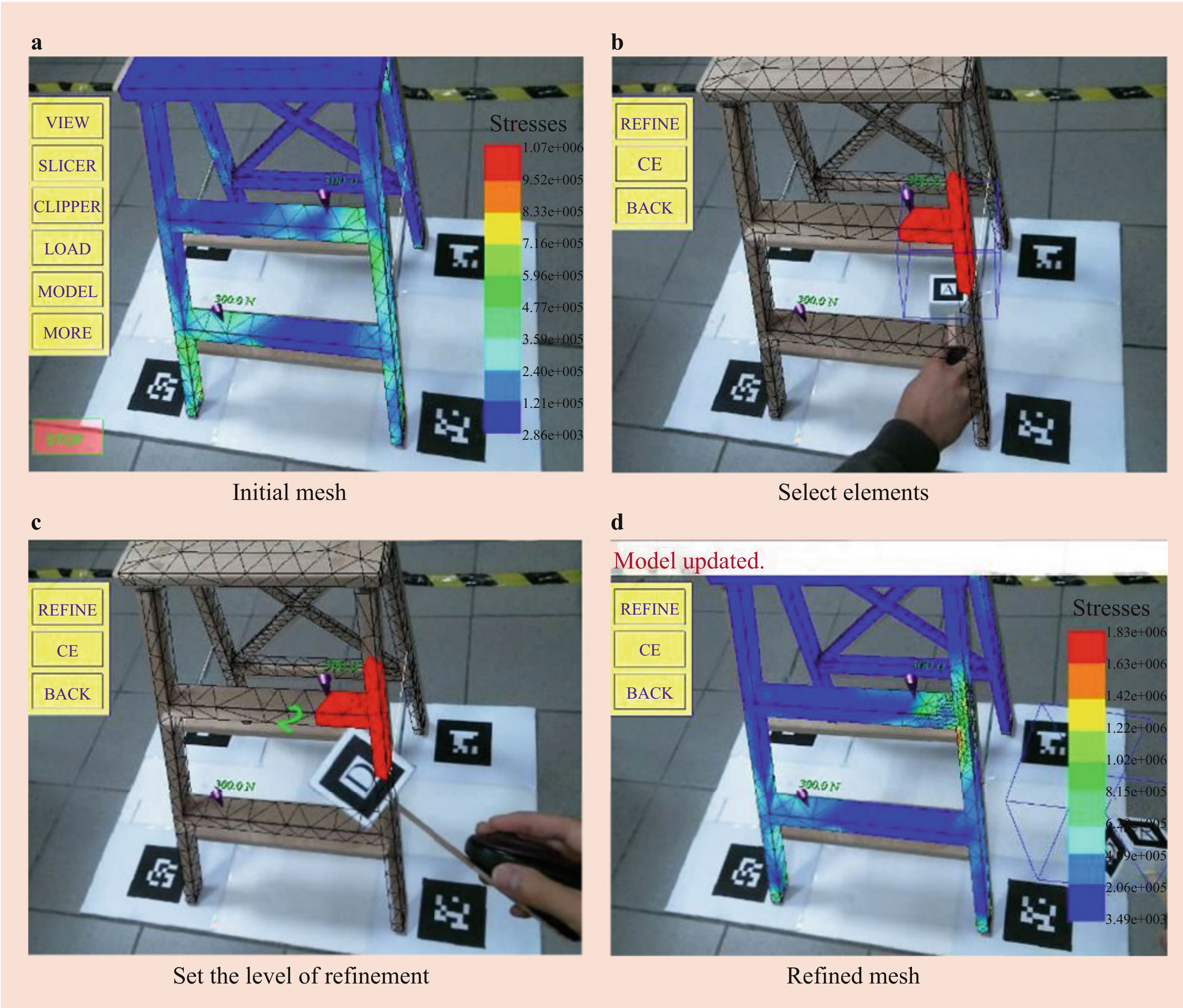


Fig. 22.45 Interactive mesh refinement: (**a**) initial mesh, (**b**) select elements, (**c**) set the level of refinement, and (**d**) refined mesh

Table 22.3 Computation time for model modification (unit: seconds)

Model descriptions	Time for model modification	Time for re-analysis	Response time
The scenario of adding beams			
Step ladder model of 944 nodes and beam models of 54 nodes	0.1	0.3	2.7
Step ladder model of 1926 nodes and beam models of 126 nodes	0.2	0.4	3.6
Step ladder model of 4132 nodes and beam models of 212 nodes	0.2	1.6	4.9
The scenario of mesh refinement			
Model of 1124 nodes (refinement level 1)	1.6	0.4	5.9
Model of 1639 nodes (refinement level 3)	2.0	0.7	6.7
Model of 4788 nodes (refinement level 4)	3.7	2.4	11.7

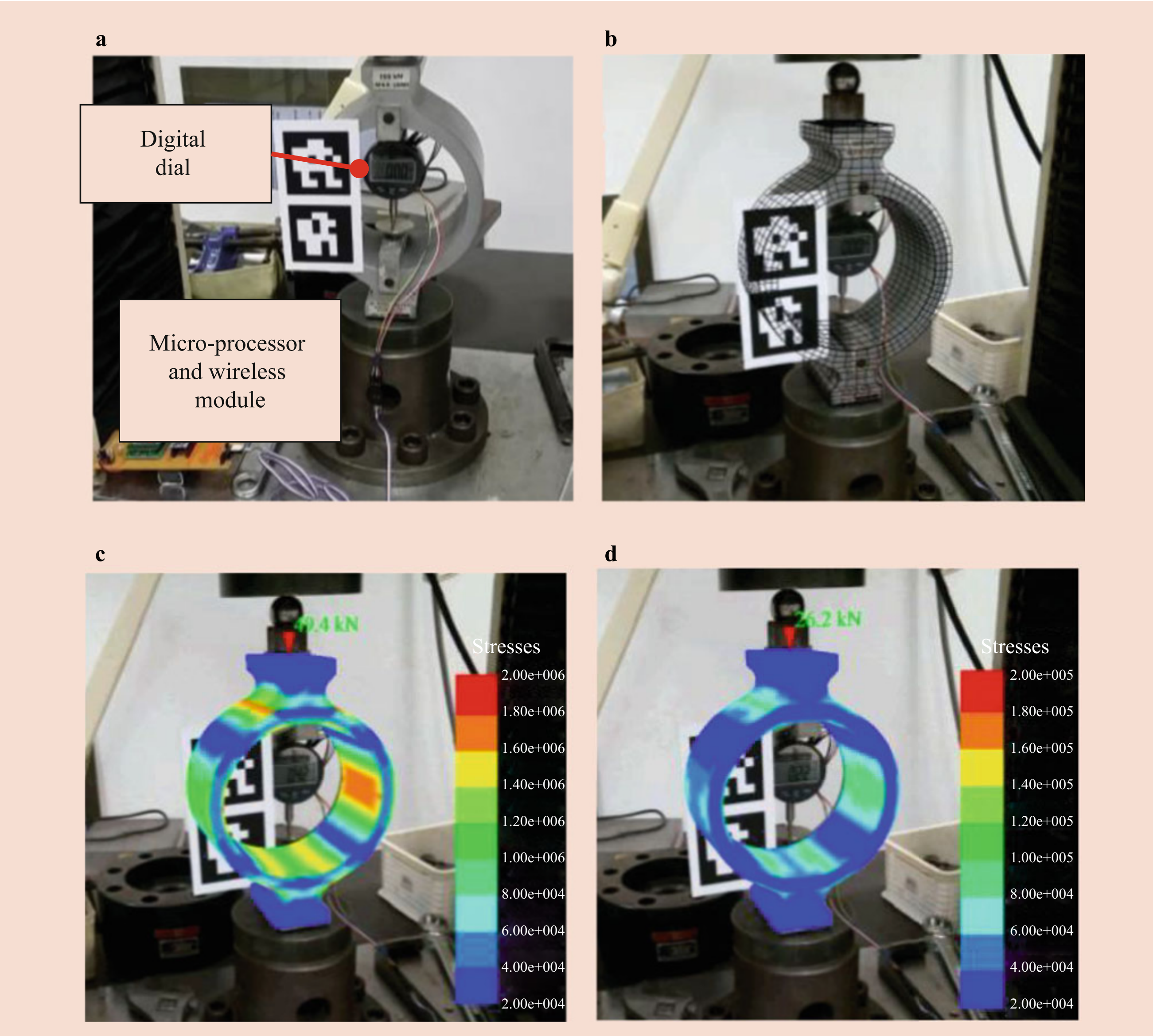


Fig. 22.46 Real-time simulation of a proving ring: (**a**) system setup, (**b**) mesh visualization, (**c**) stress visualization, and (**d**) stresses under a varying load

22.8.1 Some Open Issues

Enrich the Interactive Model Modification Methods

Besides local mesh refinement and adding geometric elements, more interactive methods can be developed to enhance the system performance, such as material minimization while maintaining strength, modification of the location and shape of features, etc. Mesh adaptation methods can be employed to modify meshes directly rather than modifying geometric models using CAD systems. Fast re-analysis techniques can be adopted to reduce the solution time in certain situations. In addition, innovative interaction methods can be achieved by sensing the context to guide model modifications. For instance, by detecting load locations, a mesh is automatically refined in the regions near the loads. By tracking the user's viewpoint, a mesh can be refined automatically in the regions that are in the user's field of view.

Apply Model Reduction Techniques for Efficient Analysis

Efficient interaction methods can be developed by combining model reduction techniques with AR interfaces. Global-local

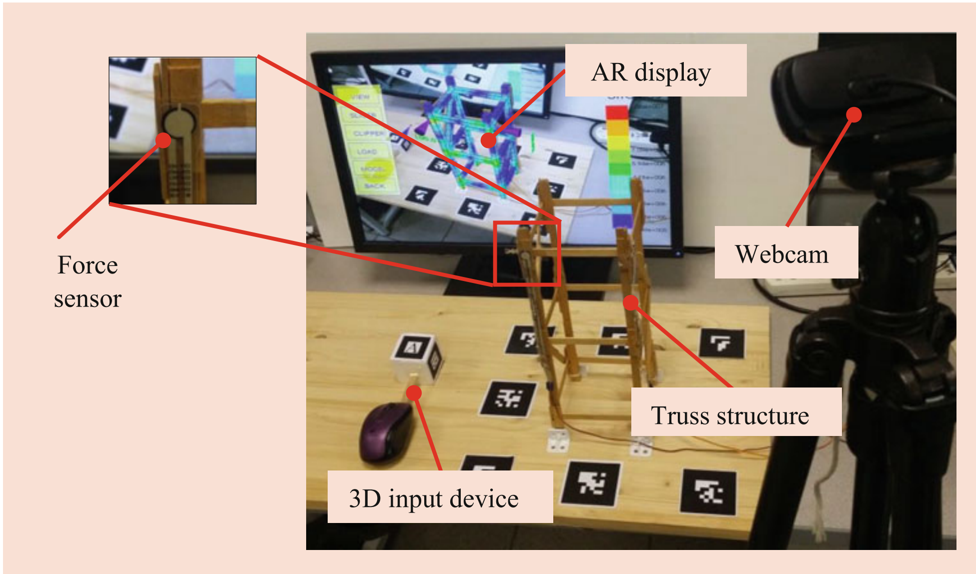


Fig. 22.47 System setup of the prototype application

analysis techniques can be employed when the analysis is to be focused on the regions of interest. For example, a coarse mesh and a fine mesh can be prepared beforehand. In the AR environment, an initial analysis is performed with the coarse mesh. The user can select the region of interest to form a local model using the fine mesh. By using global-local analysis techniques, the local model can be analyzed with the results from the coarse mesh. This interactive method would enable the user to perform accurate analysis on the regions of interest only. Besides, super-element techniques can be employed to condense the components of a structure into super-elements, i.e., substructures. In the AR environment, the user could replace a substructure to a new design for evaluation and select a substructure to recover the interior results for investigation.

Adapt the System to Mobile AR Platforms

The system can be adapted to mobile AR platforms for outdoor and collaborative applications. A client-server environment can be built. The server performs FEA computation, while the client performs tracking and provides user interfaces. The issues on task allocation can be investigated. For example, scientific visualization algorithms can be performed either on the server or on the mobile client. In addition, touchscreen-based interaction methods can be developed.

Real-Time Simulation of Soft Objects in AR

Many real-time FE modeling methods have been developed to generate visually appealing deformation of soft objects. These methods are not suitable for engineering analysis, but can be used for simulating the behavior of flexible objects in prototyping applications, such as the deformation of cables for cable harnessing, the forces needed to assemble flexible products, etc.

Structural Health Monitoring with Finite Element Model Updating

For civil engineering structures, the finite element models usually do not produce results that match with the results from field experiments. A typical reason is the variations in physical parameters due to damage or material deterioration. Finite element updating techniques can be used to update the parameters of the model, identify the locations of damage, and assess the damage condition [73, 74]. In practice, the FE model could be updated periodically with sensor measurements. AR visualization of stress distribution and deteriorated regions may be helpful for monitoring the health of infrastructures.

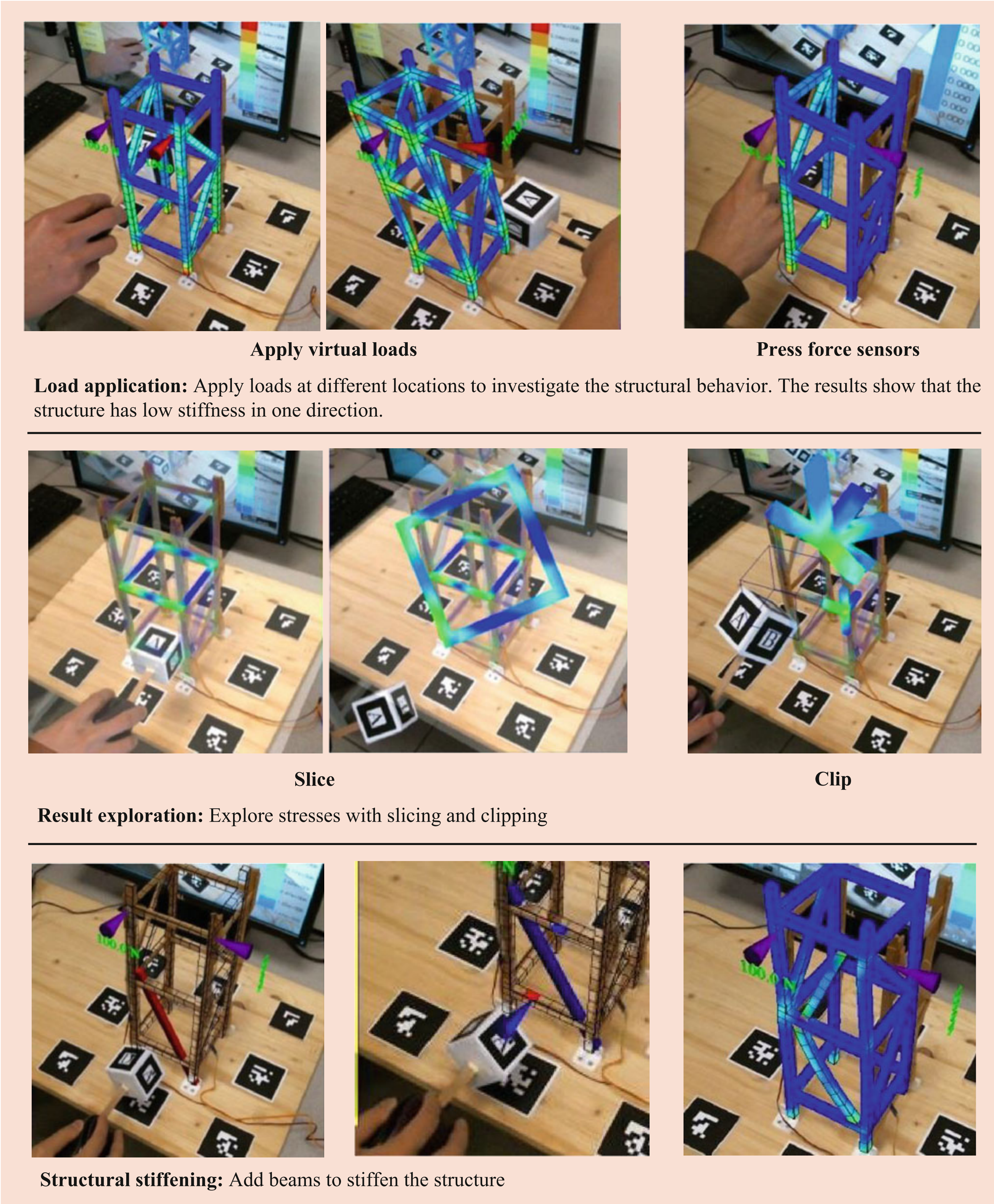


Fig. 22.48 FEA tasks in the user study

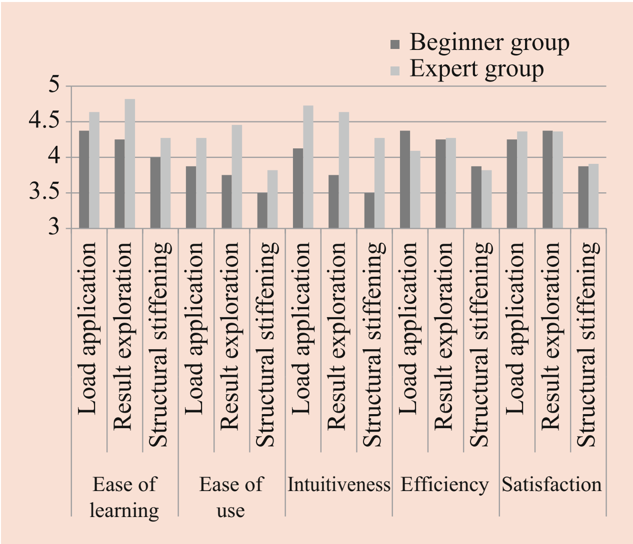


Fig. 22.49 Results of the qualitative evaluation

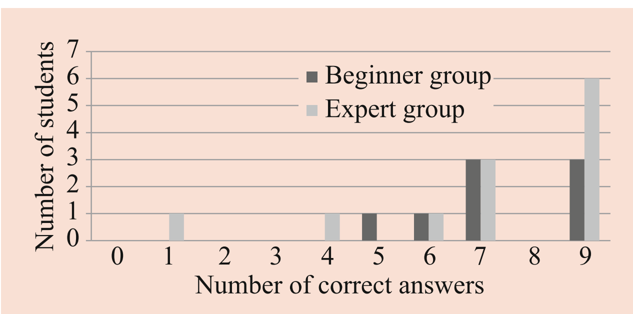


Fig. 22.50 Results for the matching question

References

1. Azuma, R.T.: A survey of augmented reality. Presence Teleoperator Virt. Environ. **6**(4), 355–385 (1997)
2. P. Milgram, H. Takemura, A. Utsumi, F. Kishino: Augmented reality: A class of displays on the reality-virtuality continuum, Proc. SPIE Telemanipulator and Telepresence Technologies, vol. 2351, Boston, United States, October 31-November 1, 1994, pp. 282–292
3. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. Comp. Graph. Appl. **21**(6), 34–47 (2001)
4. Scherer, S., Wabner, M.: Advanced visualization for finite elements analysis in virtual reality environments. Int. J. Interact. Des. Manuf. **2**(3), 169–173 (2008)
5. Hafner, M., Schoning, M., Antczak, M., Demenko, A., Hameyer, K.: Interactive postprocessing in 3D electromagnetics. IEEE Trans. Magn. **46**(8), 3437–3440 (2010)
6. Schroeder, W., Martin, K., Lorensen, B.: The Visualization Toolkit: an Object-Oriented Approach to 3D Graphics, 4th edn. Kitware, New York (2006)
7. Lee, E.-J., El-Tawil, S.: FEMvrml: an interactive virtual environment for visualization of finite element simulation results. Adv. Eng. Softw. **39**(9), 737–742 (2008)
8. Ingrassia, T., Cappello, F.: VirDe: a new virtual reality design approach. Int. J. Interact. Des. Manuf. **3**(1), 1–11 (2009)
9. Setareh, M., Bowman, D.A., Kalita, A.: Development of a virtual reality structural analysis system. J. Archit. Eng. **11**(4), 156–164 (2005)
10. A. Liverani, F. Kuester, B. Hamann: Towards interactive finite element analysis of shell structures in virtual reality. Proc. IEEE International Conference on Information Visualization, London, UK, July 14–16, 1999, pp. 340–346
11. Hambli, R., Chamekh, A., Salah, H.B.H.: Real-time deformation of structure using finite element and neural networks in virtual reality applications. Finite Elem. Anal. Des. **42**(11), 985–991 (2006)
12. Connell, M., Tullberg, O.: A framework for immersive FEM visualization using transparent object communication in a distributed network environment. Adv. Eng. Softw. **33**(7), 453–459 (2002)
13. Cheng, T.-M., Tu, T.-H.: A fast parametric deformation mechanism for virtual reality applications. Comput. Ind. Eng. **57**(2), 520–538 (2009)
14. Yeh, T.P., Vance, J.M.: Applying virtual reality techniques to sensitivity-based structural shape design. J. Mech. Des. **120**(4), 612–619 (1998)
15. Ryken, M.J., Vance, J.M.: Applying virtual reality techniques to the interactive stress analysis of a tractor lift arm. Finite Elem. Anal. Des. **35**(2), 141–155 (2000)
16. D. Rose, K. Bidmon, T. Ertl: Intuitive and interactive modification of large finite element models. Proc. IEEE Visualization, Austin, United States, October 10–15, 2004, pp. 361–368
17. H. Graf, A. Stork: Virtual reality based interactive conceptual simulations: combining post-processing and linear static simulations. Proc. 5th International Conference on Virtual, Augmented and Mixed Reality, Las Vegas, United States, July 21–26, 2013, pp. 13–22
18. L. Chen, T.W. Day, W. Tang, N.W. John: Recent developments and future challenges in medical mixed reality. Proc. IEEE International Symposium on Mixed and Augmented Reality, ISMAR, Nantes, France, Oct 9–13, 2017, pp. 123–135
19. Guha, D., Alotaibi, N.M., Nguyen, N., Gupta, S., McFaul, C., Yang, V.X.X.: Augmented reality in neurosurgery: a review of current concepts and emerging applications. Can. J. Neurol. Sci. **44**(3), 235–245 (2017)
20. Khor, W.S., Baker, B., Amin, K., Chan, A., Patel, K., Wong, J.: Augmented and virtual reality in surgery – the digital surgical environment: applications, limitations and legal pitfalls. Ann. Trans. Med. **4**(23), 454 (2016)
21. Nee, A.Y.C., Ong, S.K., Chryssolouris, G., Mourtzis, D.: Augmented reality applications in design and manufacturing. Ann CIRP. **61**(2), 657–679 (2012)
22. Egger, J., Masood, T.: Augmented reality in support of intelligent manufacturing – a systematic literature review. Comput. Ind. Eng. **140** (2020)
23. Diao, P.-H., Shih, N.-J.: Trends and research issues of augmented reality studies in architectural and civil engineering education – a review of academic journal publications. Appl. Sci. **9**(9), 1–9 1840 (2019)
24. Cardoso, L.F.S., Mariano, F.C.M.Q., Zorzal, E.R.: A survey of industrial augmented reality. Comput. Ind. Eng. **139**, 1–12 (2020)
25. Green, S.A., Billinghurst, M., Chen, X.Q., Chase, J.G.: Human-robot collaboration: a literature review and augmented reality approach in design. Adv. Robot. Syst. **5**(1), 1–18 (2008)
26. Chen, P., Liu, X.L., Cheng, W., Huang, R.H.: A review of using augmented reality in education from 2011 to 2016. In: Popescu, E., Kinshuk, Khribi, M.K., Huang, R., Jemni, M., Chen, N.-S., Sampson, D.G. (eds.) Innovations in Smart Learning, pp. 13–18. Springer Nature, Berlin (2017)

27. Yuliono, T., Sarwanto, Rintayati, P.: The promising roles of augmented reality in educational setting: a review of the literature. Int. J. Educ. Methodol. **4**(3), 125–132 (2018)
28. Karakus, M., Ersozlu, A., Clark, A.C.: Augmented reality research in education: a bibliometric study. Eurasia J. Math. Sci. Technol. Educ. **15**(10), 1755 (2019)
29. Weidlich, D., Scherer, S., Wabner, M.: Analyses using VR/AR visualization. IEEE Comput. Graph. Appl. **28**(5), 84–86 (2008)
30. V. Heuveline, S. Ritterbusch, S. Ronnas: Augmented reality for urban simulation visualization. Proc. First International Conference on Advanced Communications and Computation, Barcelona, Spain, October 23–29, 2011, pp. 115–119
31. Daponte, P., Vito, L.D., Picariello, F., Riccio, M.: State of the art and future developments of the augmented reality for measurement applications. Measurement. **57**, 53–70 (2014)
32. J. Underkoffler, B. Ullmer, H. Ishii: Emancipated pixels: real-world graphics in the luminous room. Proc. 26th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, United States, August 8–13, 1999, pp. 385–392
33. F. Niebling, R. Griesser, U. Woessner: Using augmented reality and interactive simulations to realize hybrid prototypes. Proc. 4th International Symposium on Visual Computing, Las Vegas, United States, December 1–3, 2008, pp. 1008–1017
34. O. Kreylos: Augmented Reality Sandbox. Accessed on January 13, 2016 from http://idav.ucdavis.edu/~okreylos/ResDev/SARndbox/
35. H. Kaufmann, B. Meyer: Simulating educational physical experiments in augmented reality. Proc. ACM SIGGRAPH ASIA 2008 Educators Programme, no. 3, Singapore, December 10–13, 2008, pp. 68–75
36. Valentini, P.P., Pezzuti, E.: Interactive multibody simulation in augmented reality. J. Theor. Appl. Mech. **48**(3), 733–750 (2010)
37. P.P. Valentini, E. Pezzuti: Dynamic splines for interactive simulation of elastic beams in augmented reality. Proc. International Conference on Innovative Methods in Product Design, Venice, Italy, June 15–17, 2011, pp. 89–96
38. F. Mannus, J. Rubel, C. Wagner, F. Bingel, A. Hinkenjann: Augmenting magnetic field lines for school experiments. Proc. IEEE International Symposium on Mixed and Augmented Reality, Basel, Switzerland, October 26–29, 2011, pp. 263–265
39. Malkawi, A.M., Srinivasan, R.S.: A new paradigm for human-building interaction: the use of CFD and augmented reality. Autom. Constr. **14**(1), 71–84 (2005)
40. N. Haouchine, J. Dequidt, I. Peterlik, E. Kerrien, M.-O. Berger, S. Coti: Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. Proc. IEEE International Symposium on Mixed and Augmented Reality, Adelaide, Australia, October 1–4, 2013, pp. 199–208
41. Bernasconi, A., Kharshiduzzaman, M., Anodio, L.F., Bordegoni, M., Re, G.M., Braghin, F., Comolli, L.: Development of a monitoring system for crack growth in bonded single-lap joints based on the strain field and visualization by augmented reality. J. Adhes. **90**(5–6), 496–510 (2014)
42. Fiorentino, M., Monno, G., Uva, A.E.: Interactive 'touch and see' FEM simulation using augmented reality. Int. J. Eng. Educ. **25**(6), 1124–1128 (2009)
43. Matsutomo, S., Mitsufuji, K., Hiasa, Y., Noguchi, S.: Real time simulation method of magnetic field for visualization system with augmented reality technology. IEEE Trans. Magn. **49**(5), 1665–1668 (2013)
44. Matsutomo, S., Miyauchi, T., Noguchi, S., Yamashita, H.: Real-time visualization system of magnetic field utilizing augmented reality technology for education. IEEE Trans. Magn. **48**(2), 531–534 (2012)
45. Nealen, A., Muller, M., Keiser, R., Boxerman, E., Carlson, M.: Physically based deformable models in computer graphics. Comp. Graph. Forum. **25**(4), 809–836 (2006)
46. Cotin, S., Delingette, H., Ayache, N.: Real-time elastic deformations of soft tissues for surgery simulation. IEEE Trans. Visualiz. Comp. Graph. **5**(1), 62–73 (1999)
47. Bro-Nielsen, M., Cotin, S.: Real-time volumetric deformable models for surgery simulation using finite elements and condensation. Comp. Graph. Forum. **15**(3), 57–66 (1996)
48. I. Nikitin, L. Nikitina, P. Frolov, G. Goebbels, M. Göbel, S. Klimenko, G.M. Nielson: Real-time simulation of elastic objects in virtual environments using finite element method and precomputed Green's functions. Proc. Eighth Eurographics Workshop on Virtual Environments, Barcelona, Spain, May 30–31, 2002, pp. 47–52
49. Berkley, J., Turkiyyah, G., Berg, D., Ganter, M., Weghorst, S.: Real-time finite element modeling for surgery simulation: an application to virtual suturing. IEEE Trans. Visualiz. Comp. Graph. **10**(3), 314–325 (2004)
50. Lee, B., Popescu, D.C., Ourselin, S.: Topology modification for surgical simulation using precomputed finite element models based on linear elasticity. Prog. Biophys. Mol. Biol. **103**(2–3), 236–251 (2010)
51. Bathe, K.-J.: Finite Element Procedures. Prentice Hall, Watertown (1996)
52. Joldes, G.R., Wittek, A., Miller, K.: Suite of finite element algorithms for accurate computation of soft tissue deformation for surgical simulation. Med. Image Anal. **13**(6), 912–919 (2009)
53. Miller, K., Joldes, G., Lance, D., Wittek, A.: Total Lagrangian explicit dynamics finite element algorithm for computing soft tissue deformation. Commun. Numer. Methods Eng. **23**(2), 121–134 (2007)
54. Lapeer, R.J., Gasson, P.D., Karri, V.: A hyperelastic finite element model of human skin for interactive real-time surgical simulation. IEEE Trans. Biomed. Eng. **58**(4), 1013–1022 (2011)
55. M. Müller, J. Dorsey, L. McMillan, R. Jagnow, B. Cutler: Stable real-time deformations. Proc. ACM SIGGRAPH/ Eurographics Symposium on Computer Animation, San Antonio, United States, July 21–22, 2002, pp. 49–54
56. Courtecuisse, H., Jung, H., Allard, J., Duriez, C., Lee, D.Y., Cotin, S.: GPU-based real-time soft tissue deformation with cutting and haptic feedback. Prog. Biophys. Mol. Biol. **103**(2–3), 159–168 (2010)
57. Allard, J., Cotin, S., Faure, F., Bensoussan, P.-J., Poyer, F., Duriez, C., Delingette, H., Grisoni, L.: SOFA-an open source framework for medical simulation. In: Westwood, J.D., Haluck, R.S., Hoffman, H.M., Mogel, G.T., Philips, R., Robb, R.A., Vosburgh, K.G. (eds.) Medicine Meets Virtual Reality 15, pp. 13–18. IOS Press, Amsterdam (2007)
58. Sin, F.S., Schroeder, D., Barbic, J.: Vega: non-linear FEM deformable object simulator. Comput. Graph. Forum. **32**(1), 36–48 (2013)
59. Portnoy, S., Yarnitzky, G., Yizhar, Z., Kristal, A., Oppenheim, U., Siev-Ner, I., Gefen, A.: Real-time patient-specific finite element analysis of internal stresses in the soft tissues of a residual limb: a new tool for prosthetic fitting. Ann. Biomed. Eng. **35**(1), 120–135 (2007)
60. Yuan, M.L., Ong, S.K., Nee, A.Y.C.: The virtual interaction panel: an easy control tool in augmented reality systems. Comp. Animat. Virtual Worlds. **15**(3–4), 425–432 (2004)
61. J. Looser, M. Billinghurst, R. Grasset, A. Cockburn: An evaluation of virtual lenses for object selection in augmented reality. Proc. 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia, Perth, Australia, December 1–4, 2007, pp.203-210
62. Argelaguet, F., Andujar, C.: A survey of 3D object selection techniques for virtual environments. Comput. Graph. **37**(3), 121–136 (2013)

63. J. F. Lucas: Design and evaluation of 3D multiple object selection techniques, Master's thesis, Virginia Polytechnic Institute and State University, Virginia, United States, 2005
64. R.B. Haber, D.A. McNabb: Visualization idioms: a conceptual model for scientific visualization systems, in Nielson, G.M. and Shriver, B., Visualization in Scientific Computing, LosAlamitos IEEE Computer Society Press, 1990, 74–93
65. Santos, S.D., Brodlie, K.: Gaining understanding of multivariate and multidimensional data through visualization. Comput. Graph. **28**(3), 311–325 (2004)
66. G. Debunne, M. Desbrun, M.-P Cani, A.H. Barr: Dynamic real-time deformations using space and time adaptive sampling. Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, United States, August 12–17, 2001, pp. 31–36
67. Silva, C.W.D.: Vibration and Shock Handbook. CRC Press Taylor and Francis (2005)
68. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipe in C: The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom (1992)
69. I.S. Duff, R.G. Grimes, J.G. Lewis: Rutherford Appleton Laboratory, Central Computing Department, Atlas Centre, Oxon OX11 0QX, United Kingdom. Accessed January 14, 2016 from http://www.cs.colostate.edu/~mcrob/toolbox/c++/sparseMatrix/hbsmc.pdf
70. J.F. Doyle: Modern Experimental Stress Analysis: completing the solution of partially specified problems. Wiley (2004)
71. D. Winer: XML-RPC Specification. Accessed January 14, 2016 from http://xmlrpc.scripting.com/spec.html
72. Hager, W.W.: Updating the inverse of a matrix. SIAM Rev. **31**(2), 221–239 (1989)
73. Jaishi, B., Ren, W.-X.: Damage detection by finite element model updating using modal flexibility residual. J. Sound Vib. **290**(1–2), 369–387 (2006)
74. Teughels, A., Maeck, J., Roeck, G.D.: Damage assessment by FE model updating using damage functions. Comput. Struct. **80**(25), 1869–1879 (2002)
75. Huang, J.M., Ong, S.K., Nee, A.Y.C.: Real-time finite element structural analysis in augmented reality. Adv. Eng. Softw. **87**, 43–56 (2015)

J.M. Huang received his B.Eng. degree in mechanical engineering from Northwestern Polytechnical University (China) in 2008, M.Sc. degree in electromechanical engineering from University of Macau (China) in 2011, and Ph.D. degree in mechanical engineering from National University of Singapore in 2016. Currently, he is a consultant providing operations research and artificial intelligence technology services to a variety of industry sectors.

Soh Khim Ong is Associate Professor at NUS. Her research interests are virtual and augmented reality applications in manufacturing and life cycle engineering. She is Fellow of CIRP. She has published 5 books and over 330 international refereed journal and conference papers. She received 2004 OYME Award from SME and 2009 Emerging Leaders Award in Academia (US Society for Women Engineers).

Andrew Yeh Ching Nee received his Ph.D. and D.Eng. degrees from UMIST, Manchester. He is currently Emeritus Professor in the Department of Mechanical Engineering of NUS. He is Fellow of CIRP, SME, and Academy of Engineering, Singapore. His research interests are in manufacturing engineering, augmented reality, and digital twin. He is Editor-in-Chief of IJAMT and Executive Editor-in-Chief of AIM.

Augmented Reality in Maintenance: A Review of the State-of-the-Art and Future Challenges

23

Soh Khim Ong, C. Y. Siew, and Andrew Yeh Ching Nee

Contents

23.1 Augmented Reality Applications in Maintenance . . . 575

23.2 Summary of Reviews of Augmented Reality in Maintenance . . . 578

23.3 Literature Review . . . 579
23.3.1 Applications . . . 580
23.3.2 Contextual Awareness . . . 581
23.3.3 Tracking . . . 581
23.3.4 Human Motion Tracking . . . 582
23.3.5 User Interface and Interaction . . . 583
23.3.6 Hardware . . . 584

23.4 Discussion . . . 585
23.4.1 Feasibility of AR Application in Industrial Maintenance . . . 586
23.4.2 Uniform Implementation Method for AR in Maintenance . . . 586

23.5 Research Gaps Based on Current Technological Developments . . . 587
23.5.1 A Unified and Robust Design Approach to Achieve Contextual Awareness . . . 587
23.5.2 UI Design for AR in Remote Maintenance . . . 587
23.5.3 IoT and Digital Twin Integration . . . 587
23.5.4 Improvements in User Tracking . . . 588
23.5.5 Ergonomic Assessment of Current Maintenance Practices . . . 588
23.5.6 Potential Future Research Focus Areas . . . 588

23.6 Conclusion . . . 591

References . . . 591

S. K. Ong (✉) · A. Y. C. Nee
Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore
e-mail: mpeongsk@nus.edu.sg; mpeneeyc@nus.edu.sg

C. Y. Siew
Mechanical Engineering Department, National University of Singapore, Singapore, Singapore
e-mail: mpescy@nus.edu.sg; siewchiyung@u.nus.edu

Abstract

In maintenance, the use of augmented reality (AR) has often been perceived as a solution in resolving the various problems faced by the industrial sector. For any reliable production line, maintenance is crucial in order to ensure the safe, continued, and productive operation of the entire manufacturing process. However, the difficulty of maintaining modern machinery and products is increasing exponentially such that operators may not be able to cope. As a human-computer interaction (HCI) tool, AR can potentially provide the competitive advantage to the maintenance operators to keep up with this rate of increase. This chapter presents new research and application developments in AR in maintenance. The chapter provides a holistic literature review on the recent developments in the state of the art of AR in maintenance. Current research directions and future developments are discussed.

Keywords

Maintenance · Augmented reality · Human-computer interaction · IoT

23.1 Augmented Reality Applications in Maintenance

Maintenance can be described as the task of sustaining the reliability and operating conditions of existing plant assets, products, or equipment [1]. In recent years, this definition of maintenance has been expanded to include more in-depth reliability engineering concepts, where operators proactively suggest and implement solutions in the mitigation of potential machine failures [2]. Maintenance operators are expected to be able to cope with the increasingly complicated maintenance procedures for equipment, products, and systems that have become more complex progressively. The traditional pen and paper approach of using manuals and maintenance

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_23

documentation is increasingly becoming inadequate and outpaced by the rapid evolution and increasing complexity of modern industrial machines [3]. The constant evolution and the increasing importance of maintenance pose substantial problems to maintenance operators and may hinder their ability to work productively.

Augmented reality (AR) has been perceived as a solution to many problems faced by the industrial sector [4]. In maintenance, the ability of AR in integrating virtual information with real industrial environments has resulted in AR being an attractive option for simplifying the complex workflow of maintenance operators [5]. A typical AR maintenance system structure commonly described in various research papers and commercially available systems is illustrated in Fig. 23.1. AR may provide a competitive advantage to maintenance operators to help them keep up with the problems of increasing machine complexity. Firstly, AR allows users to be more attentive and efficient in dealing with the information associated with the maintenance tasks through providing and augmenting only information that is relevant to the operators' maintenance tasks [6]. AR enables a scaled approach for guidance of maintenance tasks by providing different levels of supports based on different combinations of videos, texts, and instructions of the maintenance tasks. For expert operators, AR can be used to provide additional documentation support as and when needed; for non-experts, AR can provide more dynamic assurance support to the operators during maintenance. These aspects have been validated and discussed by various researchers [7, 8].

In contrast to the early days of AR applications in maintenance, the availability of more capable AR enabling technologies has resulted in the increasing feasibility of using AR in industrial scenarios. However, not all areas of maintenance will benefit from the use of AR. From some of the key systems that have been developed for AR in maintenance, it can be observed that most of the AR systems fall into two categories, namely, (1) AR-assisted maintenance guidance and (2) AR-assisted maintenance training.

(1) AR-Assisted Maintenance Guidance Systems

An AR-assisted maintenance guidance system is the most common type of AR maintenance applications since one of the key goals is to allow the use of the system without additional training, although it is still not fully realized in existing systems due to safety and practical concerns. Regardless, the ability of AR in superimposing useful information for the user has the potential to reduce the burden faced by existing technicians, who will not be required to carry or memorize cumbersome manuals in contrast with the current maintenance approach. Users can instead focus on the task at hand. Consequently, the system may potentially be able to save time, lower user fatigue, increase productivity, and reduce errors.

(2) AR-Assisted Maintenance Training Systems

Due to the potentially high stakes involved, a trainee may not be able to have sufficient maintenance training conducted on the job for him/her. The traditional approach to maintenance training involves the trainee going through existing manuals and conducting maintenance on a decommissioned setup whenever available. However, this results in the scenario where the users are being trained using static information instead which may have limited effectiveness in helping the trainee to understand the scenario at hand.

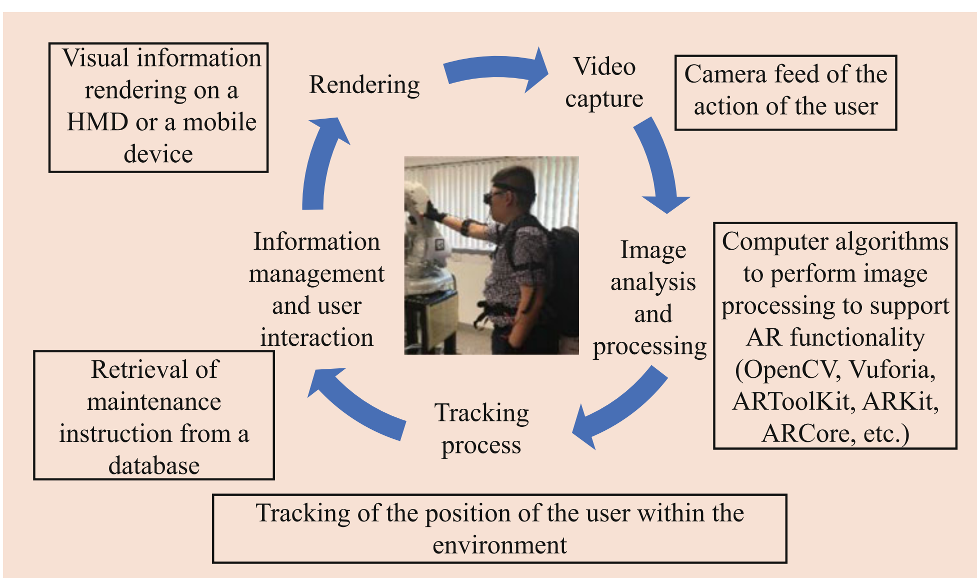


Fig. 23.1 Typical workflow of an AR maintenance system

The ability for AR to overlay virtual objects in a real scenario is able to better simulate a potential situation that the novice operator may face. This approach in turn allows the trainee to be more familiar with the actual maintenance context.

AR has long been identified as a potential panacea for the many problems faced in the industrial sector [4]. In recent years, the rise of publicly available AR-based development toolkits and related hardware has helped spur academic research and commercial interest toward AR applications in maintenance. The ability of AR to integrate virtual information into real industrial environments has resulted in AR being an attractive option for simplifying the increasingly complex workflow of a modern maintenance job operator [5]. It is a competitive advantage that AR may potentially provide to maintenance operators to help them keep pace with this evolution. The increase in interactivity that AR provides allows users to be more attentive and efficient in dealing with the information associated with the maintenance task at hand [6]. In terms of worker support, AR represents a more tangible scaled approach to maintenance tasks in general, one in which video and manuals alone may not be able to provide. For expert users, AR will be able to provide additional documentation support as and when needed; for non-experts, it may potentially provide more dynamic assurance and support to the worker during maintenance [7, 8]. The benefits of AR in maintenance can be summarized next.

(1) Reduction in Time and Effort

Apart from the benefits in assisting a maintenance operator in understanding the maintenance assignment, a digital manual will allow on-the-fly updates to the corresponding instructions and state of the machines, ensuring that an operator is always up to date. The operators will be able to concentrate on the tasks since the virtual information can be augmented at the corresponding position in the physical environment.

(2) Remote Collaboration and Guidance

The ability of AR to overlay information directly in front of users provides a more convenient way for multiple users, whether geographically apart or co-located, to collaborate on a maintenance problem as and when required. When a maintenance operator encounters a problem, he/she can request for assistance easily from experts who may be remotely located. Remote maintenance is receiving increasing attention from the industry. There are situations where a remote supervisor may need to convey information to an operator on-site. A variety of AR tele-maintenance systems have been reported. Zhu et al. [9] proposed an authorable remote tele-maintenance system. Bottecchia et al. [10] have proposed a remote AR-assisted maintenance interface which promotes collaboration between different parties over the Internet.

(3) Cost Reduction

The application of AR in maintenance can achieve cost reduction in two ways. It has been shown that the use of AR can aid in reducing the number of errors during maintenance [11]. The use of AR as a teaching/training tool has the potential of reducing the training time and cost [12] due to the increase in interactivity that an AR-assisted system provides.

Despite the effectiveness of AR in maintenance coupled with industrial pressure to address the problems faced by modern maintenance operators, the application of AR in industrial applications is not extensive. The use of AR as a support tool in maintenance and maintenance training can be traced back to the early 1990s [13]. However, most of the reported research on AR in maintenance focused primarily on investigating the applicability and viability of AR in providing guidance and instructions to the operators. Various studies have shown the potential efficacy of AR in alleviating the problems faced by the industry. However, there are many roadblocks that need to be resolved to ensure the safe and successful implementation of AR in maintenance applications. AR is an HCI tool where practical considerations of human involvement must be considered in order to maximize the utility of AR. Unfortunately, many of the reported solutions are not guided by HCI principles in designing a practical AR interface [14] and cannot be implemented safely in their current development state [15]. This indicates a lack of understanding on the requirements to achieve a practical AR in maintenance system.

Most existing AR-assisted maintenance systems that have been reported have no direct influence on the quality of the maintenance works and/or assessment of the potential improvement in the quality of the work achieved with the aid of AR. The outcome and quality of a maintenance procedure are still not assessed. This has implications especially in situations where a high level of accountability is required to ensure that a maintenance job has been performed to the required standards. This problem is further exacerbated by the increasing use of workers from various backgrounds where language barriers or experience levels may hinder communication between these workers.

This chapter provides a review of the literature with regard to AR and its application in the field of maintenance. The research focus and the practical implementation issues of AR in maintenance will be reviewed critically. This proposes solutions and potential future research and development trends of AR in maintenance. The chapter begins with an overview of the current state of the art and other related developments of AR in maintenance. The review methodology adopted and a detailed analysis of the current state of the art in AR in maintenance are presented. Issues relating to the future trends of the research are examined. The research gaps identified through this literature review are discussed.

23.2 Summary of Reviews of Augmented Reality in Maintenance

Before presenting a comprehensive review of AR in maintenance, a few earlier review articles [16–21] are discussed. A summary of the scope of each literature review and a few existing AR-assisted maintenance systems are presented in Tables 23.1, 23.2, and 23.3, respectively.

Due to the increasing popularity of AR in industrial applications, there has been a surge of AR maintenance-related review papers in recent years. However, many of the review papers focus on outlining the concepts of AR in a specific use case. They do not cover extensively all the topics involved of AR in maintenance and are generally fragmented in nature.

Lamberti et al. [16] discussed tracking-related issues and mentioned briefly the usability of the reviewed AR-assisted maintenance systems. Borsci et al. [20] reviewed extensively the benefits of AR in maintenance but did not discuss the technical implementation issues in depth. Most recently, Palmarini et al. [21] have started unifying all the different aspects

Table 23.1 Summary of the reported literature reviews

Review	Scope	Topics and issues discussed
[16]	Discussion on new challenges and opportunities of AR in maintenance	Tracking-related problems; usability and system reconfiguration of AR systems Most AR-assisted maintenance systems are developed for a specific purpose Tracking is still an issue for system realization
[17]	Brief discussion of an AR-assisted maintenance system and the current use of AR in actual industrial scenario	AR information must be content specific and dynamic Users tend to follow instructions blindly when instructions are presented as it is Limitations of current enabling technologies and the importance of HCI
[18]	Discussion on main challenges in implementing AR in industrial scenario via a case study	Technical issues involved in the usability of AR-assisted maintenance systems A streamlined framework for content generation is necessary Devices need to be easy to use and simple AR systems must fulfill the cost/benefit requirements of the organization
[19]	Brief discussion on AR to support through life engineering, which includes maintenance	Technological issues related to AR in maintenance Benefits of AR for support from external sensorial devices for visualization AR systems must fulfill the usability and portability requirements of the hardware for user comfort
[20]	Extensive review on the evaluation of virtual and mixed reality tools for automotive maintenance operator training	Extensive review on the performance review methodologies of operators during maintenance Different users have different responses when using an AR system A dynamic content generation approach is required
[21]	Review of the developments of AR in maintenance and the related issues	Various aspects of AR in maintenance and importance of proper UI development Improvements in hardware, tracking, and interaction necessary before AR can be practically implemented

Table 23.2 Summary of a few AR-/VR-assisted maintenance systems

Groups/projects	Institutions	Area	Features
Boeing [13]	Boeing	Maintenance training, ergonomics	Heads-up HMD Skill-based training
ARVIKA [22, 23]	BMBF, DLR, Airbus Deutschland, EADS, DaimlerChrysler, VW, AUDI, Ford and BMW, DS technologies, Hüller-Hille, Gühring, Index, Ex-Cell-O, Siemens	Diagnosis, maintenance, design	User-centric design Plug-in-based implementation for easier integration
BMW augmented reality [24]	BMW	Maintenance training and repair	Automatic recovery from tracking failure Mobile-based training solution
The common augmented reality platform (CAP)	Bosch, Reflekt GmbH	Diagnosis, maintenance, repair	Generic platform adaptable to multiple different implementations
http://www.armedia.it/i-mechanic [19]	AR-media	Maintenance guidance	Mobile-based Link with existing car maintenance networks (for repairs beyond the normal users)
Cranfield University [25]	Cranfield University	Maintenance training	Personalized training
ALEn3D [26]	CFE Mexico	Maintenance training	Virtual training platform for high-risk maintenance situation

Table 23.3 Summary of AR-assisted maintenance systems

System	Summary	Methodology	Pros	Cons	Passive
[3, 6]	Prototype AR application to support military mechanics in routine maintenance tasks inside an armored vehicle turret	Uses a tracked head-worn display to augment a mechanic's natural view	Allow the imposition of image and information directly in front of a user	No feedback on the state of machine repair	Yes
[11]	An AR interface designed to support bringing AR technology in aircraft maintenance	Marker-less tracking (image mosaic method. Tracking of plane). HMD	Provide information and assistance during simple aircraft maintenance	Static information	Yes
[27]	Technical maintenance with interactive AR instruction via large screen	I. Unifeye engineer (commercial). II. Screen-based video see-through display	Showcase the effectiveness of AR in maintenance guidance in a screen-based representation	No feedback on user performance	Yes
[28]	Using AR as an interactive, intuitive tool to capture current machine state and to generate multimedia maintenance report	I. Point and edge-based detection method. II. Context awareness realized via graph-based logical method	Capture current machine state so as to provide the relevant maintenance information and generate the corresponding report	No step verification	Partial
[19]	A system that aims to reduce the error of an assembly workstation via the integration of the sensors with an AR system	I. An assembly table with force sensors embedded. II. Sensors are used to determine assembly conditions. CCD camera is used to determine errors that are undetectable by the force sensors	Capable of performing in process error detection. Capable of selecting the recovery procedure	Long system setup time	Yes
[29]	A methodology to provide dynamic instructions (depending on the operator) during the training of a maintenance operator. An expert system	Open rules format for logic implementation. Level of instructions dynamically determined during maintenance situation	Able to dynamically adapt the level of details required for an operator depending on how well the operator is doing	Information dynamic rules are based on time	Partial
[30]	An AR content authoring system allows ordinary users to easily program and apply interactive features to virtual objects	RGB-D camera (Kinect) for hand detection and gesture interaction	Ease of use for users with no programming experience in authoring and modifying AR content	No user performance feedback	Yes

of AR in maintenance in their review paper and outlined the possible future directions of the research. However, there are other issues and aspects that have not been dealt with due to the rapidly evolving nature of this field. This review chapter aims to complement previous review works that have been established with regard to AR in maintenance.

23.3 Literature Review

To highlight the state of the art for AR in maintenance and to evaluate any potential research gaps, the taxonomy in Fig. 23.2 has been defined [31]. The taxonomy is selected based on fields that have been discussed in the various AR in maintenance review papers. For example, AR-assisted maintenance systems are often used as support systems for a variety of industrial operations or as training platforms [21]. Ontology is a common approach used to achieve contextual awareness [32] although various other techniques exist. For tracking-related issues, computer vision-based tracking techniques and motion tracking of the users are discussed commonly [33]. User interface (UI) developments refer to user interface development issues or methods in presenting the augmentation information.

In this present literature review, extensive research papers have been reviewed based on the taxonomy in Fig. 23.2. The following review goals have been defined, namely, (1) assessment of the feasibility of AR-assisted maintenance systems in the industry, (2) identification of common implementation frameworks for AR in maintenance, (3) identification of research gaps based on current research developments, and (4) establishment of future trends of AR in maintenance.

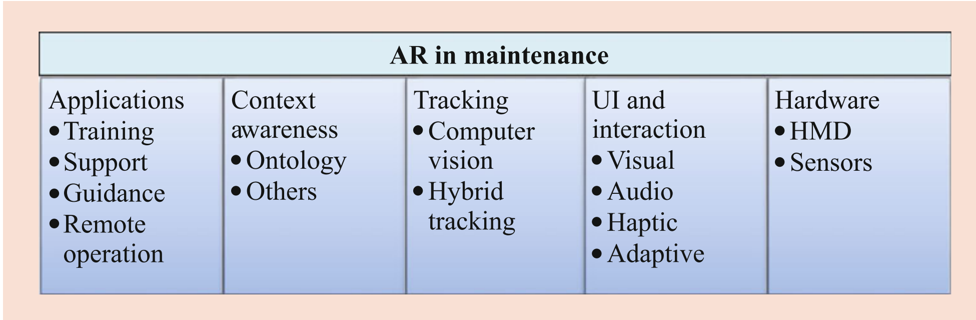


Fig. 23.2 AR in maintenance taxonomy

Table 23.4 Comparison of the different applications

Application	Training	Support	Both
Pros	Easy to implement Less safety concerns	More realistic systems can be developed	Broader applicability of the systems
Cons	Not fully representative of actual maintenance scenario/requirements	Safety regulations and concerns	Training and support requirements may conflict
References	[38–49]	[8, 10, 27, 28, 38–69]	[38–49]

23.3.1 Applications

Application refers to the common use cases of AR-assisted maintenance platforms. From the literatures, it can be noted that there has been an increase in the research on the applications of AR as compared to fundamental AR research in recent years. This observation is consistent with results presented by Kim et al. [31] and other research reported recently [34–37].

Maintenance is a process that includes operations, such as assembly, disassembly, repair, inspection, diagnosis, and/or training operations [21]. Hence, AR-assisted maintenance systems can be classified into training, support (assembly, repair, inspection, and diagnosis), or both as presented in Table 23.4.

From Table 23.4, it can be observed that many of the AR-assisted maintenance applications that have been reported are maintenance support systems rather than training systems, which is consistent with the findings by Palmarini et al. [21]. The ideal capability of an AR-assisted maintenance system is to allow a maintenance operator to conduct his/her maintenance job with minimal or no additional training [70]. The support type applications have received significant attention in the last 10 years. Traditional AR systems involve bulky setups with limited computational capabilities to process the required algorithms and the display of the necessary augmentation [71], making them more suitable as training and education platforms. With the rise of modern lightweight and increasingly powerful platforms, AR can be applied directly in support-based systems, and AR systems for maintenance support are increasingly being accepted as a potential tool for maintenance operators to enhance their current workflow. For example, companies, such as IBM and Hyperloop, have considered the implementation of AR in their maintenance workflow [72].

From the literature survey, there has been a trend in introducing additional functionalities into AR to include more aspects of maintenance. One example is the ability to conduct remote maintenance using the AR-assisted maintenance systems. Traditional AR-based maintenance systems mainly focus on supporting an operator working at a specific location. There has been an increasing interest in using AR for remote and/or off-site maintenance, either as a form of guidance via a remote authoring and collaboration framework [73], as a supervisory function, or both. Many researches in this area have been reported. Masoni et al. [74] suggested possible enhancements of existing AR-based maintenance applications via the integration of Industry 4.0 concepts together with remote maintenance capabilities. Aschenbrenner et al. [63] created the ARTab platform to help improve situational awareness for remote maintenance. Bottecchia et al. [10] proposed an Internet-based tele-assistance system for in-the-field maintenance. Wang et al. [68] introduced a multi-user framework coordination for on- and off-site users during maintenance. However, it is important to note that in these reported systems, remote operator interactions are often restricted to the WIMP (windows, icons, menus, and pointer) interface. It should be noted that most of the examples listed in Table 23.3 have not moved beyond the prototyping stage. This is because an implementation-ready application would require the system to be able to fulfill the various safety and legal requirements that exist within the various industries [45].

23.3.2 Contextual Awareness

Contextual awareness refers to the ability of a system to understand and handle the information about a current location or the task of a user [32]. A contextually aware AR-assisted maintenance system will be more efficient in presenting information to a user since the information can be labeled and paired automatically with minimal direct user input. There have been a variety of ways in which improved contextual awareness can be achieved. An approach involves the establishment of advanced ontological approaches to better model the data that is associated with the maintenance tasks being conducted. Zhu et al. [75] proposed an ontological maintenance framework that simplifies the authoring process of AR content for a maintenance scenario. Del Amo et al. [76] have taken a similar approach, where common contextual information related to maintenance is categorized systematically to model the necessary information in AR-assisted maintenance. Neges et al. [28] used a graphical-based approach to implement contextual awareness so as to track the actions that have been performed by a user.

Table 23.5 represents the different tools and approaches that have been used to implement contextual awareness in AR-assisted maintenance systems. It can be observed from Table 23.5 that there is no standardized approach in achieving contextual awareness. Most of the approaches have been tested only in specific industrial domains in maintenance due to the different domain requirements and regulation conformance issues [45]. In addition, there has been no formal approach to benchmark the different approaches in Table 23.5, making performance evaluation difficult.

The ease of implementation of the approaches varies. Ontology and rule-based approaches are well-received since the rule sets can be determined quickly, whereas algorithmic graph-based models and machine learning approaches would require extensive modelling and training, which can hinder the ease of implementation.

Researchers have taken other approaches other than those listed in Table 23.5 to implement contextual awareness in AR-assisted maintenance systems. Oliveira et al. [62] proposed a gamified interface in which contextual awareness of the user actions is necessary; however, no detailed discussion was elaborated. Benbelkacem et al. [57] presented a proof-of-concept prototype that evaluates the effectiveness of AR in maintenance prototype with some forms of contextual awareness; however, the approach used was not discussed. Many other applications investigated in this literature review proposed customized frameworks, which may only be feasible in the authors' use case and may not be considered as true contextual awareness implementations.

Contextual awareness is a crucial aspect in any AR-assisted maintenance system, and information of the actions of a user needs to be collected for contextual awareness systems to be more robust. As a context can be any entity that leads to information presented correctly [32], any ontology or data model can only be as useful as the information that can be collected. The current lack of robust motion tracking research in AR in maintenance has restricted the information that can be collected. The problem is further compounded by a lack of techniques required for industrial implementation [79].

Table 23.5 Examples of approaches used to achieve contextual awareness

Study	Approach	Ease of implementation
[28]	Algorithmic graph-based model	–
[43]	Data modelling using XML	+
[45]	Ontology-OWL (KBS)	+
[47]	Rule-based (OpenRules)	+
[50]	Ontology (SOW)	–
[52]	Data modelling (AROMA-FF)	–
[58]	Data modelling (AROS-system)	–
[69]	Machine learning	–
[75]	Ontology	+
[77]	Data modelling (DWARF)	–
[78]	Interaction modelling (IRVO)	–

23.3.3 Tracking

Tracking forms one of the core technological tenets of AR since the accurate tracking of a user's space is a requirement for the correct positioning of augmented virtual objects in the real environment [80], in addition to the detection of a user's own action. It is a topic that continues to challenge AR in maintenance research due to the difficulty of creating a suitable turn-key tracking solution in industry scenarios [81]. Tracking methods can be grouped into a few categories, namely, marker-based tracking [82], markerless tracking [83], hybrid tracking [84], or sensor-based tracking [85], as illustrated in Table 23.6.

Most developments in tracking for AR focus on the technologies to allow accurate positioning of the virtual information in a physical workspace and less on the verification of the actual steps being conducted by the users, although there has been an increase in research on this issue [79] [86]. Lima et al. [87] proposed a markerless tracking system for AR in the automotive industry. Li and Fan [88] proposed a model-based tracking approach for use in the outdoor environment. Zhou et al. [89] have reviewed a variety of different tracking methodologies. The emergence of publicly available tracking libraries, such as Vuforia [90],

Table 23.6 Advantages and disadvantages of different AR tracking algorithms

Category		Advantages	Disadvantages
Visual tracking	Marker-based tracking	Accurate; fast; stable, universalized hardware and software; commonly used	Marker requirement. Susceptible to lighting conditions
	Markerless tracking	Applicable in an unprepared environment; relieves the effect of interruptions on the registration due to the application of natural features	Trade-off between real time and accuracy of registration
Sensor-based tracking		Simple registration algorithm; high running speed [33]	Sometimes expensive; limited natural interactivity due to extra facilities; calibration required
Hybrid tracking		Complement each other to realize the maximum function	

Table 23.7 Pros and cons of marker-based and markerless approaches

Types	Marker-based	Markerless
Pros	Accurate; fast; stable, universalized hardware and software; commonly used	Applicable in an unprepared environment; relieves the effect of interruptions on the registration due to the application of natural features
Cons	Marker requirement. Susceptible to lighting conditions	Trade-off between real time and accuracy of registration

ARCore [91], and ARKit [92], has made the deployment of AR maintenance applications easier.

Table 23.7 shows the pros and cons of marker-based and markerless tracking methodologies. It can be observed that there has been a shift toward markerless and other types of tracking for AR in maintenance. For proof-of-concept applications, the marker-based method is the dominant method for information presentation [44] since it is the easiest to implement. However, the popularity of markerless tracking has shown a noted increase recently due to the rapid developments of other enabling technologies. This trend may be beneficial in industrial applications since marker-based approaches may not be practical due to markers' placement and durability issues [27]. The decrease in the use of marker-based tracking technologies in recent years may be attributed to these developments [31].

23.3.4 Human Motion Tracking

Tracking is also critical in the tracking of a user motion. The information of a user motion can be used to enhance the contextual awareness of the systems. Macro-positioning of a user may be described using techniques previously described in various literatures [93]. A more detailed tracking of a user's actual actions may be used to provide further context of the user's actions. Paz et al. [94] proposed a SURF-based algorithm for the detection of changes in the maintenance environment and detecting whether a step has been performed during maintenance. Unfortunately, there is limited research in this aspect for maintenance. Most of the researches on user motion tracking are not in AR maintenance but for gesture recognition. Stiefmeier et al. [95] described an ultrasonic hand tracking setup that may be used for maintenance activity recognition using a bicycle as an example. Wang et al. [49] have described a bare hand motion tracking system for assembly that may be extended for maintenance. A gesture-based system was proposed by Zenati-Henda et al. [66] as a method to communicate technical information by a remote user.

In general, user motion tracking methods can be categorized into three types, mechanical, optical, and inertial. There are researches focusing on other user tracking approaches, such as acoustic. These approaches have their own pros and cons [96], as shown in Table 23.8.

Mechanical tracking methods involve the use of physical gauges on a user to measure his/her motion. Despite the high accuracy that can be achieved using mechanical tracking methods, such approaches are not implemented practically and accepted in real industrial scenarios due to the high level of interference with the actual physical motion of a user. The main obstacle is users are often required to wear bulky mechanical contraptions, which have safety and practical implications.

Optical tracking approach involves the use of image processing techniques to measure the motion of a user. There are two main categories, namely, active and passive methods. Active methods involve the attachment of emitters on a user. Through the use of an array of cameras, the emitters are detected and their positions triangulated. Passive markers rely on a camera acting as both an emitter and a detection device, with markers placed on a user to reflect this signal. Passive optical methods can be based purely on the image of a user using computer vision techniques; however, the positional accuracy is generally significantly lower as compared to other optical-based approaches. Figure 23.3 illustrates the

Table 23.8 Different human motion monitoring and tracking methods

Types	Examples	Advantages	Disadvantages
Optical/computer vision	OptiTrack™ [97], leap Motion™ [98]	Highly accurate.	Subjected to lighting conditions. Require placement of cameras
Mechanical	Gypsy Suit [99]	Robust tracking data, less susceptible to outside noise	Highly intrusive form of tracking
Inertial	Perception Neuron [100]	Generally cheaper compared to optical approach	Easily subjected to magnetic field noise

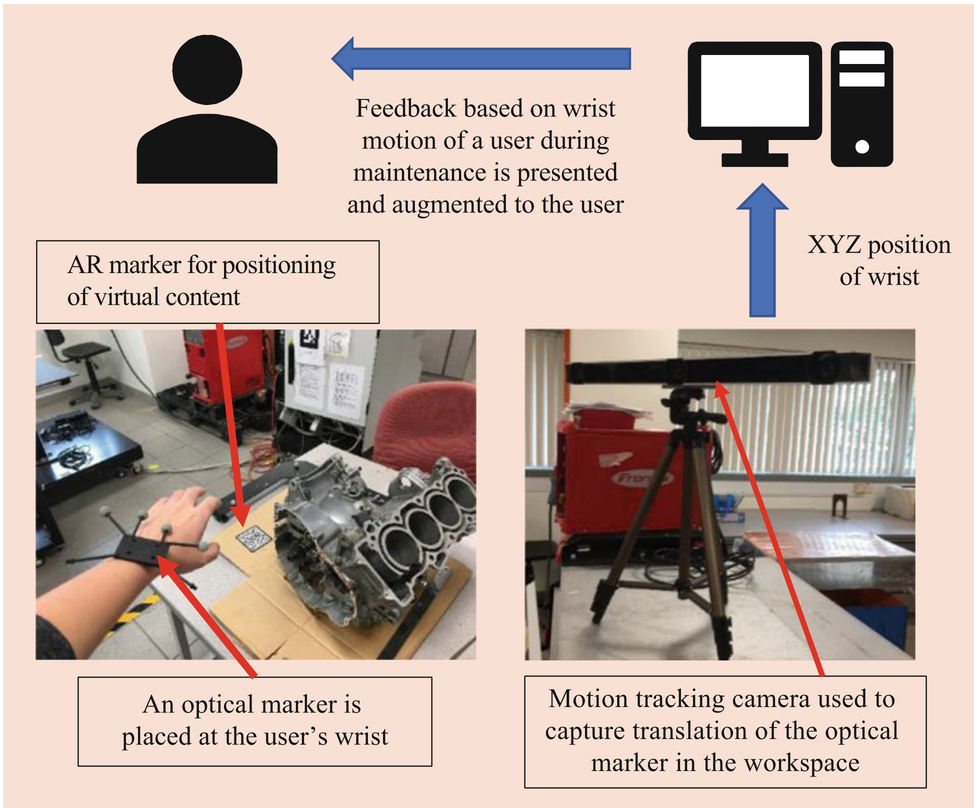


Fig. 23.3 Wrist tracking during maintenance

tracking of the wrist of a maintenance operator in an AR-based maintenance system reported by Siew et al. [79], in which sensors are attached directly onto the wrist, using optical tracking method, namely, the OptiTrack™.

The inertial approaches rely on the use of an array of accelerometers and gyroscopes to measure the pose and position of a sensor that is attached to a user. This pose information is used to estimate the pose of the user's body. Despite their lower cost, their susceptibility to magnetic interference from external devices that are prevalent in industrial settings makes inertial approaches less favorable in industrial scenarios.

23.3.5 User Interface and Interaction

This section focuses on the developments in UI and interaction methods that are related to AR in maintenance. A summary of these UI and interaction methods is presented in Fig. 23.4.

Most of the AR-assisted maintenance systems in this review use some forms of video or audio augmentation, and haptic augmentation is implemented rarely. However, in recent years, haptic augmentation has been investigated increasingly as an approach to improve human-computer interaction and applied in a variety of AR-assisted maintenance systems [101]. Webel et al. [41] suggested the integration of haptics into maintenance instruction support so that the system may be less intrusive when providing location guidance to a maintenance operator, especially when motion-related information is needed.

The development of dynamic UIs has received more attention as researchers attempt to enhance a user's experience in AR-assisted maintenance as different users may have different requirements. Currently, there are no specific guidelines

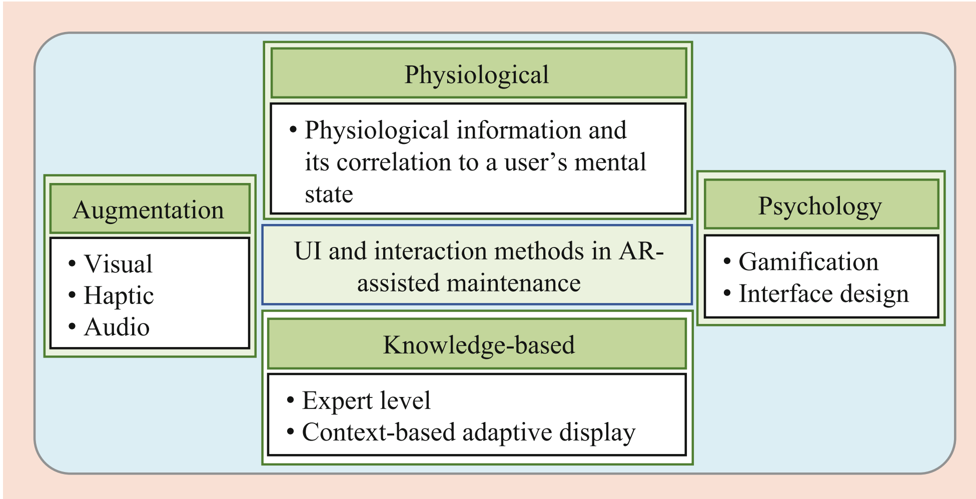


Fig. 23.4 UI and interaction methods

on the design and use of adaptive AR displays [102]. It is often assumed that users will have benefits from using augmented displays. However, there is an increase in the research on the ways to present AR contents. Bacca et al. [39] described a user-centric approach in developing AR-related content. Tsai and Huang [103] proposed the development of a system based on prior knowledge of user behavior and other external data learning approaches to display the corresponding navigation information. Oliveira et al. [62] proposed a gamified UI to make training and understanding of crucial maintenance instructions more appealing to the modern crowd. Syberfeldt et al. [47] introduced a time-based adaptive maintenance display that may be used to display appropriate content to a maintenance operator based on his/her perceived expertise level. Both Webel et al. [41] and Syberfeldt et al. [47] proposed a hard−/soft-based augmentation approach based on the expertise level of a user, where soft augmentation can also be used for non-expert users. This, in turn, will result in an increased utility and efficiency of the AR content. Siew et al. [102] developed an adaptive AR display methodology (Fig. 23.5) that can provide different levels of augmentation (Fig. 23.6) to a maintenance operator based on the expertise level of the operator, the gaze of the operator during the maintenance process, and the intention of the operator.

The UI development is often in tandem to the development in contextual awareness . A context-aware system will be able to generate appropriate content to a user automatically. Zhu et al. [73] proposed an online authoring interface supported by a contextually aware content generation system. Golparvar-Fard and Ham [104] have used automatic content generation approaches to create AR content. However, a lack of a proper design language in the UI elements may have hampered the rate of information update by a user during an actual maintenance task [105]. Many researches that have been reviewed in this study are developed based on visual and audio information.

User interaction may be improved as the tracking methodologies to detect the context of a user's actions become more mature. Westerfield et al. [106] utilized the tracking information of the markers to identify the correct placements of parts. Gesture-based interaction methods are commonly used [30, 48, 66, 107]. Human behaviors should be considered during the design of suitable UIs for a user. Although there are existing frameworks that have been developed that consider human behaviors in UI design, many of these frameworks do not account for the differences in the visualization requirements between users.

Interaction plays a key role in AR as it determines the extent of a user's ability to interact with virtual objects. The greater the number of interaction options available to a user, the better is the ability for the user to handle the information that is being presented instead of using the information passively. Various techniques and tools have been developed, such as speech [108], gaze interaction [109], and gesture interaction [30], in recent years. A summary of the different interaction methods is provided in Table 23.9. In contrast to conventional WIMP tools, such as computers and mice, the methods listed in Table 23.9 have better applicability in AR-based maintenance systems. The need for a user to conduct maintenance procedures and interact actively with the surrounding during a maintenance process coupled with safety concerns would eliminate the practicality of traditional input devices.

23.3.6 Hardware

Older researches and studies have used a desktop for AR realization [42]. Currently, mobile devices are the most

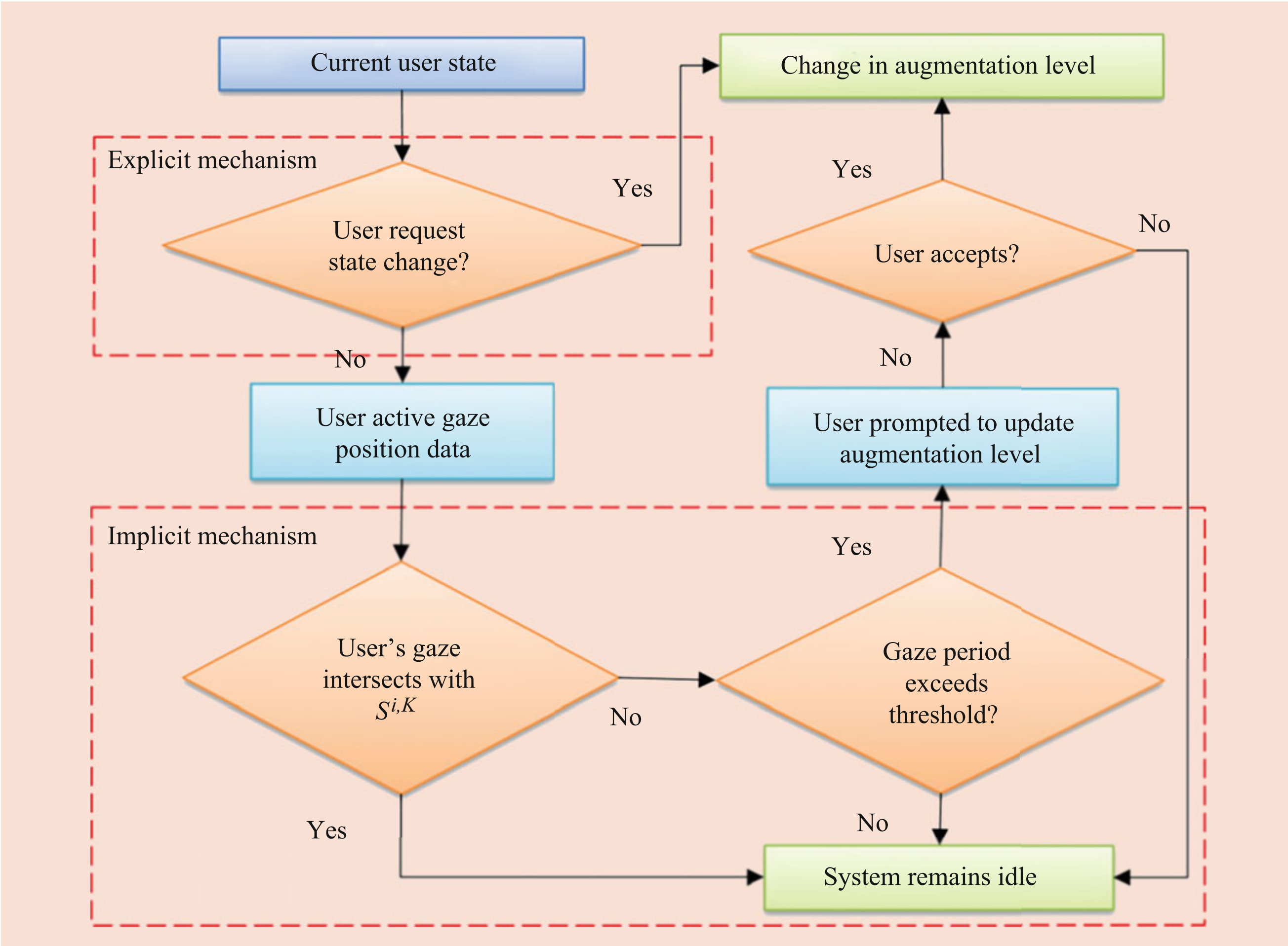


Fig. 23.5 Adaptive AR display methodology [102]

commonly implemented interface. Head-mounted devices (HMDs) are becoming more portable, lightweight, and practical for use. There is an increasing interest from commercial companies to develop HMDs, such as the HoloLens [111], for AR implementation [112]. Modern AR display technologies have shown remarkable improvements although technical challenges, such as a lack of image fidelity and motion sickness, still exist. Table 23.10 presents the advantages and disadvantages of the different AR display [113], which can be summarized as follows:

- Retinal scanning displays (RSD): The image is projected directly onto the surface of the retina [114].
- Head-mounted displays (HMDs): The image is projected directly on a screen that is mounted on the head of a user [115]; there are two types, namely, the optical see-through (OST) and video see-through (VST) HMDs.
- Handheld displays (HHDs): The augmented content is projected on a mobile screen [111].
- Spatial AR displays (SARD): The virtual content is projected either on the object itself or on a static screen [27, 116].

There has been relatively limited progress in the development of sensor-related hardware for AR applications. Few researchers have reported sensor-related hardware developments for AR in maintenance [117–119]. Microsoft Kinect has been cited as a promising low-cost tracking technology [81] that can be used in a variety of AR-assisted maintenance applications. The rise of low-cost sensors, such as the Leap Motion [120], may provide possible improvements to the application of AR in maintenance.

23.4 Discussion

Based on a thorough analysis of the existing literature, the following insights are observed.

23.4.1 Feasibility of AR Application in Industrial Maintenance

Research and development progress in tracking technologies have allowed AR to be more effectively implemented in the industry. The shift toward markerless tracking allows AR to be implemented in more industry sectors where it may not be feasible in the past [21, 33, 81]. Industrial support on the development of other enabling technologies has simplified the implementation process and allows systems' designers to concentrate on bringing the technology to the industry. While developments in contextual awareness and situational awareness have allowed the development of more robust AR-assisted maintenance applications, these systems are still unable to meet the safety requirements demanded by the industry. Technological bottlenecks still exist, such as the lack of portable user tracking technologies to detect a user's actions and motions. User tracking technologies to detect a user's actions and motions are useful in order to guide a user during a maintenance operation; however, it has received little research attention.

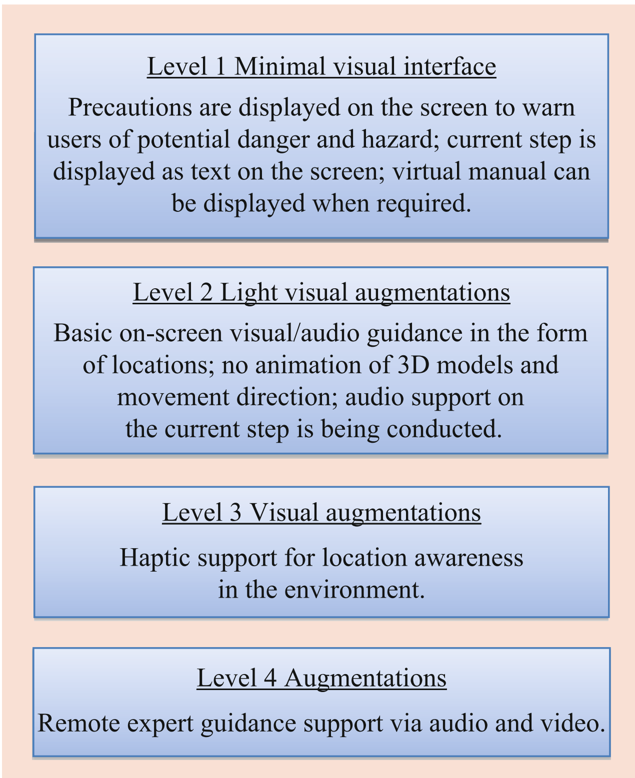


Fig. 23.6 Levels of augmentation [102]

23.4.2 Uniform Implementation Method for AR in Maintenance

There is a lack of a uniform implementation method for the design of the AR-assisted maintenance systems. Many approaches that have been reported do not result in a generic solution but use application-specific methodology to develop the AR-assisted maintenance applications. Consequently, these frameworks and implementation methods to achieve contextual awareness have not been benchmarked in terms of information transfer efficiency and safety. In addition, there is no uniformity in terms of the design language for the augmentation of information. Despite the variety of application-specific AR-assisted maintenance applications that have been reported, the interoperability of these applications is limited due to a lack of modularity during implementation as well as the non-uniform data formats adopted.

Table 23.9 Comparison of the different interaction tools

Category		Advantages	Disadvantages
Traditional devices (mouse, keyboard)		Low cost, good social acceptance, ubiquitous and already widespread applications	Restricted mobility; not suited for AR interaction
Mobile devices (touchscreen)		Lightweight, portability, ubiquitous, combinations of tracking and display, powerful computational and possessing abilities, and storage functions	Distract technicians' focus, interruption of continuous maintenance processes with the hands being engaged
Glove		Robust, intuitive, easy to use, low cost, portable, enhanced immersive feeling, accurate	Limited adjustability for altered hand sizes; not intuitive; limited user acceptability
Bare hand		Natural, intuitive, good social acceptance, cost and space reduction due to the absence of some equipment, enhanced immersive feeling [110]	Sensitive to illumination; not robust; limited application ranges (e.g., not applicable to field work due to safety issues)
Hand-free	Gaze	Free users from extra facilities; frees two hands	"Midas touch" problem gaze cannot be disengaged. Requires supplementary control action
	Speech	Free users from extra facilities; frees two hands	Poor noise immunity, poor social acceptance, possible error with differing individual speech patterns

Table 23.10 Advantages and disadvantages of the different display techniques

	HMD		HHD	SARD
	OST	VST		
Examples	HoloLens, Daqri Smart Helmet	Oculus rift, HTC vive	Mobile devices, tablets, etc.	Lightform™
Brightness	−	+	+	−
Resolution	Increasing	Increasing	Increasing	Increasing
Portability	+	+	−	+
Power consumption	−	−	+	−
Advantages	Direct view of real-world environment without passing through the screen [121]	Real-world image can be processed before being displayed [115]	Ease of use; relatively cheap	Frees user from directly using the display device, easier to set up multi-user collaboration [122]
Disadvantages	Insufficient brightness since background image cannot be modified. Possible issues of alignment of virtual objects due to difficulty in aligning with the real image	Reduces the fidelity of the real-world images. Potential safety issue due to occluded view if device fails. Possible issues of alignment of virtual objects due to difficulty in aligning with the real image	Engages at least one hand. Possible issues of alignment of virtual objects due to difficulty in aligning with the real image	Distracts user's focus; only applicable in environments requiring limited mobility and interactions

23.5 Research Gaps Based on Current Technological Developments

The following research gaps have been identified based on current technological developments.

23.5.1 A Unified and Robust Design Approach to Achieve Contextual Awareness

Cognitive load is defined as the ability of a user to comprehend the information that is presented to him/her [123]. Information display is a complex problem for the application of AR in maintenance as a variety of factors need to be considered to minimize the cognitive load experienced by the operators of the AR-assisted maintenance applications. These factors differ depending on the usage of these systems, e.g., maintenance training, guidance and support during maintenance, remote maintenance, etc. In addition, the amount and types of information to be presented to the maintenance operators by the AR systems depend based on the skill levels of these maintenance operators.

The most common augmentation is visual augmentation. Haptic and audio augmentations are alternative sources of augmentation. There is a lack of research in the integration of visual, haptic, and audio information into a single cohesive augmentation in AR-assisted maintenance applications. A unified design language was reported by Scurati et al. [105]. This unified design language is useful in consolidating the design requirements of a particular maintenance application, helping developers understand the usefulness of different presentation styles, and expediting the creation of AR-based maintenance manuals.

23.5.2 UI Design for AR in Remote Maintenance

In maintenance, AR has been implemented commonly as a maintenance guidance tool for the operators. However, there are situations where a remote operator may need to convey information to an operator on-site. Hence, AR application in remote maintenance is receiving increasing research attention. Many AR-assisted maintenance systems have been reported for tele-maintenance. Zhu et al. [75] proposed an authorable remote tele-maintenance system. Bottecchia et al. [10] have reported a remote AR-based maintenance interface which promotes collaboration between different parties over the Internet. Unfortunately, in many of these implementations, the remote parties are also often limited to the WIMP interface [73]. There is significant room for improvement in the interactions and conveyance of information between the multiple parties to maximize information transfer efficiency. Although there have been researches reported on the development of more robust remote interfaces, the WIMP interface is still the dominant interaction method for remote interaction.

23.5.3 IoT and Digital Twin Integration

In recent years , there has been an increasing shift toward the integration of the concepts of digital twin and IoT in AR-based maintenance applications [74]. Digitization and the concept of a digital twin [124] for modern machines would mean that an additional amount of information, which was previously unavailable, will be available to support human-centric maintenance. The integration of such information into existing AR-based maintenance implementations may potentially increase the information that can be used to improve the contextual awareness of these systems. This is particularly relevant in closed-loop maintenance.

23.5.4 Improvements in User Tracking

In the development of context-aware AR-based maintenance systems, information relating to the actions of a user need to be collected so that such systems can be more robust. Most AR-based systems adopt tracking methods that will allow the successful implementation of AR techniques. The availability of better user motion tracking methods may lead to the realization of more adaptive AR-based systems.

From the literature review, there are other information that can be tracked and utilized to understand the actions of a user and therefore enhance the contextual awareness of the AR-based systems. A few examples are as follows:

- Eye motion, which may be used to indicate the level of attentiveness of the user and present the correct information accordingly [125]
- Physiological data (heart rate, galvanic skin response) [126, 127]

23.5.5 Ergonomic Assessment of Current Maintenance Practices

Research in the application of AR in maintenance mainly focuses in the development of new applications. There are other research areas that can be explored, such as the development of new techniques and modules for these AR-based maintenance systems, e.g., assessment of the ergonomic health of a user.

Ergonomics is a discipline that considers the well-being of a user in the design of a workplace or job [128]. When a potential ergonomically unhealthy situation has been identified, a physical ergonomic analysis work is commonly conducted in two levels [129, 130]. As shown in Table 23.11, the first level offers a quick preliminary analysis of the ergonomic situation. The second level is a more in-depth analysis of the situation, such as a complete force analysis of the situation.

Many work-related musculoskeletal diseases (WMSD) are built up from long-term exposure to unhealthy working environments, which may not be apparent immediately if ergonomic considerations have not been taken into account during maintenance, assembly, disassembly, etc. [131]. An operator who has been working in conditions that may not be conducive to his/her physical and mental well-being for an extended period may suffer permanent body harm. The harm may consequently hinder him/her from performing his/her tasks [130]. With careful consideration and sufficient

Table 23.11 Examples of physical ergonomic assessment tools [129, 130]

Level 1	Dutch Musculoskeletal Questionnaire (DMQ)
	Rapid Upper Limb Assessment (RULA)
	Ergonomic Assessment Worksheet (EAWS)
	Scaling Experiences During Work
	Posture Checklist
Level 2	Occupational repetitive actions (OCRA)
	Strain index
	Assessment of exposure to manual patient handling MAPO (Movement and Assistance of Hospital Patients) index

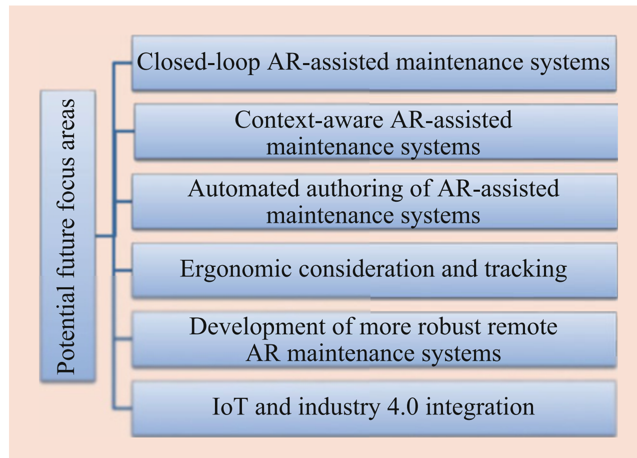


Fig. 23.7 Potential future focus areas

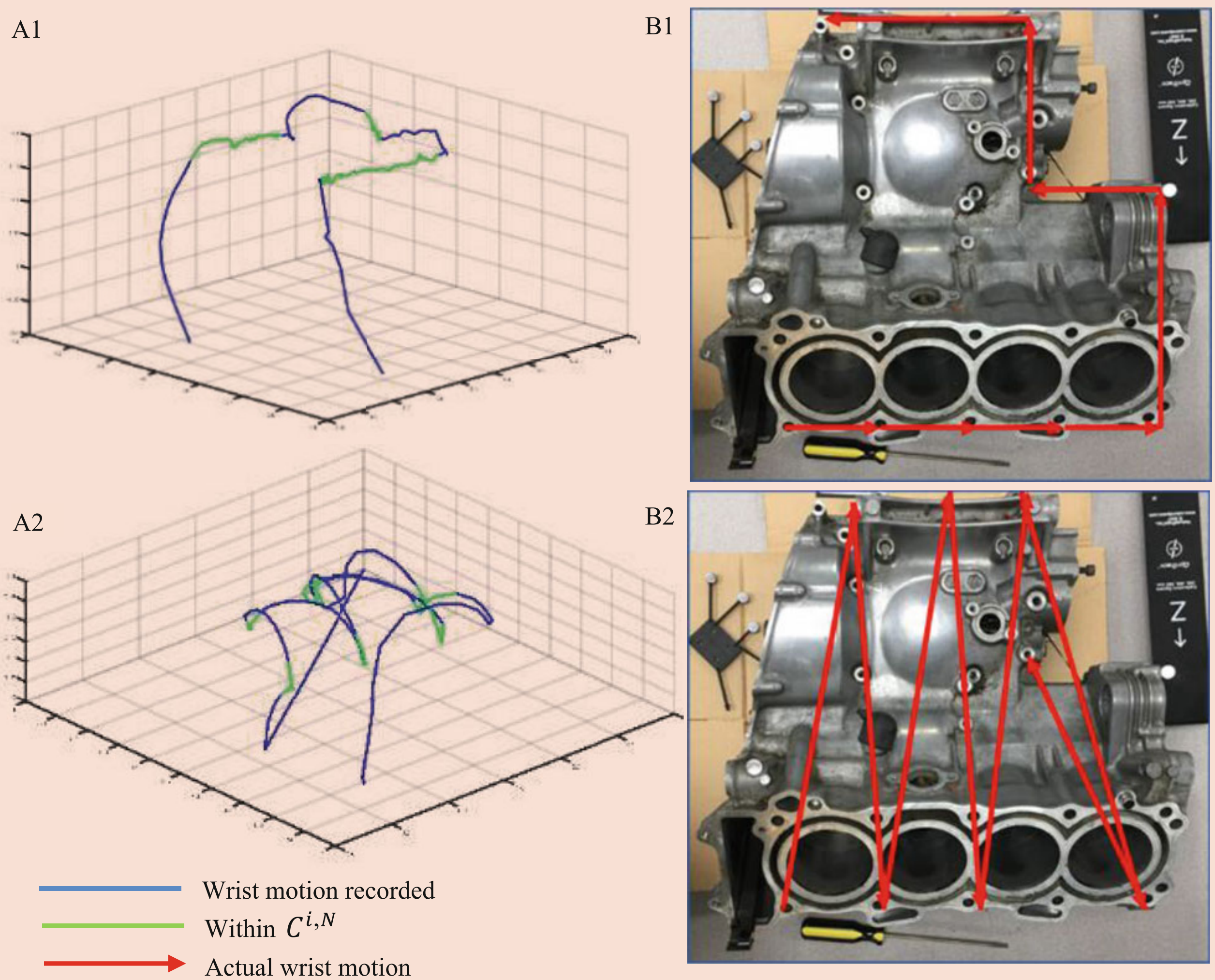
planning, all the steps in a task can be designed in an ergonomically viable way to avoid problems, such as back pain [132]. However, this may result in additional required training/retraining for operators. Thus, an AR-based module or mechanism that can monitor the ergonomics state of a user dynamically is useful, so as to potential ergonomic issues during maintenance.

23.5.6 Potential Future Research Focus Areas

With many technological advancements, the practicality of applying AR in maintenance has increased. A few potential future research areas have been identified (Fig. 23.7) and will be discussed in this section.

Closed-Loop AR-Assisted Maintenance Systems

Closed-loop maintenance systems are systems with a feedback loop to allow the systems to provide feedback to the users during the maintenance process. Most traditional AR-based maintenance systems can be considered as open-loop systems, in which information is only provided to a user as he/she performs a step in the maintenance process, and no information is provided to the user after he/she has completed the step with regard to, for example, the correctness of the



	A1	A2
Time (s)	43	50
Distance (m)	2.38	4.79
Within expected regions (line colour)	Yes	Yes

Fig. 23.8 User performance [79]

step or the maintenance quality. An open-loop maintenance process is not desirable, and AR-based maintenance systems that are open-loop do not utilize fully the advantages that the AR technology provides. Maintenance quality can be defined based on a wide variety of factors. A common definition of maintenance quality is the correctness of the steps that have been conducted for a maintenance job with respect to time used to perform these steps. Siew et al. [79] implemented a methodology to assess the correctness of the maintenance steps performed by an operator through tracking the wrist motion of the operator; guidance can be provided to the operator when the system detects that the wrist motion of the operator has deviated frequently from the expected movements (Fig. 23.8). A closed-loop AR-based maintenance system can improve maintenance quality as such a system can enhance a user's understanding of the information augmented. Development of closed-loop AR-based maintenance systems can provide more effective guidance and feedback to a user as and when required.

Context Awareness of AR-Based Maintenance Systems

Many AR-based maintenance applications utilize static information, i.e., information is retrieved directly from a database and presented to a user without any consideration of the user's current action, skill level, etc. This approach of presenting information with limited consideration of the context of a user decreases the effectiveness of the information presented [103]. During a maintenance process, an operator will typically move toward the maintenance target to perform the maintenance procedures; he/she may move around the shop floor to acquire the necessary tools, mull over, assess the next procedure to perform, and rest after performing a procedure [133]. Thus, there is a need to consider the users' context, states, and actions in order to apply AR successfully to a variety of industrial maintenance issues. In an ideal AR-assisted maintenance system, the system is able to understand the context and states of a user to support the actions of the user. The user tracking module provides information of the actions of the user, and this information enhances the contextual awareness of the system. In this way, the UI and interaction module will be able to provide useful and relevant information and interactivity to the user. Siew et al. [133] have formulated an operator behavior model (Fig. 23.9) that provides a contextual awareness approach to achieve an adaptive AR-assisted maintenance system.

Automated Authoring of AR-Assisted Maintenance Systems

One of the major problems in the AR-assisted maintenance systems is that the authoring of the systems is not automated. Reported systems have developed and used a schema to categorize and translate the content to be used for the AR systems. However, these approaches typically still require user intervention for the creation of the related content, which represents a significant amount of time in the implementation process. To simplify the authoring process, a methodology in which this can be done automatically is thus desirable. This may be implemented based on the identification of the CAD models and parts presented.

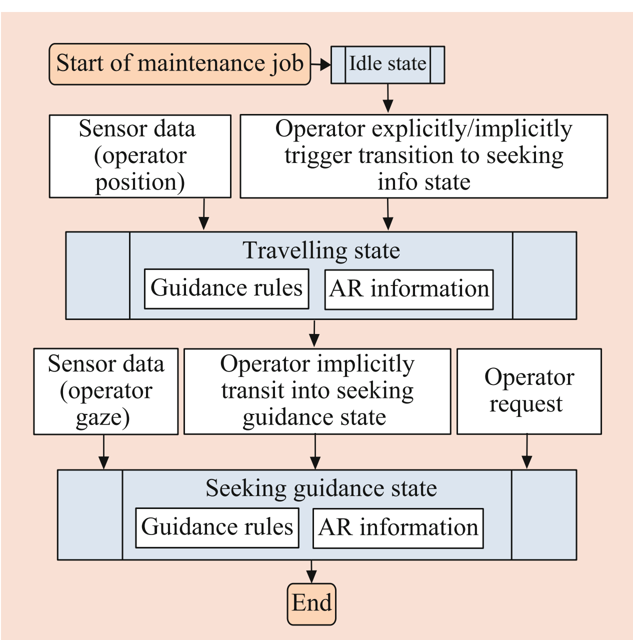


Fig. 23.9 Operator behavior model [133]

Ergonomic Consideration and Tracking

The ergonomics of the maintenance operators is an important consideration in industrial maintenance environments. A few ergonomic concepts have been implemented for the monitoring of the postures of the operators in some of the reported AR-assisted maintenance systems. However, ergonomic considerations must include other possible causes to make the evaluation more holistic so as to increase workplace safety. For example, the vibration from hand tools may result in long-term hand-related ergonomic issues. An analysis of the hand ergonomics by the tracking of the hand motion of the operators is another area that still requires substantial work.

Development of More Robust Remote AR Maintenance Systems

Currently, most of the reported systems focus on the current on-site operators. However, there has been an increased interest in providing remote access support to AR maintenance systems. While AR aims to act as a guidance tool, there are situations where a remote user needs to be able to convey information to the operator on-site. In this current literature review, a variety of AR systems have been considered to aid in the field of tele-maintenance. Zhu et al. [9] proposed an authorable remote tele-maintenance system. Bottecchia et al. [10] have also proposed a remote AR maintenance interface which promotes collaboration between different parties over the Internet using AR. Unfortunately, many of the implementations suffer from a lack of unified design language. Remote parties are also often limited to the WIMP interface as well [73]. There are significant rooms to improve the interaction and conveying of information between the two parties to maximize information transfer efficiency. Although there are studies indicating a move toward achieving a more robust remote interface, as discussed, the WIMP interface is still the dominant interaction method for remote interaction. It is foreseeable that via the utilization of a tracking methodology for the remote user, it is possible to improve the robustness of remote instructions.

IoT and Industry 4.0 Integration

Currently, there is a trend toward the integration of Big Data, digital twin (DT), and IoT in AR in maintenance applications [74]. A number of systems have been reported that attempted to integrate AR with IoT. The increasing digitization, the advancements in sensor technologies, and the DT concept for modern machines and equipment would mean that a large amount of data and information previously unavailable can now be obtained to guide and support the operators in preventive and corrective maintenance processes. Many systems that reported the utilization of the sensors and IoT information are still limited to proof-of-concept research. Integrating the information into existing AR maintenance implementation may have the potential to further increase the information that can be used to improve the contextual awareness of the system , such as system states, or tapping into the DT information of the maintenance tasks.

23.6 Conclusion

This chapter has presented a holistic review of the research on AR application in maintenance. A description of the current state of AR application in maintenance is first presented, followed by a description of the current researches that have been reported. From the review, it can be observed that many of the researches on AR in maintenance have been carried out independently, resulting in a large variety of techniques and applications. There are no unified methods and frameworks in the reported researches, for example, in context awareness-related developments. Currently, there are difficulties in implementing these AR-assisted maintenance systems to the industry and satisfying the regulatory requirements of industrial applications. In tracking research, there has been an increased trend toward markerless tracking and human motion tracking. However, the physiological changes that a user may experience during maintenance have not been addressed adequately in AR in maintenance research. In addition, a few reported researches consider the effectiveness of the information that is presented to a user. Thus, a successful AR-based maintenance implementation would need to consider human requirements in the design and development of such a system. Developments of AR-based maintenance systems should adopt a human-centric approach to ensure successful system deployment and practical user experience.

References

1. Higgins, L.R., Mobley, R.K., Smith, R.: Maintenance Engineering Handbook. McGraw-Hill Professional, New York (2002)
2. Lee, J., Holgado, M., Kao, H.A., Macchi, M.: New thinking paradigm for maintenance innovation design. IFAC-PapersOnline. **19**, 7104–7109 (2014)
3. Henderson, S.J., Feiner, S.: Evaluating the benefits of augmented reality for task localization in maintenance of an Armored Personnel Carrier Turret. In: Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality, Orlando, FL, pp. 135–144 (2009)
4. Nee, A.Y.C., Ong, S.K., Chryssolouris, G., Mourtzis, D.: Augmented reality applications in design and manufacturing. Ann. CIRP. **61**, 657–679 (2012)
5. Azuma, R.: A survey of augmented reality. Presence Teleop. Virt. **6**, 355–385 (1997)
6. Henderson, S.J., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE Trans. Vis. Comput. Graph. **17**(10), 1355–1368 (2011)
7. De Crescenzio, F., Fantini, M.: Implementing augmented reality to aircraft maintenance: A daily inspection case study. In: Proceedings of the IDMME Virtual Concept Conference, Bordeaux, France, pp. 1–5 (2010)
8. Dalle, M., Roy, R., Dini, G.: Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality. Ann. CIRP. **66**, 465–468 (2017)
9. Zhu, J., Ong, S.K., Nee, A.Y.C.: A context-aware augmented reality system to assist the maintenance operators. Int. J. Interact. Des. Manuf. **8**, 293–304 (2014)
10. Bottecchia, S., Cieutat, J.M., Jessel, J.P.: T.A.C: Augmented reality system for collaborative tele-assistance in the field of maintenance through internet. In: Proceedings of the 1st International Conference on Augmented Human, Megeve, France, pp. 1–14 (2010)
11. De Crescenzio, F., Fantini, M., Persiani, F., Di Stefano, L., Salti, S., Azzari, P.: Augmented reality for aircraft maintenance training and operations support. IEEE Comput. Graph. Appl. **31**, 96–101 (2010)
12. Perdikakis, A., Araya, A., Kiritsis, D.: Introducing augmented reality in next generation industrial learning tools: a case study on electric and hybrid vehicles. Procedia Eng. **132**, 251–258 (2015)
13. Caudell, T.P., Mizell, D.W.: Augmented reality: An application of heads-up display technology to manual manufacturing processes. In: Proceedings of the 25th International Conference on System Sciences, Hawaii, pp. 659–669 (1992)
14. Dünser, A., Grasset, R., Seichter, H., Billinghurst, M.: Applying HCI principles to AR systems design. In: Proceedings of the 2nd International Workshop on Mixed Reality User Interfaces: Specification, Authoring, Adaptation, Charlotte, NC, pp. 37–42 (2007)
15. Grabowski, M., Rowen, A., Rancy, J.P.: Evaluation of wearable immersive augmented reality technology in safety-critical systems. Saf. Sci. **103**, 23–32 (2018)
16. Lamberti, F., Manuri, F., Sanna, A., Paravati, G., Pezzolla, P., Montuschi, P.: Challenges, opportunities, and future trends of emerging techniques for augmented reality-based maintenance. IEEE Trans. Emerg. Top. Comput. **2**, 411–421 (2014)
17. Martinetti, A., Rajabalinejad, M., Van Dongen, L.: Shaping the future maintenance operations: reflections on the adoptions of augmented reality through problems and opportunities. Procedia CIRP. **59**, 14–17 (2017)
18. Porcelli, I., Rapaccini, M., Espíndola, D.B., Pereira, C.E.: Technical and organizational issues about the introduction of augmented reality in maintenance and technical assistance services. IFAC Proc. Vols. **46**(7), 257–262 (2013)
19. Dini, G., Mura, M.D.: Application of augmented reality techniques in through-life engineering services. Procedia CIRP. **38**, 14–23 (2015)
20. Borsci, S., Lawson, G., Broome, S.: Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance. Comput. Ind. **67**, 17–26 (2015)
21. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. Robot. Comput. Integr. Manuf. **49**, 215–228 (2018)
22. Friedrich, W., Jahn, D., Schmidt, L.: ARVIKA: Augmented reality for development, production and service. In: Proceedings of the 1st International Symposium on Mixed and Augmented Reality, Darmstadt, Germany, pp. 3–4 (2002)
23. Weidenhausen, J., Knoepfle, C., Stricker, D.: Lessons learned on the way to industrial augmented reality applications. A retrospective on ARVIKA. Comput. Graph. **27**, 887–891 (2003)

23

24. De Groote, P.: Maintenance performance analysis: a practical approach. J. Qual. Maint. Eng. **1**, 4–24 (1995)
25. Regenbrecht, H.T., Specht, R.: A mobile passive augmented reality device – mPARD. In: Proceedings of the IEEE and ACM International Symposium on Augmented Reality, Munich, 2000, pp. 81–84 (2000)
26. García, A.A., Bobadilla, I.G., Figueroa, G.A., Ramírez, M.P., Román, J.M.: Virtual reality training system for maintenance and operation of high-voltage overhead power lines. Virtual Reality. **20**, 27–40 (2016)
27. Fiorentino, M., Uva, A.E., Gattullo, M., Debernardis, S., Monno, G.: Augmented reality on large screen for interactive maintenance instructions. Comput. Ind. **65**, 270–278 (2014)
28. Neges, M., Wolf, M., Abramovici, M.: Secure access augmented reality solution for Mobile maintenance support utilizing condition-oriented work instructions. Procedia CIRP. **38**, 58–62 (2015)
29. Syberfeldt, A., Danielsson, O., Holm, M., Wang, L.: Dynamic operator instructions based on augmented reality and expert systems. Procedia CIRP. **41**, 346–351 (2016)
30. Shim, J., Yang, Y., Kang, N., Seo, J., Han, T.D.: Gesture-based interactive augmented reality content authoring system using HMD. Virtual Reality. **20**, 57–69 (2016)
31. Kim, K., Billinghurst, M., Bruder, G., Duh, H.B.L., Welch, G.F.: Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008–2017). IEEE Trans. Vis. Comput. Graph. **24**, 2947–2962 (2018)
32. Zhu, J., Ong, S.K., Nee, A.Y.C.: A context-aware augmented reality assisted maintenance system. Int. J. Comput. Integr. Manuf. **28**(2), 213–225 (2015)
33. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graph. Appl. **21**(6), 34–47 (2001)
34. Gattullo, M., Scurati, G.W., Fiorentino, M., Uva, A.E., Ferrise, F., Bordegoni, M.: Towards augmented reality manuals for industry 4.0: a methodology. Robot. Comput. Integr. Manuf. **56**, 276–286 (2019)
35. Reardon, C., Lee, K., Rogers, J.G., Fink, J.: Augmented reality for human-robot teaming in field environments. In: Proceedings of the International Conference on Human-Computer Interaction, Orlando, pp. 79–92 (2019)
36. Castellanos, M.J., Navarro-Newball, A.A.: Prototyping an augmented reality maintenance and repairing system for a deep well vertical turbine pump. In: Proceedings of the International Conference on Electronics, Communications and Computers, Cholula, Mexico, pp. 36–40 (2019)
37. Wang, X., Ong, S.K., Nee, A.Y.C.: A comprehensive survey of ubiquitous manufacturing research. Int. J. Prod. Res. **56**, 604–628 (2019). (2018)
38. Haritos, T., Macchiarella, N.D.: A mobile application of augmented reality for aerospace maintenance training. In: Proceedings of the AIAA/IEEE Digital Avionics Systems Conference, Washington, DC, pp. 5.B.3–5.1 (2005)
39. Bacca, J., Baldiris, S., Fabregat, R., Kinshuk, Graf, S.: Mobile augmented reality in vocational education and training. Procedia Comput. Sci. **75**, 49–58 (2015)
40. Nakajima, C., Itho, N.: A support system for maintenance training by augmented reality. In: Proceedings of the 12th International Conference on Image Analysis and Processing, Mantova, Italy, pp. 158–163 (2003)
41. Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., Preusche, C.: An augmented reality training platform for assembly and maintenance skills. Robot. Auton. Syst. **61**, 398–403 (2013)
42. Macchiarella, N.D., Vincenzi, D.A.: Augmented reality in a learning paradigm for flight aerospace maintenance training. In: Proceedings of the 24th Digital Avionics Systems Conference, Avionics in Changing Market Place: Safe and Secure, Salt Lake City, Washington, DC, pp. 5.D.1–5.1-9 (2005)
43. Schwald, B., DeLaval, B.: An augmented reality system for training and assistance to maintenance in the industrial context. In: Proceedings of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Czech Republic, pp. 425–432 (2003)
44. Segovia, D., Ramírez, H., Mendoza, M., Mendoza, M., Mendoza, E., González, E.: Machining and dimensional validation training using augmented reality for a Lean process. Procedia Comput. Sci. **75**, 195–204 (2015)
45. Jo, G.S., Oh, K.J., Ha, I., Lee, K.S., Hong, M.D., Neumann, U., You, S.: A unified framework for augmented reality and knowledge-based systems in maintaining aircraft. In: Proceedings of the National Conference on Artificial Intelligence, Quebec City, pp. 2990–2997 (2014)
46. Neges, M., Adwernat, S., Abramovici, M.: Augmented Virtuality for maintenance training simulation under various stress conditions. Procedia Manuf. **19**, 171–178 (2018)
47. Syberfeldt, A., Holm, M., Danielsson, O., Wang, L., Brewster, R.L.: Support systems on the industrial shop-floors of the future: Operators' perspective on augmented reality. In: Proceedings of the 6th CIRP Conference on Assembly Technologies and Systems, Gothenburg, pp. 108–113 (2016)
48. Abdelnaby, M., Abd Elazem, M., Aly, H.A., Kaboudan, A.: Augmented reality maintenance training with intel depth camera. In: Proceedings of the International Conference on Machine Vision and Information Technology, Los Alamitos, pp. 116–122 (2017)
49. Wang, X., Ong, S.K., Nee, A.Y.C.: Real-virtual components interaction for assembly simulation and planning. Robot. Comput. Integr. Manuf. **41**, 102–114 (2016)
50. Toro, C., Sanin, C., Vaquero, J., Posada, J., Szczerbicki, E.: Knowledge based industrial maintenance using portable devices and augmented reality. In: Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Vietri sul Mare, Italy, pp. 295–302 (2007)
51. Nakagawa, T., Sano, T., Nakatani, Y.: Plant maintenance support system by augmented reality. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Tokyo, Japan, pp. I-768–I-773 (1999)
52. Lee, S., Akin, Ö.: Augmented reality-based computational fieldwork support for equipment operations and maintenance. Autom. Constr. **20**, 338–352 (2011)
53. Savioja, O., Järvinen, P., Karhela, T., Siltanen, P., Woodward, C.: Developing a mobile, service-based augmented reality tool for modern maintenance work. In: Proceedings of the International Conference on Virtual Reality, Beijing, China, pp. 554–563 (2007)
54. Ke, C., Kang, B., Chen, D., Li, X.: An augmented reality-based application for equipment maintenance. In: Proceedings of the First International Conference on Affective Computing and Intelligent Interaction, Beijing, China, pp. 836–841 (2005)
55. Schlagowski, R., Merkel, L., Meitinger, C.: Design of an assistant system for industrial maintenance tasks and implementation of a prototype using augmented reality. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, pp. 294–298 (2017)
56. Zenati, N., Zerhouni, N., Achour, K.: Assistance to maintenance in industrial process using an augmented reality system. In: Proceedings of the IEEE International Conference on Industrial Technology, Hammamet, Tunisia, pp. 848–852 (2004)
57. Benbelkacem, S., Belhocine, M., Bellarbi, A., Zenati-Henda, N., Tadjine, M.: Augmented reality for photovoltaic pumping systems maintenance tasks. Renew. Energy. **55**, 428–437 (2013)

58. Tatić, D., Tešić, B.: The application of augmented reality technologies for the improvement of occupational safety in an industrial environment. Comput. Ind. **85**, 1–10 (2017)
59. Garza, L.E., Pantoja, G., Ramírez, P., Ramírez, H., Rodríguez, N., González, E., Quintal, R., Pérez, J.A.: Augmented reality application for the maintenance of a Flapper valve of a Fuller-Kynion type M pump. Procedia Comput. Sci. **25**, 154–160 (2013)
60. Mooser, J., Wang, L., You, S., Neumann, U.: An augmented reality interface for mobile information retrieval. In: Proceedings of the IEEE International Conference on Multimedia and Expo, Beijing, pp. 2226–2229 (2007)
61. Behringer, R., Chen, S., Sundareswaran, V., Wang, K., Vassiliou, M.: Novel interface for device diagnostics using speech recognition, augmented reality visualization and 3D audio auralization. In: Proceedings of the International Conference on Multimedia Computing and Systems, Florence, Italy, pp. 427–432 (1999)
62. Oliveira, A., Caetano, N., Botega, L.C., de Araújo, R.B.: A head-up display with augmented reality and gamification for an e-maintenance system: using interfaces and gamification to motivate workers in procedural tasks. In: Proceedings of the International Conference on Human Interface and the Management of Information, Los Angeles, 2015, pp. 499–510 (2015)
63. Aschenbrenner, D., Maltry, N., Kimmel, J., Albert, M., Scharnagl, J., Schilling, K.: ARTab – Using virtual and augmented reality methods for an improved situation awareness for telemaintenance. IFAC-PapersOnLine. **49**(30), 204–209 (2016)
64. Aschenbrenner, D., Latoschik, M.E., Schilling, K.: Industrial maintenance with augmented reality: Two case studies. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Munich, pp. 341–342 (2016)
65. Abramovici, M., Wolf, M., Adwernat, S., Neges, M.: Context-aware maintenance support for augmented reality assistance and synchronous multi-user collaboration. Procedia CIRP. **59**, 18–22 (2016)
66. Zenati-Henda, N., Bellarbi, A., Benbelkacem, S., Belhocine, M.: Augmented reality system based on hand gestures for remote maintenance. In: Proceedings of the International Conference on Multimedia Computing and Systems, Marrakech, Morocco, pp. 5–8 (2014)
67. Regenbrecht, H., Baratoff, G., Wilke, W.: Augmented reality projects in the automotive and aerospace industries. IEEE Comput. Graph. Appl. **25**(6), 48–56 (2005)
68. Wang, W., Qi, Y., Wang, Q.: An augmented reality application framework for complex equipment collaborative maintenance. In: Proceedings of the 8th International Conference on Cooperative Design, Visualization, and Engineering, Berlin, 2011, pp. 154–161 (2011)
69. Petersen, N., Stricker, D.: Cognitive augmented reality. Comput. Graph. **53**, 82–91 (2015)
70. Chimienti, V., Iliano, S., Dassisti, M., Dini, G., Failli, F.: Guidelines for implementing augmented reality procedures in assisting assembly operations. In: Proceedings of the International Precision Assembly Seminar, Chamonix, pp. 174–179 (2010)
71. Macchiarella, N.D.: Augmented reality as a medium for job task training. J. Instr. Deliv. Syst. **19**(1), 21–24 (2005)
72. Greenhalgh, P., Mullins, B., Grunnet-Jepsen, A., Bhowmik, A.K.: Industrial deployment of a full-featured head-mounted augmented-reality system and the incorporation of a 3D-sensing platform. In: Proceedings of the Society for Information Display Symposium Digest of Technical Papers, San Francisco, pp. 448–451 (2016)
73. Zhu, J., Nee, A.Y.C., Ong, S.K.: Online authoring for augmented reality remote maintenance. In: Proceedings of the 12th IASTED International Conference on Computer Graphics and Imaging, Innsbruck, Austria, pp. 87–94 (2011)
74. Masoni, R., Ferrise, F., Bordegoni, M., Gattullo, M., Uva, A.E., Fiorentino, M., Carrabba, E., Di Donato, M.: Supporting remote maintenance in industry 4.0 through augmented reality. Procedia Manuf. **11**, 1296–1302 (2017)
75. Zhu, J., Ong, S.K., Nee, A.Y.C.: An Authorable context-aware augmented reality system to assist the maintenance technicians. Int. J. Adv. Manuf. Technol. **66**, 1699–1714 (2013)
76. del Amo, I.F., Erkoyuncu, J.A., Roy, R., Wilding, S.: Augmented reality in maintenance: an information-centred design framework. Procedia Manuf. **19**, 148–155 (2018)
77. Bauer, M., Bruegge, B., Klinker, G., MacWilliams, A., Reicher, T., Riß, S., Sandor, C., Wagner, M.: Design of a component-based augmented reality framework. In: Proceedings of the IEEE and ACM International Symposium on Augmented Reality, New York, pp. 45–54 (2001)
78. Benbelkacem, S., Zenati-Henda, N., Belhocine, M., Malek, S.: Augmented reality system for e-maintenance application. In: Proceedings of the 2nd Mediterranean Conference on Intelligent Systems and Automation, Zarzis, Tunisia, pp. 185–189 (2009)
79. Siew, C.Y., Ong, S.K., Nee, A.Y.C.: A practical augmented reality-assisted maintenance system framework for adaptive user support. Robot. Comput. Integr. Manuf. **59**, 115–129 (2019)
80. Siltanen, S.: Theory and Applications of Marker-Based Augmented Reality. VTT Technical Research Centre of Finland, Espoo (2012)
81. Wang, X., Ong, S.K., Nee, A.Y.C.: A comprehensive survey of augmented reality assembly research. Adv. Manuf. **4**, 1–22 (2016)
82. Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, Francisco, pp. 85–94 (1999)
83. Wang, R., Paris, S., Popović, J.: 6D hands: Markerless hand-tracking for computer aided design. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, pp. 549–558 (2011)
84. Lee, T., Höllerer, T.: Hybrid feature tracking and user interaction for markerless augmented reality. In: Proceedings of the IEEE International Symposium on Virtual Reality, Reno, pp. 145–152 (2008)
85. Yu, L., Ong, S.K., Nee, A.Y.C.: A tracking solution for mobile augmented reality based on sensor-aided marker-less tracking and panoramic mapping. Multimed. Tools Appl. **75**, 3199–3220 (2016)
86. Manuri, F., Pizzigalli, A., Sanna, A.: A state validation system for augmented reality based maintenance procedures. Appl. Sci. **9**(10), 2115 (2019)
87. Lima, J.P., Roberto, R., Simões, F., Almeida, M., Figueiredo, L., Teixeira, J.M., Teichrieb, V.: Markerless tracking system for augmented reality in the automotive industry. Expert Syst. Appl. **82**, 100–114 (2017)
88. Li, J., Fan, X.: Outdoor augmented reality tracking using 3D city models and game engine. In: Proceedings of the 7th International Congress on Image and Signal Processing, Dalian, China, pp. 104–108 (2014)
89. Zhou, F., Duh, H.B.L., Billinghurst, M.: Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, Cambridge, pp. 193–202 (2008)
90. PTC. Vuforia. https://www.vuforia.com/ (2018)
91. Google. ARCore. https://developers.google.com/ar/discover/ (2018)
92. Apple. ARKit. https://developer.apple.com/arkit/ (2018)
93. Flatt, H., Koch, N., Röcker, C., Günter, A., Jasperneite, J.: A context-aware assistance system for maintenance applications in smart factories based on augmented reality and indoor

localization. In: Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation, Luxembourg, pp. 1–4 (2015)
94. Paz, A., Guenaga, M.L., Eguiluz, A.: Augmented reality for maintenance operator training using surf points and homography. In: Proceedings of the 9th International Conference on Remote Engineering and Virtual Instrumentation, Bilbao, Spain, pp. 1–4 (2012)
95. Stiefmeier, T., Ogris, G., Junker, H., Lukowicz, P., Troster, G.: Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario. In: Proceedings of the 10th IEEE International Symposium on Wearable Computers, Montreux, pp. 97–104 (2006)
96. Welch, G., Foxlin, E.: Motion tracking: no silver bullet, but a respectable arsenal. IEEE Comput. Graph. Appl. **22**, 24–38 (2002)
97. Natural Point: OptiTrack – Motion capture systems. http://optitrack.com/
98. Leap Motion Inc.: Leap motion. https://www.leapmotion.com/
99. Meta Motion: Gypsy 7 electro-mechanical motion capture system. http://metamotion.com/gypsy/gypsy-motion-capture-system.htm
100. Noitom. Perception neuron by Noitom | Perception neuron motion capture for virtual reality, animation, sports, gaming and film. https://neuronmocap.com/
101. Boessenkool, H., Thomas, J., Wildenbeest, J.G.W., Heemskerk, C.J.M., de Baar, M.R., Steinbuch, M., Abbink, D.A.: Where to improve in human-in-the-loop tele-operated maintenance? A phased task analysis based on video data of maintenance at JET. Fusion Eng. Des. **129**, 309–319 (2018)
102. Siew, C.Y., Nee, A.Y.C., Ong, S.K.: Improving maintenance efficiency with an adaptive AR-assisted maintenance system. In: Proceedings of the 4th International Conference on Robotics, Control and Automation, Guangzhou, China, pp. 74–78 (2019)
103. Tsai, C.H., Huang, J.Y.: Augmented reality display based on user behaviour. Comput. Stand. Interf. **55**, 171–181 (2018)
104. Golparvar-Fard, M., Ham, Y.: Automated diagnostics and visualization of potential energy performance problems in existing buildings using energy performance augmented reality models. J. Comput. Civ. Eng. **28**, 17–29 (2013)
105. Scurati, G.W., Gattullo, M., Fiorentino, M., Ferrise, F., Bordegoni, M., Uva, A.E.: Converting maintenance actions into standard symbols for augmented reality applications in industry 4.0. Comput. Ind. **98**, 68–79 (2018)
106. Westerfield, G., Mitrovic, A., Billinghurst, M.: Intelligent augmented reality training for motherboard assembly. Int. J. Artif. Intell. Educ. **25**, 157–172 (2015)
107. Fiorentino, M., Uva, A.E., Monno, G., Radkowski, R.: Natural interaction for online documentation in industrial maintenance. Int. J. Comput. Aided Eng. Technol. **8**(1/2), 56–79 (2016)
108. Goose, S., Sudarsky, S., Zhang, X., Navab, N.: Speech-enabled augmented reality supporting mobile industrial maintenance. IEEE Pervasive Comput. **2**(1), 65–70 (2003)
109. Park, H.M., Lee, S.H., Choi, J.S.: Wearable augmented reality system using gaze interaction. In: Proceedings of the 7th IEEE International Symposium on Mixed and Augmented Reality, Cambridge, pp. 175–176 (2008)
110. Wang, X., Ong, S.K., Nee, A.Y.C.: Real-virtual interaction in AR assembly simulation based on component contact handling strategy. Assem. Autom. **35**(4), 376–394 (2015)
111. Lee, A.S., Swift, M., Tang, J.C., Chen, H.: 3D collaboration method over HoloLens and Skype endpoints. In: Proceedings of the 3rd International Workshop on Immersive Media Experiences, Brisbane, pp. 27–30 (2015)
112. Pringle, A., Hutka, S., Mom, J., van Esch, R., Heffernan, N., Chen, P.: Ethnographic study of a commercially available augmented reality HMD App for industry work instruction. In: Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments, Rhodes, pp. 389–397 (2019)
113. van Krevelen, R., Poelman, R.: A survey of augmented reality technologies, applications and limitations. Int. J. Virtual Real. **9**, 1–20 (2010)
114. Johnston, R.S., Willey, S.R.: Development of a commercial retinal scanning display. In: Proceedings of the Symposium on OE/Aerospace Sensing and Dual Use Photonics, Orlando, pp. 2–14 (1995)
115. Kiyokawa, K.: Trends and vision of head mounted display in augmented reality. In: International Symposium on Ubiquitous Virtual Reality, Daejeon, South Korea, pp. 14–17 (2012)
116. Factura, B., LaPerche, L., Reyneri, P., Jones, B., Karsch, K.: Lightform: Procedural effects for projected AR. In: Proceedings of the SIGGRAPH Special Interest Group on Computer Graphics and Interactive Techniques Conference, Vancouver, BC, pp. 6-1–6-2 (2018)
117. Lakshmprabha, N.S., Kasderidis, S., Mousouliotis, P., Petrou, L., Beltramello, O.: Augmented reality for maintenance application on a mobile platform. In: Proceedings of the IEEE Annual International Symposium on Virtual Reality, Arles, Camargue-Provence, pp. 355–356 (2015)
118. Fang, W., Zheng, L., Wu, X.: Multi-sensor based real-time 6-Dof pose tracking for wearable augmented reality. Comput. Ind. **92–93**, 91–103 (2017)
119. Herold, R., Vogel, U., Richter, B., Kreye, D., Reckziegel, S., Scholles, M., Lakner, H.: OLED-on-CMOS integration for augmented-reality systems. In: Proceedings of the International Students and Young Scientists Workshop – Photonics and Microsystems, Wroclaw-Szklarska Poreba, Poland, pp. 19–22 (2008)
120. Lv, Z., Halawani, A., Feng, S., Ur Réhman, S., Li, H.: Touch-less interactive augmented reality game on vision-based wearable device. Pers. Ubiquit. Comput. **19**, 551–567 (2015)
121. Haller, M., Thomas, B.H., Billinghurst, M.: Emerging Technologies of Augmented Reality. IGI Global, Hershey (2007)
122. Ohshima, T., Satoh, K., Yamamoto, H., Tamura, H.: RV-border guards: a multi-player mixed reality entertainment. Trans. Virtual Real. Soc. Jpn. **4**(4), 699–705 (1999)
123. Sweller, J.: Cognitive load during problem solving: effects on learning. Cogn. Sci. **12**(2), 257–285 (1988)
124. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F.: Digital twin-driven product design, manufacturing and service with big data. Int. J. Adv. Manuf. Technol. **94**, 3563–3576 (2018)
125. Faro, A., Giordano, D., Pino, C., Spampinato, C.: Visual attention for implicit relevance feedback in a content based image retrieval. In: Proceedings of the Symposium on Eye-Tracking Research & Applications, Austin, pp. 73–76 (2010)
126. Thayer, J.F., Åhs, F., Fredrikson, M., Sollers, J.J., Wager, T.D.: A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci. Biobehav. Rev. **36**(2), 747–756 (2012)
127. Tattersall, A.J., Hockey, G.R.: Level of operator control and changes in heart rate variability during simulated flight maintenance. Hum. Factors. **37**(4), 682–698 (1995)
128. Kroemer, K.H.E., Grandjean, E.: Fitting the Task to the Human: A Textbook of Occupational Ergonomics, 5th edn. CRC Press, London (1997)
129. Colombini, D., Occhipinti, E.: Scientific basis of the OCRA method for risk assessment of biomechanical overload of upper limb, as preferred method in ISO standards on biomechanical risk factors. Scand. J. Work Environ. Health. **44**(4), 436–438 (2018)
130. Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H.W.: Handbook of Human Factors and Ergonomics Methods, 1st edn. CRC Press, London (2004)

131. Salvendy, N.G.: Handbook of Human Factors and Ergonomics, 4th edn. Wiley, Hoboken (2012)
132. Hoy, D., Brooks, P., Blyth, F., Buchbinder, R.: The epidemiology of low Back pain. Best Pract. Res. Clin. Rheumatol. **24**(6), 769–781 (2010)
133. Siew, C.Y., Ong, S.K., Nee, A.Y.C.: Improving maintenance efficiency and safety through a human-centric approach. Adv. Manuf. **9**(1), 104–114 (2021)

Soh Khim Ong is Associate Professor at NUS. Her research interests are virtual and augmented reality applications in manufacturing and life cycle engineering. She is Fellow of CIRP. She has published 5 books and over 330 international refereed journal and conference papers. She received 2004 OYME Award from SME and 2009 Emerging Leaders Award in Academia (US Society for Women Engineers).

Siew C.Y. obtained his PhD in 2019 from the National University of Singapore Graduate School of Integrative Science and Engineering, investigating the applicability of augmented reality in maintenance applications. His research interests include augmented reality, IoT systems, and human-computer interaction.

Andrew Yeh Ching Nee received his Ph.D. and D.Eng. degrees from UMIST, Manchester. He is currently Emeritus Professor in the Department of Mechanical Engineering of NUS. He is Fellow of CIRP, SME, and Academy of Engineering, Singapore. His research interests are in manufacturing engineering, augmented reality, and digital twin. He is Editor-in-Chief of IJAMT and Executive Editor-in-Chief of AIM.

23

24 Augmented Reality for Maintenance and Repair

Salvatore D'Onofrio, Raffaele Vertucci, Stefano Ricciardi, and Maurizio De Nino

Contents

24.1 Introduction . . . 597

24.2 AR for Maintenance . . . 598

24.3 AR Solutions for Repairs and Troubleshooting . . . 600
24.3.1 AR and Industry 4.0 . . . 600
24.3.2 Techniques and Methodologies Behind AR-Aided Servicing . . . 602
24.3.3 Augmentation Strategies for AR-Based Guidance to Repair Procedures . . . 602
24.3.4 Near-Future Scenarios: Collaborative Repair and Troubleshooting Through AR and Virtual Assistants . . . 603

24.4 Design of a Virtual Toolbox: The iEngine Framework . . . 604
24.4.1 Layered Architecture of iEngine . . . 605
24.4.2 iEngine Module Architecture . . . 606
24.4.3 Novelty of the Developed Approach in iEngine . . . 608

24.5 Use Case and Experimentation . . . 608
24.5.1 VTE Modules . . . 610
24.5.2 VTE Application AI Module . . . 611
24.5.3 VTE Final Result . . . 612

24.6 Conclusions . . . 612

References . . . 614

Abstract

As an emerging and powerful technology, augmented reality (AR) combines the real world with virtual and computer-generated objects in order to give to users useful data about the task they are performing. Due to its features, AR is a promising way to enhance many activities requiring the intervention of specialized personnel combining, sometimes, the use of other tools such as real-time monitoring, fault diagnosis, and communication systems.

This chapter presents the use of augmented reality (AR) methodologies and technologies to improve maintenance, repair, and troubleshooting activities by enhancing the technicians' performance from the training to the on-site interventions to reduce time and costs. Then, it presents an innovative framework designed to facilitate the development of high-performance 3D real-time solutions for Industry, Marketing, Cultural Heritage, and Healthcare. Finally, the chapter ends with use case, test, and experimentation.

Keywords

Augmented reality (AR) · Virtual reality · AR applications · AR technologies · Maintenance · Repairs

24.1 Introduction

The term augmented reality (AR) indicates the superimposition, on what a person perceives with his senses in a real environment, of more sensory information generated through devices that the user generally wears or that are in his nearby. The term "augmented" refers to the fact that other information is added in addition to what the user naturally perceives. Ideally, augmented reality could be applied to each of the five senses, but today, the one with the highest potential for use is certainly sight. Nowadays, when we talk about augmented reality, we refer to the fact that the user sees additional computer-generated information in his real environment.

Unlike the virtual reality (VR), the AR does not aim to replace itself to the real environment, but to take advantage

S. D'Onofrio (✉) · R. Vertucci
Leonardo S.p.A, Giugliano in Campania, Italy
e-mail: salvatore.donofrio@leonardocompany.com; raffaele.vertucci@leonardocompany.com

S. Ricciardi
Department of Biosciences and Territory, University of Molise, Pesche, Italy
e-mail: stefano.ricciardi@unimol.it

M. De Nino
Digitalcomoedia S.r.l, Naples, Italy
e-mail: m.denino@digitalcomoedia.com

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_24

of it. To achieve this goal, AR generates new ideas for interaction between the environment and the user [1], increasing the connection between the senses and the real phenomenon observed, with particular interest in the perception of the place, of movement, and of information.

Nowadays, AR has become a cutting-edge technology in many sectors. For example, the technological evolution, the search for greater efficiency, and the development of technological processes have led to the construction of increasingly complex equipment and have changed the approach to the solving methodologies for some problems. In particular, the need has developed to maintain the efficiency of the systems and equipment that compose them through the activity necessary to ensure the continuity of operation of the systems: the maintenance. The term maintenance refers to the company function, which is responsible for the constant control of the plants, and the set of repair and overhaul works necessary to ensure the smooth operation and good state of conservation of the production plants, services, and equipment.

The aim of this chapter is to discuss about the application of AR in the industrial area where it can be very useful to assist technicians in the maintenance and repair of complex systems and equipment.

In fact, many industries, more and more often, try to move the design and construction phases to the computer, in order to replace the physical prototypes with virtual mockups for assembly and to carry out safe tests. The usefulness of this is particularly evident in the aircraft and automotive sectors, where physical prototypes are very expensive and time to market is a crucial factor. In fact, the instructions for assembling or repairing specific systems or equipment are easier to understand if they are available in the form of 3D images and not as paper manuals. This is the case of maintenance and repairing instructions, which are superimposed on the real equipment and shown to the user by means of a head-mounted display (HMD), smart glasses, or mobile device. In this way, the maintenance and repairing process can be significantly speeded up, and the probability of a worker making a mistake or a damage to the system can also be reduced.

Maintenance and repair activities represent an interesting and full of opportunities field for augmented reality. Most of these activities are carried out by qualified and trained personnel who perform procedures established by technical documentation in a relatively static environment and predictable. These procedures are generally organized in sequences of elementary operations to be carried out on particular components in a predetermined position. The combination of these and other features allows the development of numerous systems and technologies that can assist a technician during the performing of maintenance activities. Both technological systems and industrial products require global, rapid, and appropriate maintenance services. Unfortunately, too often the level of competence of the operators in the field of these systems does not prove adequate and in step with the continuous technological change, and therefore, a continuous updating is required by the staff and the assistance manuals, not always available or easy to understand.

In this context, augmented reality should be seen as a technology to provide highly effective on-the-job training. In fact, it provides the right amount of configuration information and tasks to perform the activity. Thus, the technician uses this information to increase preliminary general knowledge and then manually apply it to the repair activity. Moreover, combining digital Internet of Things (IoT) solutions with an augmented reality device, technicians can optimize how they prepare for and carry out the task, resulting in faster maintenance and repairing times and less system downtime. The potential of this technology is enormous, and, in addition to several academic studies [2], there are already some uses in the industrial field [3–5].

24.2 AR for Maintenance

Maintenance activities are of fundamental importance for the logistical support of the equipment and must be carried out periodically or in the event of breakdowns on equipment that varies in complexity from the simple replacement of a spare part to the more complex and time-consuming repair. The opportunities for integrating AR into maintenance processes are vast. The ability to track and document the work could be enhanced by leveraging wearable technologies that provide engineers with a high view of assets.

The main benefits of augmented reality applied to maintenance activities are related to the reduction of costs, human errors, execution time, breakdowns, downtime of system, and, in the increase of productivity, operation speed, fix rates, compliance, and profits. Moreover, the use of smart glasses leaves the worker's hands free to perform maintenance and repair activities while viewing the instructions superimposed on glasses.

AR-based solutions for maintenance operations include preventive maintenance, service inspection instructions, step-by-step instructions for unfamiliar procedures, corrective maintenance, predictive maintenance, and real-time access to data.

In particular, preventive maintenance involves setting downtime in convenient times for production and guaranteeing the supply of all spare parts in time and finding the necessary personnel. However, this strategy requires careful planning, and, if not optimized, the frequency of interventions ends up being too high or too low, compared to the actual needs of the equipment. In this kind of maintenance activity, augmented reality is useful to support the technician in performing assembly instructions, service inspection instructions, and compliance checklists and to access to detailed instructions for unfamiliar procedures. Corrective maintenance instead is related to the correction of faults by individuating the damaged part and in performing the right procedure for its substitution and the recovery of the system. In this context, augmented reality comes in helping with the providing of service manual instructions and remote assistance to retrieve useful data from a remote expert and to visualize them through smart glasses or mobile devices. Finally, for the predictive maintenance activities, where the user has the aim to predict the fault causing the downtime of the system, the augmented reality could be used to access to performance panels and real-time data for monitoring, control actions, disruptions, and analytics.

Nowadays, technicians know the maintenance procedures by heart, and they have confidence in their own skills so that often they do not carry out refresher courses for the maintenance of skills. However, it is precisely this excessive confidence that can often lead to risky operations that can compromise the system or cause damage to their health. In this context, AR is useful as it allows with minimal effort and little invasiveness, thanks to the use of light smart glasses and mobile devices, to guarantee continuous help to technicians during their work and training in the workplace that otherwise they would not have been carried out. Applying the Industry 4.0 paradigm, machines are increasingly interconnected according to the IoT paradigm, and continuous monitoring allows technicians to be notified in advance before maintenance is needed.

Moreover, technological safety systems for critical applications such as aircraft or navigation radar systems require global, rapid, and appropriate maintenance services. Unfortunately, too often, the level of competence of the operators in the field of these systems does not prove adequate, due to the continuous technological change, and therefore, they require a continuous updating by the staff. The so-called collaborative working environments [6] provide excellent opportunities to improve service and maintenance procedures, significantly reducing the time required for repairs as well as safety risks connected to the inactivity of the malfunctioning system.

In support of these activities carried out with the help of augmented reality, communication systems are used more and more often which allow a remote expert technician to give instructions in the form of text, labels, or sketches directly on the technician's on-site display; it superimposes its view or positions itself directly on the apparatus thanks to the spatial recognition through markers. In fact, moving forward on the reality continuum toward mixed reality, we are faced with a large number of applications in which augmented reality can be applied. It is the case of the overlapping of dynamic objects or avatars that are anchored to the real environment, and in this, they move respecting the physical constraints around the operator. One of the typical applications could be that of a virtual assistant that indicates procedures, tasks, and maintenance workflow to be performed (Fig. 24.1).

Fig. 24.1 Virtual assistant

According to this approach, support for maintenance procedures is provided by an avatar or virtual assistant. The word avatar is now widely used; this popularity is mainly due to the world of cinema, entertainment, games, and 3D simulations and even to social networks. Technically, an avatar is a particular structure used in computer graphics, mainly composed of a deformable three-dimensional model, depicting a human form, capable of performing very complex animations; more generally, this term indicates the synthetic representation of a character. The virtual assistant is displayed on the operating field and is able to communicate with the human operator through an interaction paradigm close to the way humans are used to collaborating and cooperating.

Another application of augmented reality, always concerning the maintenance of systems, is that of remote support in real time in which the technician on-site, who is in front of the apparatus, communicates with an expert technician remotely through a mobile device or an HMD device to ask for help on maintenance procedures and tasks to be performed. The remote expert technician can send instructions and information that are superimposed on the visual field of the technician on-site, thus increasing his perception of reality with useful data for his activity. For this purpose, the devices used are equipped with front cameras that allow you to send images and videos of the current state of the apparatus and the view of the technician in front of you.

However, aside from the aforementioned AR application ideas in the field of maintenance, many companies have developed their own hardware and software AR solutions to support maintenance activities. Moreover, the number of research projects clearly implies that AR for maintenance is a challenging area of investigation. The most advertised devices are definitely the Google Glass [7] and Microsoft HoloLens [8] even if more specific devices are developed by other companies all over the world.

24.3 AR Solutions for Repairs and Troubleshooting

Industry 4.0, including a technological mix of robotics, sensors, connection, and programming, represents a new revolution in the way of manufacturing products and organizing work. The continuous evolution of technologies is diversifying the declination of Industry 4.0 on multiple levels and operating areas, and the most innovative digital technologies will have a profound impact in the main development directions. One of these directions is represented by an intense interaction between man and machine, which involves natural interaction models (touch, gestures) and the use of augmented reality to achieve objectives such as increasing productivity and reducing errors in industrial processes.

As regards the first objective, the integration of AR with the Industrial Internet of Things (IIoT) represents one of the most disruptive scenarios in this revolution. AR technologies and methodologies have always been close to the industrial world, promising tangible advantages in terms of increased efficacy and efficiency [9, 10]. This synergy allows the industrial plant operator to view and interact with data in real time from any IoT-enabled machine. Workers can perform repairs, diagnose problems, and become aware when data reacts in response to their interactions. By using the display of mobile and wearable devices to view 3D content and 2D information overlaid onto the real world, it is possible to provide instructions and orientation to operators. Such systems can effectively support a wide variety of workflows, for example, by displaying the information required for a given procedure onto the operator's field of view. Superimposed checklists can be used to guide the operator to perform unfamiliar tasks, but they can also serve to verify that the activities are complete according to the requirements and in the right order. This operational paradigm allows all users to have an updated list of their tasks and orders and to create and assign, for example, daily work orders and service and maintenance tickets. At a more basic level, the visualization of the messages, relating to the interactions between different operators involved in a common task in progress, can also take place through the see-through device.

With reference to the second objective, namely, reduction of errors related to industrial processes, remote video assistance with the sharing of the visual field within an industrial environment represents the most obvious and most widespread functionality required by most cases of industrial use [11]. It allows immediate access to a more experienced remote expert who can support the operator in the field with advice, detailed information, schematics, and animated drawings visualized in the visual field. For this purpose, it is possible to make audio and video recordings and take pictures to share best practices, document the quality of the work, or take notes on repairs and inspections. The sharing with the work team and the automatic insertion into the compliance and documentation processes represent further collateral benefits. The possibility of a precise localization of the operator, through tracking systems, allows to provide the operator with an augmented view of the spot in which he is located and of the most important aspects of the surrounding area, as well as to provide work instructions contextualized to that particular location. A digitized and possibly animated version of manuals and technical data sheets can be displayed according to the spatial and operational context, even in co-registration with physical objects. The overlay display of alerts can immediately notify operators if a problem or error occurs in the workflow, to update the status of activities or to warn them of potential safety hazards in their environment.

From a more technical point of view, the aforementioned goals imply adequate strategies and technologies. The precision of tracking and visualization strategy adopted represent two key aspects that have to be carefully taken into account in industrial component augmentation. Computer vision , in both the marker-based [12, 13] and marker-less [14, 15] variants, is generally recognized as the only tracking methodology that has the potential to yield noninvasive, accurate, and low cost co-registration between virtual and real. Concerning the visualization strategy, how and when to augment real objects and environment may have a great impact on the quality of user assistance provided. For instance, in some situations, the scene observed should be possibly simplified rather than augmented.

Another aspect to be considered is represented by the interaction level available and the related interaction paradigm: the user should be able to select what kind of augmenting content to display according to his/her needs by interacting with the AR environment without complicated gear. To this aim, voice commands can be used for interacting with an AR framework [16, 17], but a hand-based interface would rather be more suited to the scope. As a hardware solution (instrumented gloves plus wrists tracking) would provide accurate hand capturing but would also reduce system's acceptability, a more feasible option is to exploit image-based techniques to track hands in real time [18], possibly exploiting the same cameras used for inside-out tracking [19].

24.3.1 AR and Industry 4.0

In the context of the typical operating scenario of Industry 4.0, augmented reality, if possible, should be a ubiquitous tool, which allows access to information and digital content and to tackle complex situations from any workplace. Whether it is a procedure, a fault, a machine downtime, or a component malfunction, the operator should be able to easily use a see-through head-mounted display (or any other see-through device) to become connected with the system, interacting with it through a touch-based [20] or gesture-based interface [21].

According to this AR-based guidance paradigm, relevant or explicitly required content is retrieved by the knowledge base and displayed into the operator's field of vision, according to the operating context. An AR platform collects data from the machinery and transmits the information necessary to complete even the most complex production and maintenance activities. These capabilities are supported by devices such as smartphones, tablet, and smart glasses that overlay 3D content directly on a live video capture of the real environment (video see-through) or directly in the operator's visual field (optical see-through). The operator collects information and guided instructions in AR mode, locates any errors and malfunction, and plans the interventions with great simplicity and precision during the running, maintenance, and repair of the machinery (see Fig. 24.2). To achieve this ideal result, starting from the user interface, everything should operate in the most natural and effective way. The operator interacts in a simple and efficient way with the machinery while simultaneously accessing the knowledge base required for solving the problem. The platform collects the data and conveys it from a server or system to mobile or wearable devices. The display of the guided instructions in AR mode generates better performance and contributes to making the production system more efficient. This type of application favors the sharing of information in real time (remote expert) and the centralization of data to make the maintenance process easier and more efficient. It facilitates the operator's identification of the fault, allowing to overcome the limits deriving from the use of paper manuals which slow down the repair process. The smart guidance approach allows you to integrate augmented reality functions with other skilled technologies of Industry 4.0 such as IoT, artificial intelligence, Big Data analytics, and machine learning.

Fig. 24.2 A technician performing an AR-aided servicing procedure on an industrial rack containing multiple electronic boards. The colored cap on the pointing finger allows tracking its position via the see-through cameras aboard the head-mounted display

The main advantages associated with the use of this type of industrial AR application include a measurable increase in productivity and efficiency of the production line thanks to the collection of data (information and guided instructions) that are conveyed by a server or plant in operation according to the context and transmitted directly into the operator's field of vision during routine or extraordinary production or maintenance activities.

Another key aspect is represented by cost optimization and error reduction, since the system provides support to operators by promptly detecting any process errors or malfunctions and possibly by planning interventions according to component usage history and failure statistics. A further advantage comes, then, from the support offered to the digitalization, as this type of environment integrates easily with existing cloud infrastructures and generates reporting archives with analytics data and shared or exclusive access areas to streamline workflows and automate production processes.

A different kind of emerging AR application targeted (not only) to the industry world is represented by the **AR-based virtual showcase**. This refers to an AR platform enabling to bring an unlimited number of machinery or equipment of any size at international events and fairs, enabling to carry out realistic demonstrations in the physical space in which they will be shown or used. Instead of bringing complex and/or bulky products to trade fair events with related economic and organizational costs deriving from packaging and transportation to distant locations, they can be viewed anywhere as they were physically present. By selecting the desired machinery or equipment, the interested visitor can have access to all its features through a 3D photorealistic animated model, visualized in the real environment in real size, providing an accurate and effective showcase of any industrial product or components, even the most complex ones. It is also possible to add and superimpose additional text or graphics, along with manuals and commercial and technical documentation.

The main advantages include the possibility of displaying large machinery on stands with limited exhibition space, the consequent reduction in logistics and packaging costs, the possibility of interactively selecting and displaying different product configurations, and the generation of analytical reports.

24

24.3.2 Techniques and Methodologies Behind AR-Aided Servicing

In the case of a technical intervention, the classic printed manual still represents a valid source of support, but often its paper nature makes it difficult to consult. An augmented reality application that highlights, in the context of a particular maintenance procedure, what are the points on which to operate certainly represents a great alternative to the aforementioned manual with the added value of being immediately usable even by not expert operators and is easily adaptable to any changed operating conditions.

Relying on the aid of such an application, indeed, releases a constraint that can be often compelling in some industrial contexts, namely, the prompt availability of a maintenance technician who has precise skills on the equipment to be maintained. In the following lines, a typical use case, the maintenance procedure of a complex electronic unit in the form of an industrial rack, provides a good example of the main technical aspects behind AR-aided servicing. A virtual representation of the actual location of the industrial racks inside the room is stored in the system's knowledge base. A tracking system intercepts the movements of the maintenance technician, to obtain his current viewpoint required to co-register virtual contents to real world. From here, the system enriches the scene with all the useful visual aids. More precisely, once the type of procedure to be performed is chosen, the system shows, step by step, what are the points on which to work (hotspot) enriching the information with textual descriptions of the task and also showing the virtual tools, if necessary, to complete the task. The ideal maintenance procedure is the one that offers to the operator the right directions about how to carry out his work in the shortest time and with no errors. This paradigm implies that a lot of effort is required to design and implement effective AR-based assistance procedures. These procedures are composed of a sequence of steps , each to be completed to pass on the next one (see Table 24.1 and Fig. 24.3). It is worth remarking that the accuracy of the virtual-to-real co-registration is of paramount importance, so that the visual aids can be correctly superimposed onto the real scene.

24.3.3 Augmentation Strategies for AR-Based Guidance to Repair Procedures

With regard to visual field augmentation, different strategies have been proposed and can be applied according to different needs and applicative contexts. The most diffused approach to delivering visual aids by means of AR consists of displaying different types of computer-generated graphics (e.g., text, arrows, labels, 3D models, and so on) onto the working field. This visual modality literally "augments" the

Table 24.1 Breakdown of basic tasks in a maintenance procedure

Index	Description
T1	Unscrew four fixing screws
T2	Check if LED is *on*
T3	Unplug the 41-pin connector
T4	Check if LED is *off*
T5	Change the state of the switch
T6	Screw four fixing screws
T7	Change the state of the switch
T8	Plug the 41-pin connector
T9	Unscrew two fixing screws
T10	Change the state of the switch
T11	Check if LED is *on* and if it is yellow
T12	Screw two fixing screws

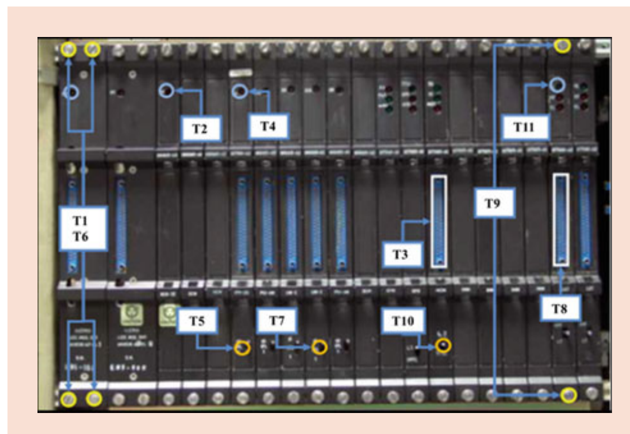


Fig. 24.3 Labels identifying locations on the rack associated with the tasks described in Table 24.1

real environment. However, there could be particular cases where the presence of additional information to the operating field may result visually confusing and detrimental for the task execution.

This situation can easily arise wherever the working environment is characterized by the presence of many interaction points or even many small items in a relatively small area (e.g., high-density electronic boards, complex devices, etc.). In this particular case, the superimposition of additional (virtual) objects to those already present may be counterproductive. To address the aforementioned issues effectively, the "diminished reality" paradigm has been proposed. The underlying idea is to subtract part of the distracting elements of the real world instead of adding something to it. The goal is facilitating the user to focus solely on the elements strictly necessary to perform a particular task. This methodology is based on the selective occlusion of unwanted elements rather than on image-based object removal [22]. Figure 24.4 provides a visual example of this modality.

Both augmented and diminished reality strategies can improve user's operational capabilities, yet there are contexts

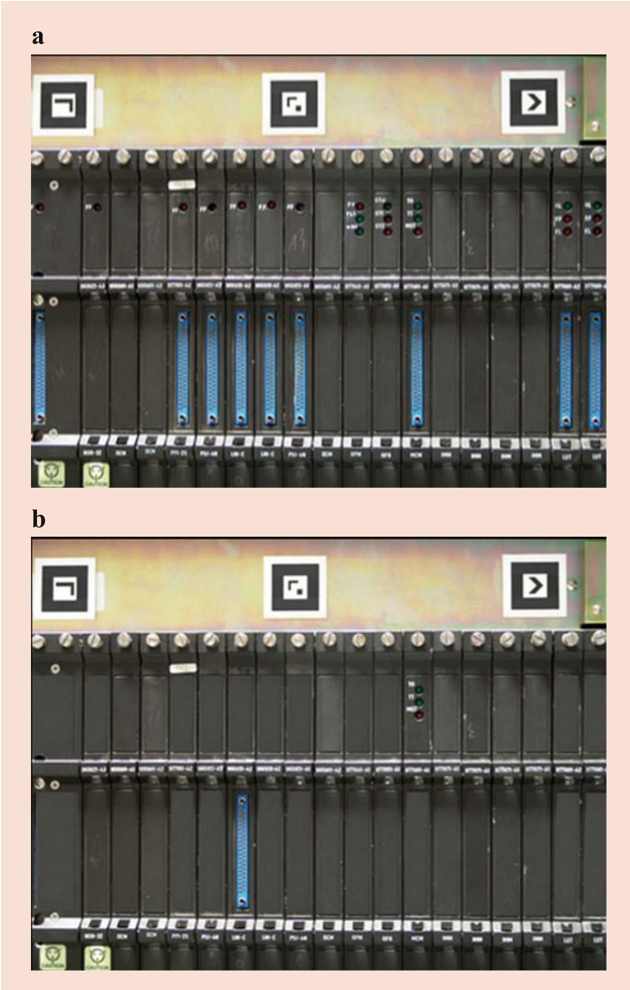


Fig. 24.4 Before (**a**) and after (**b**) view of visual field when diminished reality is applied

in which one is more suited than the other. Which of the two visual approaches should be used depends on aspects such as the total number of hotspots and/or their spatial proximity. If these parameters exceed a threshold, then the diminishing modality is preferred. Anyhow, the user may always switch to the other modality in any moment by means of a vocal command. Finally, by combining both augmented and diminished reality, a third hybrid visualization approach could be realized, providing a simplified view of the operating field in which only the elements left visible are augmented with additional info.

24.3.4 Near-Future Scenarios: Collaborative Repair and Troubleshooting Through AR and Virtual Assistants

Mission-critical high-tech complex systems typically require timely and effective maintenance and repair capabilities, but operators in the field often do not complete the necessary skills also due to the constantly updated knowledge base involved. In this context, the latest generation of collaborative work environments provides excellent opportunities for improving the maintenance and service procedure, including significantly the repair time and the risks associated with the system's activity. In fact, through the development of a collaborative 3D environment for the maintenance of industrial machinery, it is possible to simulate the work of operators in the field during the identification, diagnosis, and repair of system malfunctions [23]. Such a system can simplify the maintenance procedure by displaying various types of information content (labels, manual pages, schematic diagrams, images, or example of animations) that release the operator from the need of searching for the solution through a printed manual.

A collaborative environment provides technological means and information exchange opportunity to simplify and foster collaboration among people toward the fulfillment of a given problem too. However, this does not necessarily imply that collaboration should operate only at a human-to-human level. It is possible, indeed, to define a scenario in which an operator is alone but assisted by the intervention of a virtual assistant (an avatar) seen within the operating field and is able to collaborate through an interaction paradigm similar to that normally adopted among humans.

This avatar, whose actions are included on a semantic representation of the service/maintenance workflow, can provide the necessary assistance according to the user's needs through a proactive behavior, helping him during the research and repairing the malfunction in a much more effective way than conventional on-screen managed interfaces. A key aspect of this solution is the presence of a wearable computing unit, possibly integrated within the AR visor, or, alternatively, a wireless connection to a host computer, to the aim of freeing the user from wiring and umbilical cords.

According to this operative paradigm, an example of avatar-assisted maintenance session is depicted in Fig. 24.5, where a virtual technician shows to the human operator the diagnosis (and eventually repair) workflow (for instance, as a floating flowchart) and what step has to be performed next, aiding him to locate specific hardware components or to measure operational parameters in specific spots. The avatar could even recall and visualize a particular hardware assembly and the related mounting/un-mounting/replacing instructions, avoiding a search through extensive paper manuals. During intervention, voice synthesis and voice recognition technologies (over a reduced context-specific vocabulary) are exploited to allow a friendlier avatar-/field-operator interaction.

Besides the AR architecture, the avatar-based assistance framework involves another fundamental aspect, an interaction paradigm, regulating the different ways in which avatar/

Fig. 24.5 A third-person view of human-avatar interaction in a collaborative environment

field-operator collaboration is performed during each phase of the maintenance/repair procedure. Such a virtual assistant can be considered as a real-time system capable of reacting to the operator's requests. This explains why one of the main challenges to fully exploit the avatar's potential as an anthropomorphic interface is related to the development of an advanced Chabot capability to achieve an effective collaboration of the subjects (real and virtual) involved.

Moreover, as the typical interaction paradigm between humans is based on known rules involving the reciprocal attention of both subjects involved (by means of eye movements, gestures, etc.), an important aspect of the proposed approach is the avatar's ability to look at the operator and to underline the verbal communication through a coherent body language. These requirements could be addressed by exploiting biometrics and context-aware adaptive methods to capture operator's speech and to elaborate his/her position/orientation data provided by the AR architecture to adapt avatar's action accordingly (for instance, the avatar will wait for an operator's confirmation before proceeding with assistance workflow). The high-level behavioral model should be capable to represent avatar's appearance and activity through static and dynamic aspects.

The static aspects of a virtual assistant can be defined as the set of morphological features that the avatar shows in a given moment. This set of morphological features is dealt as a hierarchical structure composing of different subtrees. Each subtree is devoted to characterize a homogenous set of avatar's morphological feature (e.g., voice features, body features, etc.). In other words, the avatar static view defines a sequence of avatar snapshots characterizing its morphological representation during its life cycle.

However, in order to define a coherent and domain-oriented sequence of static avatar snapshots, an inference engine, representing the avatar dynamic aspects, is required. In particular, the avatar inference engine has to be capable of driving the avatar's action in response to user's needs and requirements.

Though the aforementioned challenges represent hot research topics and many partial results have been achieved in various applicative contexts, the full paradigm of human-avatar collaboration in an AR environment is still not ready to be applied in real industrial contexts. However, this scenario is closer than it may appear.

24.4 Design of a Virtual Toolbox: The iEngine Framework

The value of mixed reality for training and maintenance assistance has been clearly proven, and an increasing number of organizations are beginning to incorporate MR technology. In fact, the industry is becoming one of the market leaders for MR applications. However, there are different requirements for each industrial sector, so it becomes difficult to have one application suitable for all. For this reason, the idea is to have a toolbox on which to build specific implementations adaptable to various needs. iEngine (interactive Engine) is a framework whose goal is to facilitate the development of high-performance 3D real-time solutions for Industry, Marketing, Cultural Heritage [24], and Healthcare [25].

iEngine represents a toolbox capable of maximizing the benefits associated with virtualization and the enjoyment of immersive experiences not only on the desktop side but also through advanced virtual reality and mixed reality equipment and related interaction technologies.

For the industry, it represents an effective tool for physical familiarization with company products and for training of staff, both for maintenance and for assembly.

In addition to the well-known desktop applications, iEngine allows to take advantage of two application paradigms: MRE (mixed reality environment) and VRE (virtual reality environment).

The iEngine framework provides a set of tools, modules, and libraries in a single environment to allow and facilitate the development of complex projects.

iEngine represents advanced environments for interaction with three-dimensional virtual objects. This technology implements a high-performance real-time architecture, with high visual quality capable of providing powerful virtualization tools. The technological kernel uses a 3D graphics engine and other third-party software, allowing the interactive use of 3D digital environments. The goal of the platform is to have a range of tools that allow, through iEngine's high-level APIs, to quickly create 3D and 2D applications.

The idea of building iEngine on the top of a "game engine" gives the possibility to use a complex set of modules and

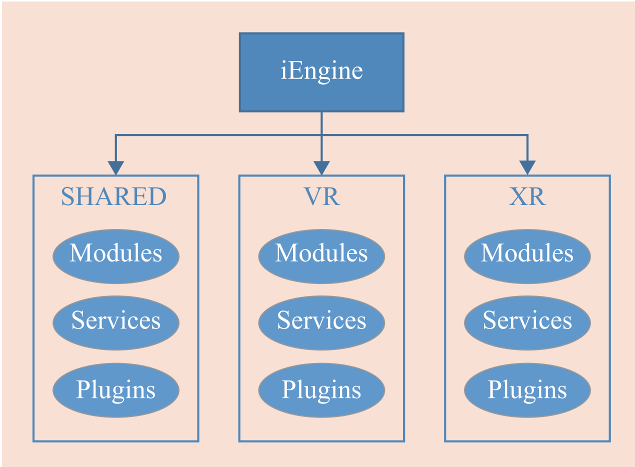


Fig. 24.6 iEngine's areas

functions that go beyond just graphic rendering. A game engine is a software framework designed for the creation and development of video games, which is based on a reusable architecture that therefore does not contain any part of the game but includes a suite of tools and components that can be used at runtime and which allow both low-level and high-level programming (Fig. 24.6).

To meet all the needs of the various platforms, iEngine provides three areas:

- SHARED
- VR
- XR

The SHARED area represents the basis for all the other areas so it is the largest set of tools. It contains all the tools useful for the development of desktop, WebGL applications, and a set of generic libraries; many modules in this environment are also present in others.

The VR area represents the container of all the tools useful for the development of applications that make use of HMD. For example, it contains modules for VR management of inputs, interactions, and locomotion.

The XR area represents the container of all the tools useful for the development of mixed reality applications. It contains XR-specific libraries and modules for XR management of inputs and interactions.

The presence of a specific module within multiple areas introduces the concept of MultiModule. An example of a MultiModule can be seen in Fig. 24.7.

A MultiModule offers various advantages, including the most obvious one, that is, the possibility to switch from one development area to another without too much effort, while maintaining the interface and functionality of the module, since these are shared.

24.4.1 Layered Architecture of iEngine

The architecture of iEngine is organized as a modular layered structure, where the modules are suitably connected and each level encapsulates the services and functionality of the underlying layers (Fig. 24.8).

Each layer has specific tasks and activities, according to the methodology defined as the "value added principle."

The architecture consists of a set of modules organized in five levels:

- User Layer
- Hardware Layer
- Low-Level/Interface Layer
- Middle-Level/Manager Layer
- High-Level/Decision-Making Layer

User Layer
The User Layer is the layer that represents the actors that are present in our applications. To be more general and to design an architecture that can be used in many different situations, this layer includes the following users: the User, the Admin, the App and/or Website that interfaces with the framework, and the Operator.

Hardware Layer
The Hardware Layer is the layer that contains everything that is hardware, i.e., physical, and that is used by the applications developed using the iEngine framework. Several components are part of it and are application dependent.

Low-Level/Interface Layer
The low-level software layer translates all the data that comes from the Hardware Layer into more suitable information and to communicate with the upper layer at the highest software level.

Middle-Level/Manager Layer
The medium-level software layer includes the management modules. For optimization reasons, this layer is divided into two sub-layers. The first level includes mainly the managers who are connected with the Interface Layer and which act as inputs to the upper-layer modules, usually Decision-Making. The second level communicates only with the modules of the Middle Level and offers higher-level functions than the managers of the first layer, however, requiring some features from the latter, so it cannot exist individually.

High-Level/Decision-Making Layer
The Decision-Making Layer is in charge of all the application logic including also the AI components. This layer is application dependent.

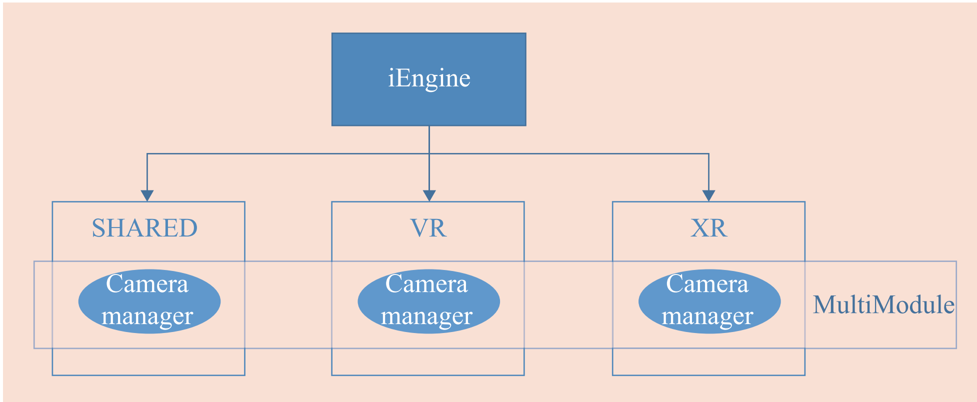


Fig. 24.7 iEngine MultiModule

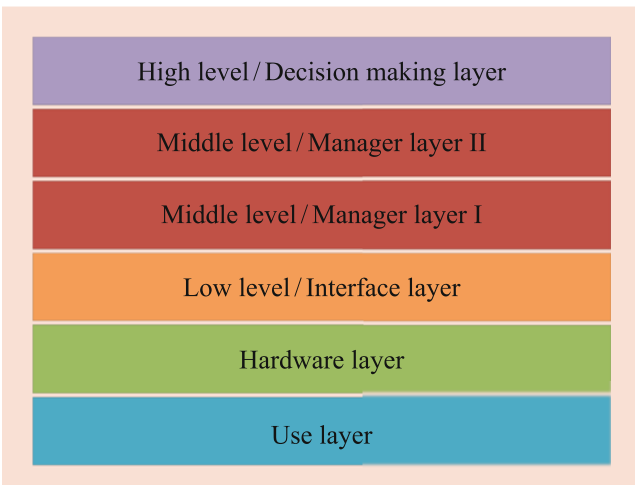


Fig. 24.8 iEngine layer architecture

24.4.2 iEngine Module Architecture

Every layer of layered architecture can include one or more iEngine modules. Each module contains services and plugins, and a certain number of modules can be grouped in packages (see Fig. 24.9).

Modules

A module of iEngine can be considered as an entity that extends the abstract BaseModule class by implementing its related abstract functions. The BaseModule class implements:

- A Module Connection System: the system that takes care of connecting the modules in such a way that they can receive messages from other modules or send messages to other modules
- A Module Error System: that is, the exception handler and notifier

The BaseModule interface defines all the properties and methods that must be present within a generic iEngine module. In it, the Enabled field indicates whether the module is enabled or disabled.

When a new instance of BaseModule is created, by default, the Enabled field assumes a value of false, and it is necessary to call the Enable method to enable it. The BaseModule destructor, on the other hand, disables the module and removes all connections with the other modules before deleting the class instance.

The IBaseModule interface extends the IBaseEntity, ISystemEntityPluginHandler, and ISystemEntityServiceHandler interfaces. The first is concerned with defining that a module is an entity that provides an Input Entity System, an Output Entity System, and an Error Entity System. The Module Connection System is defined through the implementation of the methods and properties contained in the Input Entity System and Output Entity System interfaces. The Module Error System is defined through the implementation of the methods and properties contained in the Error Entity System interface. The ISystemEntityPluginHandler and ISystemEntityServiceHandler interfaces contain the part of the Module Connection System that describes how Modules can be connected with Plugins and Services, respectively.

The ErrorOccured event of the ISystemErrorHandler interface is used to notify that an exception has occurred within the module. This event is defined within the BaseModule class and is used to generate a call to the abstract OnError method that must be implemented in each module that extends BaseModule.

The OutputChanged event of the ISystemOutputHandler interface must be defined in each module that extends BaseModule and is used to send messages to other modules.

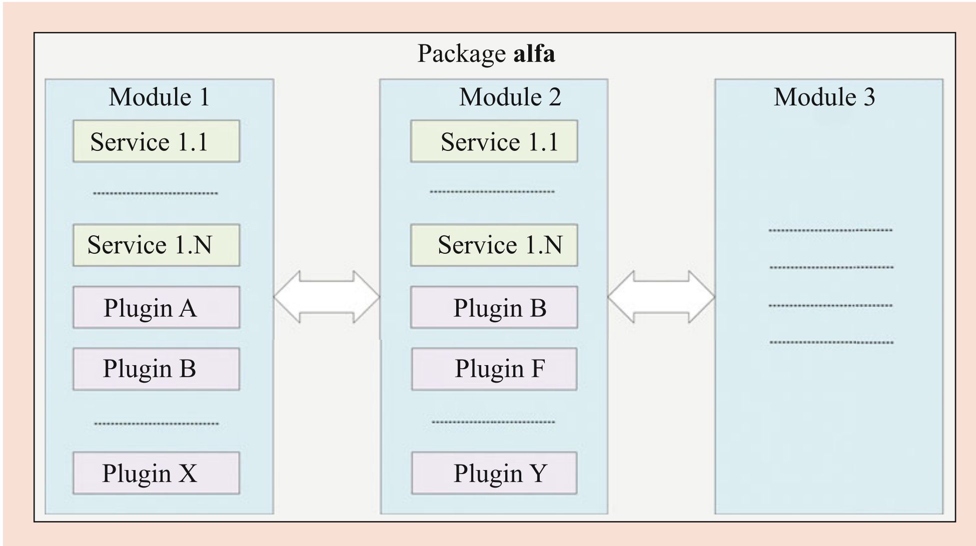


Fig. 24.9 iEngine hierarchy

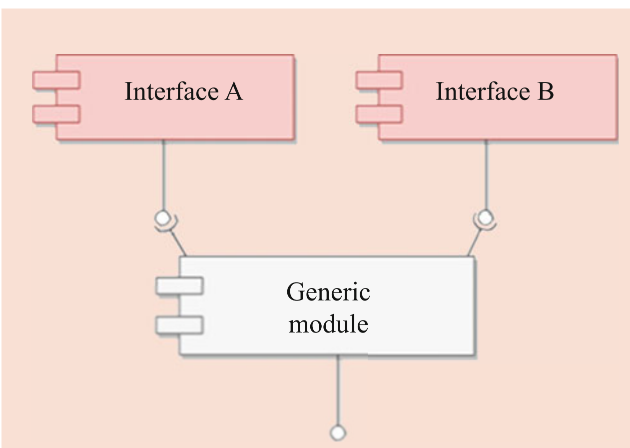


Fig. 24.10 Example of a generic module

The generic ISystemInputHandler <T> interface defines the part of the Module Connection System relating to the reception of messages. For the reception of messages, the interface specifies that a generic collection of T-type objects (InputEntity field) and methods for acting on this list at runtime (AddInput and RemoveInput methods) are required (Fig. 24.10).

With this configuration, a generic iEngine module can receive messages from other modules by registering with the respective OutputChanged events and can send messages to other modules through the SendMessage callback. The abstract BaseModule class implements a system of self-registration (RegisterHandlers method) and deregistration (RemoveHandlers method), respectively, when calling the Enable and Disable methods. These methods are virtual methods of the BaseModule and can be redefined.

To extend the abstract BaseModule class, it is necessary to implement the required abstract methods:

- OnInputChange: the method that takes care of handling messages received from other modules and possibly sending the feedback after analyzing the inputs
- OnError: the method that is concerned with notifying the exception that has occurred within the form

It may be useful to define the OnInputChange method by adding the keyword “async” if there are awaitable tasks inside it (e.g., I/O operations on the disk, RestAPI call, or in case it is necessary to wait for the termination of one or more IEnumerable functions).

The architecture described so far is implemented with the C# and dotNet 4.6 language, and the classes do not include components of the Mono subset, so it is highly portable on any platform that makes use of the aforementioned languages.

However, due to the absence of multiple inheritance in C#, it is not possible to directly implement classes within Unity3D which can be serialized through the Inspector of a GameObject (the object must derive at least from UnityEngine.Object), i.e., a module cannot directly extend the MonoBehaviour class to be attached to a GameObject and at the same time extend the BaseModule class which defines the general architecture of a module for iEngine.

For this reason, a new abstract MonoBaseModule class has been defined which wraps all the functions described in the BaseModule class in a serializable environment having a structure equivalent to it.

Therefore, to build a module on Unity3D, it is possible to:

- Extend the MonoBaseModule class, implementing all abstract methods (as for BaseModule).
- Extend the BaseModule class, and wrap the basic methods as implemented in the MonoBaseModule class.
- Build all the modules you need by extending the BaseModule class, and use them within a MonoBehaviour class by suitably connecting all the modules in a procedural way.

Services

It is possible to extend the functionality and behavior of a module at runtime by defining a service within it and defining the wrapper methods associated with it.

A service is a particular entity characterized by the fact that it is contained within a module, to which it uniquely connects. In iEngine, each service must extend the abstract BaseService class by implementing its related abstract functions.

The IBaseService interface defines all the fields and methods that must be present within a generic iEngine service. The Enabled field indicates whether the service is enabled or disabled. To enable a service, the Enable method must be called, while to disable a service, it is possible to use Disable.

To connect a service to a module, it is necessary to call the AddInput function, while RemoveInput must be called to remove it. Each service communicates via the Module Connection System only with other modules and not with the services. The communication can be of three types: Input only, Output only, and Bidirectional (both Input and Output). The communication takes place by calling the SendMessage method as for modules. All the messages that are sent to the module are then forwarded to all the services of the module itself.

Plugins

If it is necessary to define generic behaviors and / or functions not directly identifiable with a specific module, it is possible to implement plugins.

A plugin is a particular entity very similar to services but which is not contained directly within a module. Each plugin can be connected to one or more modules. Each plugin communicates via the Module Connection System only with modules and not with the services or other plugins. The communication mode with the module can be of three types: Input only, Output only, and Bidirectional (both Input and Output).

Any plugin extends the abstract BasePlugin class by implementing its related abstract functions. The IBasePlugin interface defines all the fields and methods that must be implemented in a generic iEngine plugin.

Packages

An iEngine package is a type of compound entity. It is created by assembling modules, services, and plugins to guarantee a predefined set of functions and behaviors. Each package of iEngine extends the BasePackage class by implementing its related abstract functions and specifying internally all modules, services, and plugins included and how they are connected. Therefore, a package can offer all the services implemented by modules.

Within an application, only one instance of a package at a time must be loaded. The IBasePackage interface defines all the fields and methods that must be present within a generic iEngine package. When a new instance of BasePackage is created, by default, it is disabled. Enabling or disabling a package involves activating or deactivating all the modules, services, and plugins inside it.

Inside a package, the modules, the services, and the plugins communicate through the Module Connection System, while to communicate with the external world, it is necessary to define an input module and an output module in the package. A package can also communicate with another package.

24.4.3 Novelty of the Developed Approach in iEngine

The novelty of iEngine resides mainly in the fact that the used approach can be replicated for different applications sharing the common modules and services. Typically, only the application logic module should be implemented, while all the remaining part can be obtained by just connecting in the right way the other modules. Of course, more applications will be developed with this approach, more modules will be implemented for iEngine, and less work will need to be done to develop new applications based on it.

With this point of view, iEngine can be considered much more than a VR Toolbox, but it can be considered as a complete, modular, and expandable framework and a general architectural paradigm that can be reused many times, maximizing reuse and obviously reducing not only development but also design times.

24.5 Use Case and Experimentation

To facilitate maintenance, two phases are important: a training phase and an "on-site" support phase.

The first phase in which the training is carried out could be assisted by the use of an immersive virtual reality environment where the maintenance technician can interact with the virtual clone of the system subject to maintenance.

The second phase takes place directly in the place where the system is installed to support and to help with the maintenance procedures. In this case, mixed reality comes into play, which is well suited to the purpose; in fact, the maintenance procedures to be performed can be virtually superimposed on the real objects, and the maintenance technician can "simply" follow them step by step independently from his specific knowledge of the system and avoid having to consult the manual from time to time.

To implement this use case, it has been developed the virtual training environment (VTE) system that represents a framework for training and assisted maintenance of complex equipment.

The use of the training and training scenarios takes place in VRE (virtual reality environment) mode. In this mode, the user can move freely in the real environment, and the display takes place via a special stereoscopic VR viewer (HTC Vive), able to show the user a completely virtual space within which there is the asset subject of the training activity.

The use of assisted scenarios during maintenance operations on the real object takes place in MRE (mixed reality environment) mode. In this mode, the user is in front of the device to be serviced, and with a special mixed reality viewer (Microsoft HoloLens), after the maintenance procedure to be carried out has been activated, he will see virtualized, overlapping the real elements, the different parts subject to maintenance and the different animations that will indicate the correct disassembly or assembly procedure.

From a software architecture point of view, VTE consists of three software applications: VR, MR, and Editor (Fig. 24.11).

In this specific case, we will focus exclusively on the MR application of the assisted maintenance (Fig. 24.12).

The workflow that allows you to make the most of all its functions is schematically represented in Fig. 24.8:

- The Administrator creates the Instructor and Maintainer users.
- The Instructor imports the 3D models, creates the procedures, and configures the VR training and MR maintenance environments.
- The Maintainer uses the two environments of training VR and maintenance MR.

The platform supports three categories of users:

- *The Admin*: is the system administrator who manages the Asset Repository and, therefore, modifies it in order to keep it updated by loading, removing, or modifying the various elements present within it.
- *The Instructor*: is the one who manages the authoring of the application and imports the 3D models, creates the procedures, and manages the Maintainer user.
- *The Maintainer*: is the user that follows the training via the VR application and uses the MR application as a guide during the maintenance phases. In the architecture graph, it is shown twice because it performs a double interaction: the first is to provide input, and the second is to take advantage of the generated outputs. In fact, when the Maintainer uses the application, through the use of different peripherals, he sends input data to the application. These input data are the information used by the application to understand the way in which the user is interacting with the application and to react according to the stimuli.

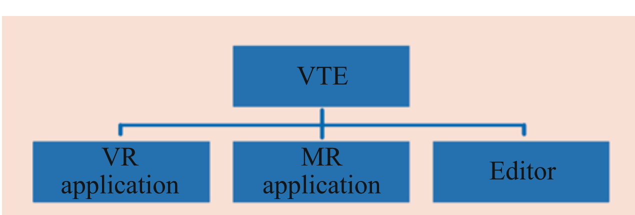


Fig. 24.11 VTE software architecture

The VTE platform was implemented starting from iEngine framework using three packages that together cover the whole architecture:

- *Authorial package*: includes the modules that allow the Instructor to configure the application, through one or more configuration files.
- *Content Management package*: includes modules that allow the loading and management of both internal and external contents of the application. These contents are kept in the Asset Repository and managed by the Asset Manager.
- *Fruition package*: includes the modules that allow the use of the application itself.

The head-mounted display (HMD) and the monitor are the output devices of the application. The HMD is the VR or AR device that the user uses to view the virtual or augmented world, while the monitor is used both to allow the operator to control what is happening and to give anyone who is assisting the possibility to view the same scene visible to the user. Both receive the scene rendered from the top layer via the Asset Renderer.

The *Asset Repository* hosts all the assets that the Admin can load using the Asset Loader. It gives the possibility to load, remove, or modify all the assets present in the repository.

The *Depth Camera*, the *controllers*, and the *trackers* are the input devices of the application. The information received from these devices are all addressed to the upper layer in the Input Interface.

The *Web Server* is the server that contains all the information shared with the App and the Website.

The *Config file* is the physical file on the disk through which the operator can configure the application. It is managed by the Config Interface of the upper layer (Fig. 24.13).

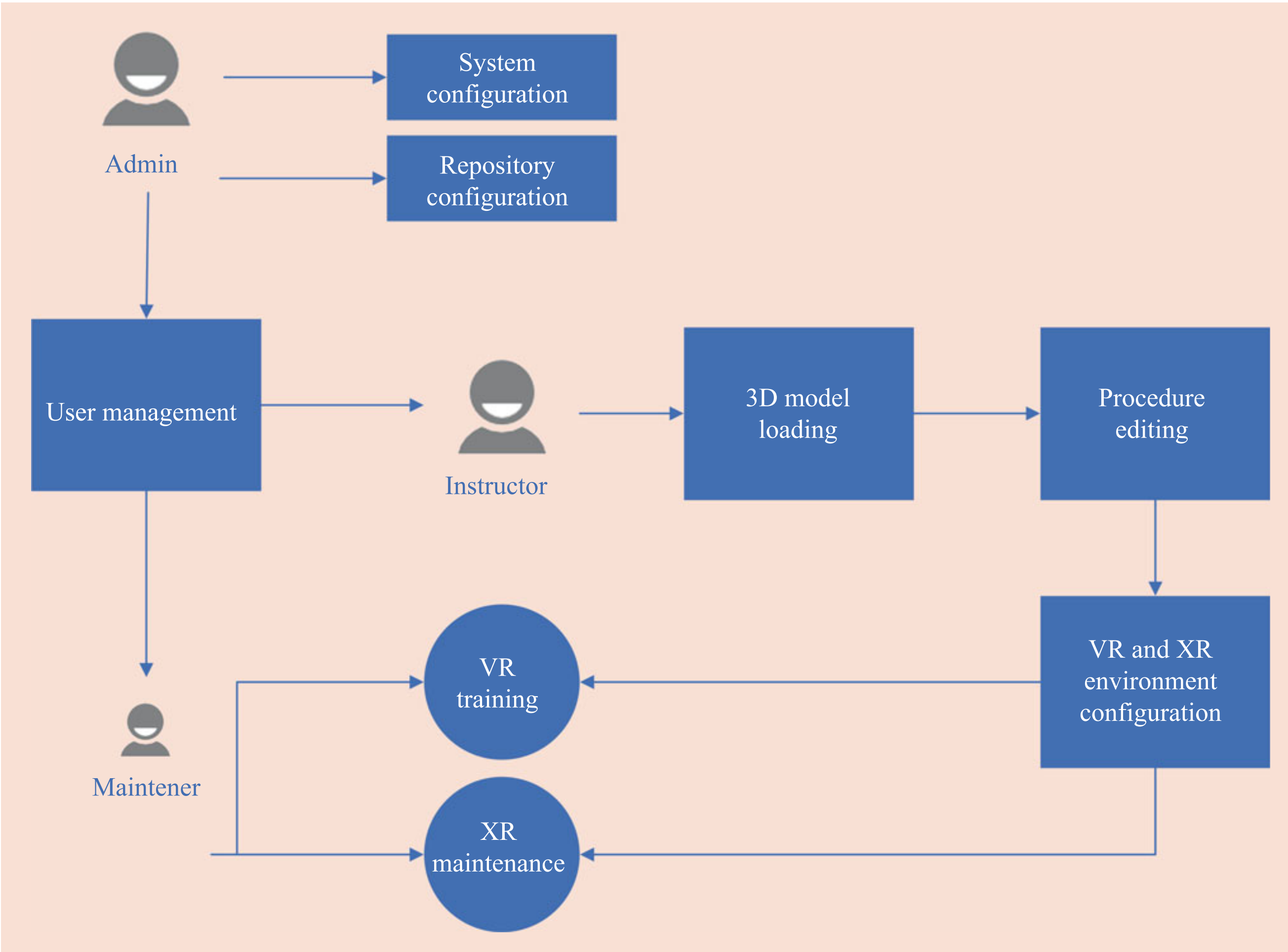


Fig. 24.12 VTE workflow

24.5.1 VTE Modules

Hereinafter are listed and described the iEngine modules used by VTE.

The *Asset Renderer* is the module that reconstructs the scene and all the components inside it, which are then sent to the viewer (HMD) and/or camera of the application and enjoyed by the user. It receives all the information that it needs from the top layer thanks to the Asset Manager.

The *Asset Loader* is the module that allows the Asset Manager to load a specific asset, which can be a new model or even an entire scene, in order to take advantage of a modularity of the application and have the possibility to expand it without the need to reinstall everything.

The *Asset Manager* takes care to create, modify, or delete objects, lights, sounds, etc., in order to create the final scene.

The *Input Interface* is the interface connected with the input peripherals such as the depth camera, the controller, and the trackers of the Hardware Layer that send all the information relating to the user inputs.

The *Web Interface* is the interface that receives and sends information to the Web Server of the lower layer and acts as an intermediary with the managers of the user profiles and metadata of the upper layer.

The *Config Interface* is the interface that receives the changes made to the Config file by the Instructor and that sends the data to the Config Manager.

The *Input Manager* translates the information that comes from the inputs into interactions that can be managed by the application and communicates these information to the Application Logic module.

The *Profile Manager* takes care of managing the profile of the registered user. It stores and manages all configurations, information, and user preferences. The Profile Manager communicates with the Web Interface to receive and/or send user data from/to the web portal. The data received and related to personalization are sent to the Application Logic module.

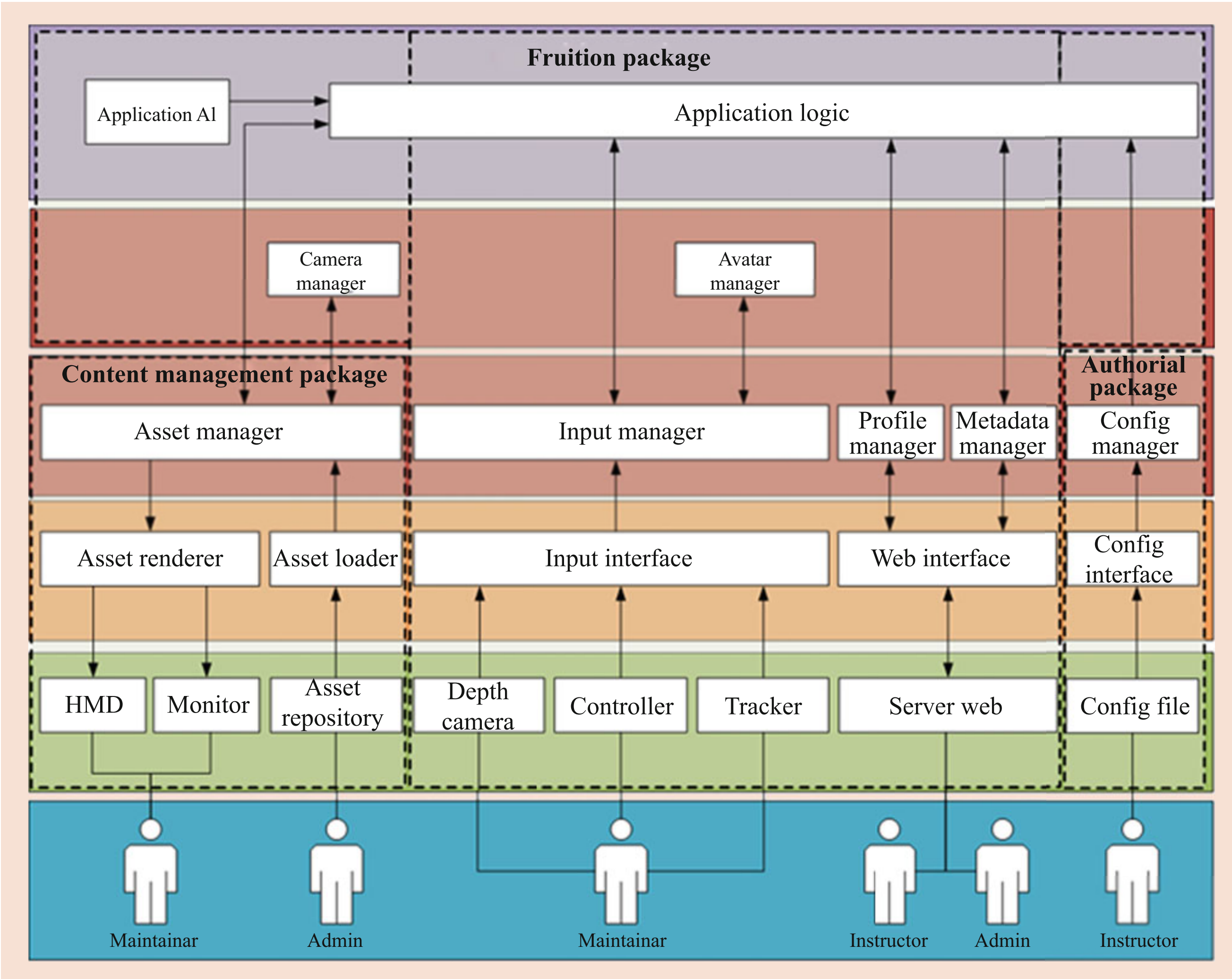


Fig. 24.13 View of the packages in iEngine architecture

The *Metadata Manager* receives the metadata from the web portal. The metadata contains information relating to the scene or to the objects in the scene.

The *Config Manager* manages the configurations done by the Instructor and read from the configuration files via the Config Interface. These information are processed by the manager who then communicates them to the Application Logic module.

The *Avatar Manager* (used only for VR application) is the module that manages the user's avatar, which is the digital representation of the user in the application. It receives information from the Input Manager, which indicates the movements to be performed and how to perform them, and from the Profile Manager, where the customization of the avatar is registered.

The *Camera Manager* is the module that manages the camera of the scene in the application. It receives its inputs from the Input Manager.

The *Application AI* module is the artificial intelligence module that allows making decisions for the management of behaviors, animations, and movements of the object according to the actions performed by the user. Furthermore, it receives inputs from the Application Logic module that could influence the decisions on the behaviors. The Application Logic module deals with the management of all application logic and manages the logic of events generated by user actions.

24.5.2 VTE Application AI Module

The decision system, as in most of the video games currently produced, is based on:

- Finite state machines
- Decision trees

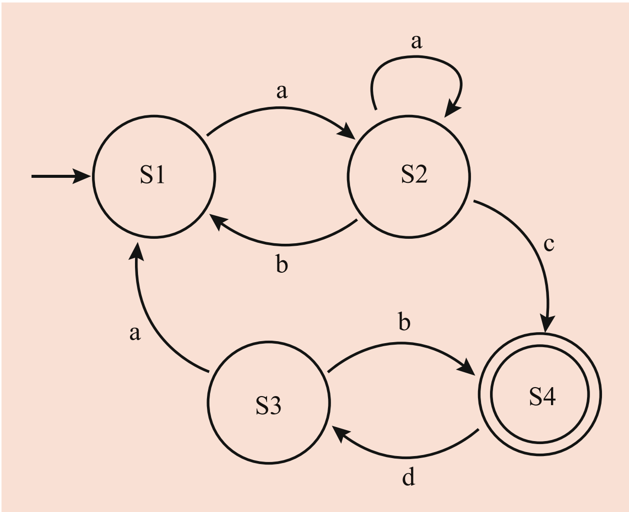


Fig. 24.14 Example of a finite state machine

Finite State Machines

Finite state machines are one of the least complicated and, at the same time, most effective and most used methods for programming artificial intelligence. For each object, it is possible to distinguish a number of states identifying a moment of its life.

Depending on their state, objects respond differently to a finite set of external stimuli. The finite state machine method allows dividing the behavior of each application object into smaller fragments, in order to facilitate its implementation, debugging, and extension.

Each state has the code responsible for managing incoming or outgoing state transitions and a code for processing and interpreting messages from the environment (Fig. 24.14).

There are two types of finite state machines (FSM): deterministic finite state machines, often called deterministic finite automata, and nondeterministic finite state machines, often called nondeterministic finite automata. These vary in the way they are represented visually, but the basic idea is the same.

Decision *Trees*

A way to implement a decision system is to use decision trees . Decision trees are easy to interpret even when they describe very complex systems. They are mainly used in the management of non-playable characters in video games and Application Logic and for the control of animations.

They have the advantage of presenting a tree-shaped structure and are composed of connections and decision points (tree nodes) (Fig. 24.15).

The tree starts from a starting decision called root (Q1 in Fig. 24.8), and for each decision (each node), it is possible to select one from a set of possible options.

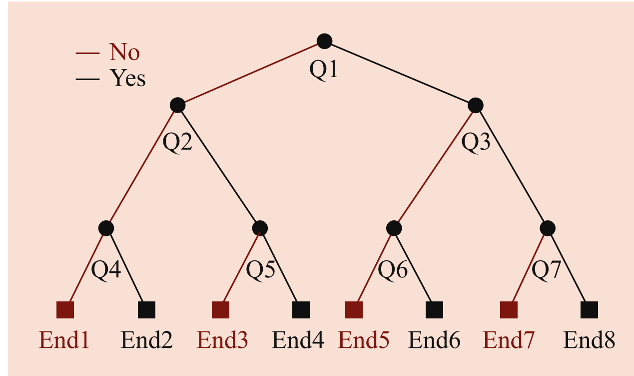


Fig. 24.15 Example of a binary decision tree

24.5.3 VTE Final Result

The result reached with the VTE framework is visible in the below figures. The figures show some overimposed information that are visualized during the maintenance operation of a rack.

The first figure shows where and how to connect an Ethernet cable on the switch, and the latter shows the status information of all devices installed in the rack (Figs. 24.16 and 24.17).

The implementation of VTE based on iEngine framework required much more time than the case without the framework. The additional effort estimated was about 30%. The reasons of this are due to the fact that it was necessary to develop a certain number of new modules and services in addition to some plugins to implement all the requirements of VTE. However, while we are writing this article, a new application is following this approach, and this application is benefiting of thousand lines of code written in the already developed services and modules. The benefits, foreseen for the new application, are estimated in a time reduction (considering the time spent in design development and testing) of around 50%. We are confident that in the future developments, it will be possible to reach reduction of the time of about 70% along the entire life cycle.

24.6 Conclusions

AR-based guidance, maintenance, and repair are rapidly becoming common in many industrial areas, also due to the push of the Fourth Industrial Revolution. According to some research, more than 80% of the companies that currently use this technology claim to have found benefits equal to or greater than their expectations. However, the lack of internal skills and the lack of sufficient back-end infrastructure represent strong barriers to growth. It also revealed that 50% of companies that have not yet implemented augmented and

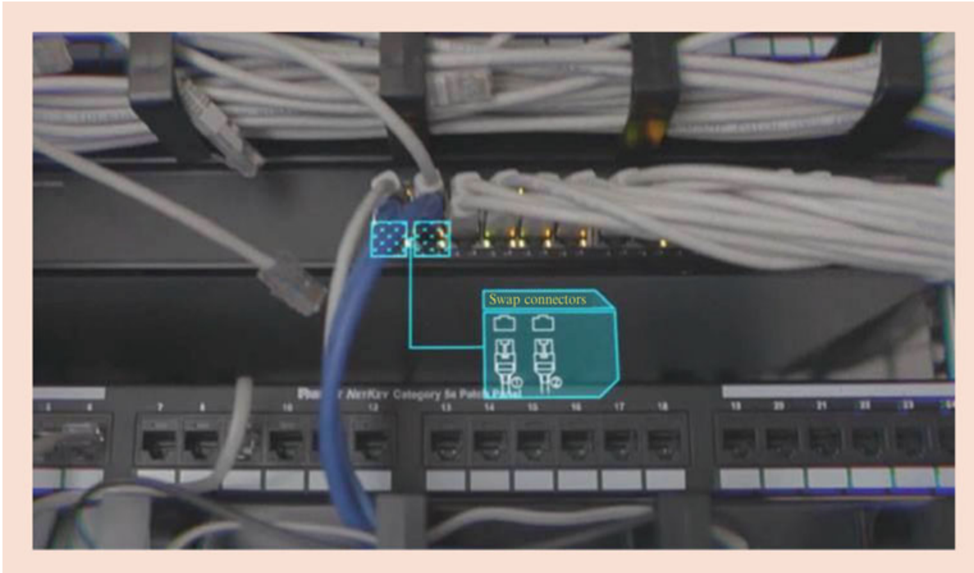


Fig. 24.16 Example of VTE in mixed reality

Fig. 24.17 Another example of VTE in mixed reality

virtual reality solutions are planning to start exploring their potential within their operations within a time horizon of 3–5 years. It is significant to note that, although augmented reality is more difficult to implement, companies consider it more advantageous than virtual reality: AR in fact generates benefits in terms of productivity thanks to the simplification of workflows. While VR has proven capable of improving a single immersive user experience isolated from reality, AR connects the digital world with the real world, thus supporting a greater number of innovative use cases. In this scenario, some critical issues remain which have hindered so far and could still slow down the process of capillary diffusion of AR in the industrial context. A clear example of these critical issues is represented by the preparation of the industrial technological infrastructure for the integration of AR/VR solutions: the lack of data and technologies to be used immediately constitutes a serious obstacle to the adoption of augmented reality. Furthermore, it is very important to identify the use cases for which the implementation of the AR results in a tangible and lasting advantage for the operators who represent its main users. Finally, for many industrial areas, a fundamental challenge is to ensure uniform integration with existing technologies and corporate culture.

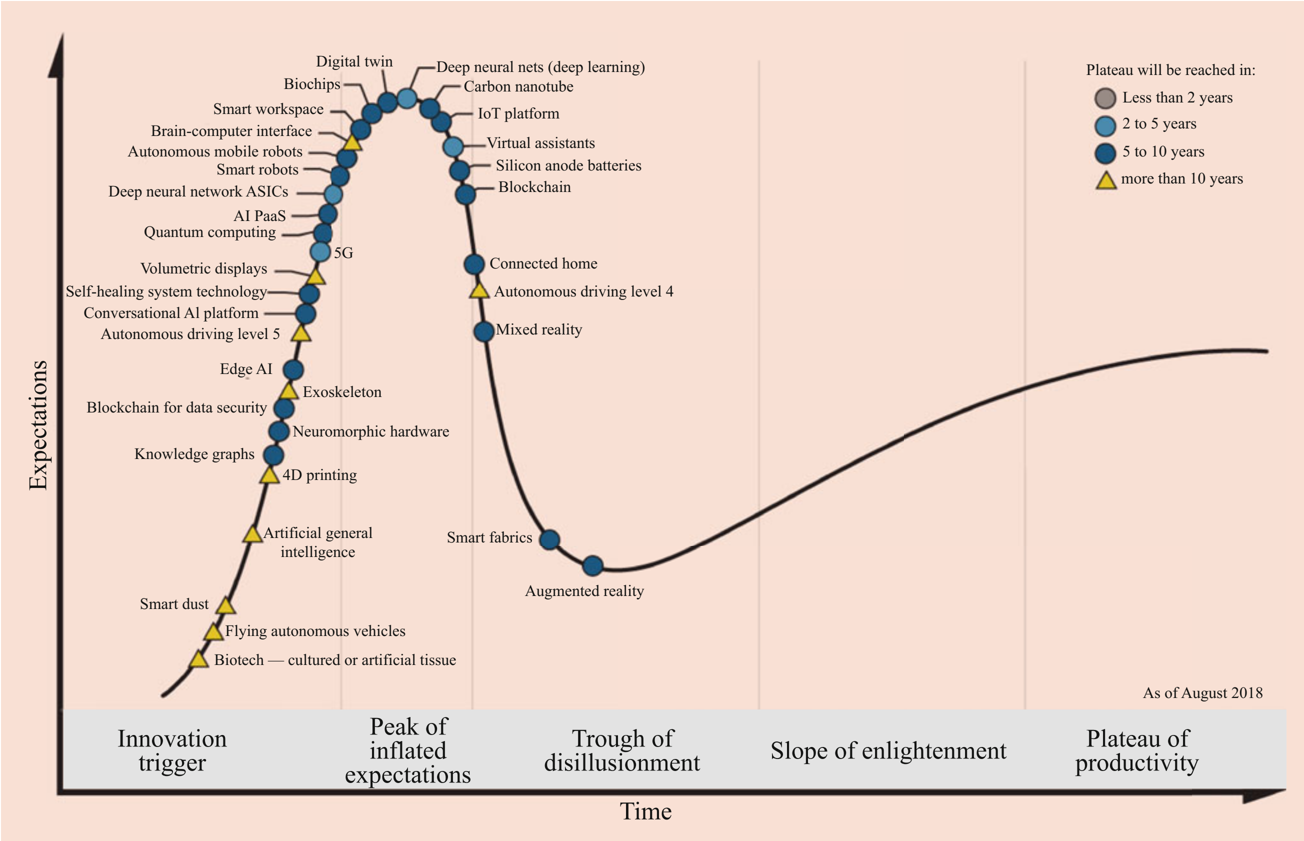


Fig. 24.18 Gartner Hype Cycle for emerging technologies (2018 [27])

AR can therefore be introduced in companies to solve the most common problems regarding the maintenance and repair of systems and to make certain activities for technicians working on the field less risky. Augmented reality is a technology that until a few years ago was seen as a game, but which today has reached a level of maturity that is ready for use in many production environments.

In fact, according to the Hype Cycle forecasts of the information technology research company Gartner [26], it is noted that in the current state of emerging technologies for 2019, there is no longer AR and MR. This means that for Gartner, AR and VR are now both mature technologies (Fig. 24.18).

Finally, from the video concepts released on YouTube [28], it is confirmed that augmented reality is already sufficiently ready to be introduced into technical Maintenance, Repair, and Overhaul (MRO) networks, as tools for daily use by the field technician. There is therefore a strong interest of IT managers and service managers in understanding the opportunities for applying this technology for the maintenance and customer support of complex equipment and systems, both in the B2B and consumer sectors. Many applications based on augmented reality have been studied by academics and industry for many years now.

Acknowledgments This work and the research behind it would not have been possible without the non-profit support of Raffaele Vertucci and Salvatore D'Onofrio who shared their knowledge and experiences on the VR/AR technologies gained in their company Leonardo S.p.A, a leading company in the aerospace industry.

References

1. Chang, Y.J., Liu, H.H., Kang, Y.S., Kao, C.C., Chang, Y.S.: Using augmented reality smart glasses to design games for cognitive training. In: Remote Engineering and Virtual Instrumentation (REV), 2016 13th International Conference on, pp. 252–253. IEEE (2016)
2. https://doi.org/10.1016/j.compind.2013.01.002
3. Thielman, S.: HoloLens: the virtual reality headset for elevator repair. The Guardian (September 19, 2016). https://www.theguardian.com/business/2016/sep/19/hololens-the-virtual-reality-headset-for-elevator-repair

4. Erickson, S.: Microsoft HoloLens enables thyssenkrupp to transform the global elevator industry (September 15, 2016). https://blogs.windows.com/devices/2016/09/15/microsoft-hololens-enables-thyssenkrupp-to-transform-the-global-elevator-industry/
5. http://www.mwpowerlab.com
6. Mulder, J.D., Boschker, B.R.: A modular system for collaborative desktop VR/AR with a shared workspace. In: Virtual Reality Conference. IEEE (2004)
7. https://www.google.com/glass/start/
8. https://www.microsoft.com/it-it/hololens
9. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. Visual. Comput. Graph. IEEE Trans. **17**(10), 1355–1368 (2011)
10. de Souza Cardoso, L.F., Mariano, F.C.M.Q., Zorzal, E.R.: A survey of industrial augmented reality. Comput. Ind. Eng. **139**, 106159 (2020)
11. Zhiqiang, B.I.A.N., Ishii, H., Shimoda, H., Yoshikawa, H., Morishita, Y., Kanehira, Y., Izumi, M.: Development of a tracking method for augmented reality applied to NPP maintenance work and its experimental evaluation. IEICE Trans. Inform. Syst. **90**(6), 963–974 (2007)
12. Fiorentino, M., Uva, A.E., Monno, G., Radkowski, R.: Natural interaction for online documentation in industrial maintenance. Int. J. Comput. Aided Eng. Technol. **8**(1–2), 56–79 (2016)
13. Lima, J.P., Roberto, R., Simões, F., Almeida, M., Figueiredo, L., Teixeira, J.M., Teichrieb, V.: Markerless tracking system for augmented reality in the automotive industry. Expert Syst. Appl. **82**, 100–114 (2017)
14. Alvarez, H., Aguinaga, I., Borro, D.: Providing guidance for maintenance operations using automatic markerless augmented reality system. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 181–190. IEEE (2011)
15. Aouam, D., Benbelkacem, S., Zenati, N., Zakaria, S., Meftah, Z.: Voice-based augmented reality interactive system for car's components assembly. In: 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS), pp. 1–5. IEEE (2018)
16. Wei, W., Yue, Q., QingXing, W.: An Augmented Reality Application Framework for Complex Equipment Collaborative Maintenance, vol. 6874, pp. 154–161. Springer, Berlin/Heidelberg (2011). ISBN: 978-3-642-23733-1
17. Mistry, P., Maes, P.: SixthSense: a wearable gestural interface. In: ACM SIGGRAPH ASIA 2009 Sketches, p. 11. ACM
18. Park, G., Ha, T., Woo, W.: Hand tracking with a near-range depth camera for virtual object manipulation in an wearable augmented reality. In: International Conference on Virtual, Augmented and Mixed Reality, pp. 396–405. Springer, Cham (2014)
19. Bai, H., Lee, G.A., Billinghurst, M.: Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand, pp. 126–131 (2012)
20. Buchmann, V., Violich, S., Billinghurst, M., Cockburn, A.: FingARtips: gesture based direct manipulation in Augmented Reality. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 212–221. ACM (2004)
21. Abate, A.F., Narducci, F., Ricciardi, S.: Mixed reality environment for mission critical systems servicing and repair. In: International Conference on Virtual, Augmented and Mixed Reality, pp. 201–210. Springer, Berlin/Heidelberg (2013)
22. Abate, A.F., Loia, V., Nappi, M., Ricciardi, S., Boccola, E.: ASSYST: Avatar baSed SYStem maintenance. In: 2008 IEEE Radar Conference, pp. 1–6. IEEE (2008)
23. Bozzelli, G., Raia, A., Ricciardi, S., De Nino, M., Barile, N., Perrella, M., … Palombini, A.: An integrated VR/AR framework for user-centric interactive experience of cultural heritage: the ArkaeVision project. Digit. Appl. Archaeol. Cult. Herit. **15**, e00124 (2019)
24. Pagano, A., Palombini, A., Bozzelli, G., De Nino, M., Cerato, I., Ricciardi, S.: ArkaeVision VR game: user experience research between real and virtual Paestum. Appl. Sci. **10**(9), 3182
25. Popolo, V., Di Nardo, M., Di Leo, R., De Cristofaro, A., De Nino, M.: Wellbeing and Smart Working in the new industry era. XXV Summer School "Francesco Turco" – Industrial Systems Engineering
26. Gartner Hype Cycle. https://www.gartner.com/en/information-technology/research/hype-cycle
27. Hype Cycle for Emerging Technologies: (2018). https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018
28. https://www.youtube.com/results?search_query=augmented+reality+maintenance

24

Salvatore D'Onofrio received the PhD in Computer Science from the University of Sannio in 2007. He works now as CTO member at Leonardo company in Italy. He is mainly involved in R&D activities focused on Virtual/Augmented Reality technologies applied to training and maintenance, Modeling & Simulation and Predictive Analysis for maintenance activities.

Raffaele Vertucci Graduated with honors in Electronic Engineering, Raffaele Vertucci gained his experience working in the aerospace industries. As CTO he has been involved in research and development activities related to image processing, computer vision and virtual/augmented reality technologies, with particular focus on their application for support, maintenance and training, writing articles, managing research programs, making dissemination activities, registering a patent and developing software tools.

Stefano Ricciardi received the PhD in Sciences and Technologies of Information, Complex Systems and Environment from the University of Salerno. Main research interests include biometry, virtual and augmented/mixed reality, haptics systems and human-computer interaction. He is currently a research fellow at the University of Molise, member of IEEE and GIRPR/IAPR. and co-author of more than 80 peer reviewed research papers.

Maurizio De Nino Graduated with honors in physics with a specialization in artificial intelligence and data acquisition system, for 20 years he did his experience working for scientific centre and hitech company with a specific experience in R&D for real-time embedded system, image processing and virtual reality. In the last 10 years, he is working with Digitalcomoedia as CTO for virtual and augmented reality system development.

Check for updates

Augmented Reality for Naval Domains

25

Mark A. Livingston, Joshua Kvavle, Corey J. Countryman, Kendy K. Vierling, Nathan Jones, Michael J. Confessore, Ryan Bowersock, and Aristides Staikos

Contents

25.1 Historical Review of Selected Research 617
25.1.1 Implementations of the Head-Up Display Concept 618
25.1.2 Maintenance and Repair Applications 620

25.2 Research and Development Efforts 621
25.2.1 Dismounted Infantry Training 621
25.2.2 Other Applications to Military Operations 627
25.2.3 Maintenance and Repair 628
25.2.4 Other Applications and Technologies 631

25.3 Discussion 635
25.3.1 Augmented Reality in the Navy: A Roadmap 635
25.3.2 Human Factors Engineering Technical Advisory Group 635
25.3.3 Lessons Learned 636

25.4 Conclusions and Future Directions 638

References 638

M. A. Livingston (✉)
Naval Research Laboratory, Washington, DC, USA
e-mail: mark.livingston@nrl.navy.mil

J. Kvavle
Naval Information Warfare Center, Pacific, San Diego, CA, USA
e-mail: kvavle@spawar.navy.mil

C. J. Countryman
Naval Undersea Warfare Center Division Keyport, Keyport, WA, USA
e-mail: corey.countryman@navy.mil

K. K. Vierling
Naval Education Division, Office of the Assistant Secretary of the Navy (Manpower and Reserve Affairs), Washington, DC, USA
e-mail: kendy.k.vierling.civ@us.navy.mil

N. Jones
U.S. Marine Corps Systems Command, Training Systems, Quantico, VA, USA
e-mail: nathan.jones1@usmc.mil

M. J. Confessore · R. Bowersock · A. Staikos
Naval Air Warfare Center, Aircraft Division, Patuxent River, MD, USA
e-mail: michael.confessore@navy.mil; ryan.bowersock@navy.mil; aristides.staikos@navy.mil

Abstract

Since the pioneering work of Ivan Sutherland and the success of Tom Furness' "Super Cockpit," augmented reality (AR) has been an interest of the US Department of Defense (DoD) for some time. As hardware advances enabled more ambitious applications, the US Navy and US Marine Corps (USMC) have carried this research forward into new application areas and into development programs to explore the practical potential of the technology to assist in accomplishing tasks required in Naval applications (including land-based operations of USMC). We review a selection of AR research and development conducted by the US Navy, ranging from basic research to prototype development. We draw lessons from the progression of projects through the R&D spectrum and explore common threads through varied applications. Challenges that currently limit the application of AR technology to naval domains are construction of environmental models that enable integration of the virtual and real elements, ability to operate in mobile scenarios, and usability concerns such as information overload.

Keywords

Military augmented reality · Mobile augmented reality · Military training · Situation awareness · Information overload · Human factors

25.1 Historical Review of Selected Research

As early as 1980, Newman [1] attributed the development of the head-up display (HUD) directly to the development of reflective gunsights in World War II. He describes the use of projection onto a semitransparent mirror to optically combine the view through the aircraft windshield with the virtual imagery. He further described the advantages of such

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_25

a system through the ability to compensate for various factors in aiming weapons and for environmental factors such as ambient brightness. Weintraub and Ensing [2] quoted a 1946 report titled "Psychological Requirements in Aviation Equipment Design," by Paul M. Fitts [3], for the US Army Air Forces Aero-Medical Laboratory at Wright Field; Weintraub and Ensing believe the following requirement was the first published conception of the HUD:

> *Displays which provide related information, or which are referred to in rapid succession should be grouped.* It is the opinion of the pilots, supported by some research data, that associated instruments should be located close to each other to minimize eye movements and fixation time. It is especially important to make it easier for the pilot to look out of the cockpit and back again at the instruments. One factor that may contribute to vertigo is the movement of the head required in some airplanes when looking from one instrument to another. Proper grouping would eliminate this condition. It has been proposed, for example, to throw the image of certain instruments onto the wind screen so that they can be viewed while looking out of the plane. [italics as in [3, p. 272] and [2, p. 1]]

Despite the obvious connection between the concept as described by Fitts and the early implementations of see-through displays, there is no evidence of Fitts' description motivating the early development of see-through displays [4]. We do note that Weintraub's work was cited by Ellis as an influence on his work at NASA Ames Research Center, through published material [5], earlier work at NASA [6] on HUDs by Dick Haines, and personal interaction [7]. But for more direct technical influence, we must look elsewhere.

Newman [1] and, later, Weintraub and Ensing [2] noted the pioneering work of J. Michael Naish, first for the British Royal Aircraft Establishment and later for Douglas Aircraft. In a retrospective introduction to a review of HUD formats [8], Naish noted the expanding definition of the term. He attributed his early work for the (British) Royal Aircraft Establishment, starting in 1956, to a requirement to reduce the demands created by an attentional shift for the pilot from the instrument panel to the forward view out of the aircraft. Despite the shared motivating statement, there was no mention of this observation by their US counterparts. Naish and his team rejected a design using a small television to replicate the forward view within the instrument panel because of the limitations of field of regard (seeing in all directions from the aircraft), the cognitive challenge created by the displacement of the camera position from the pilot's eye, as well as limits of resolution, field of view, color, and frame rate of the equipment then available for the camera and television. This led to the design "using a reflecting collimator, as in a weapon sight." Then (as would be reflected in much later research), remaining issues included losses in brightness and edge effects from the partially reflective plate. Tests were done in 1958 in real aircraft and then in 1960 using simulators. None of this appears to have had direct influence on the work of Sutherland, however.

Would these systems fit the now-accepted definition [9] of AR? There are three components to the definition. The system must combine real and virtual; certainly, the projection of the graphical displays onto the cockpit windshield qualifies these systems as meeting this requirement. The system must be interactive in real time. Since these graphics updated with the movement of the aircraft, they satisfy this requirement as well. This is a lower form of interactivity than many modern AR systems, however. The final definition component is that the system must be registered with the 3D world. Most of the early systems rendered the horizon line at the proper angle given the roll angle of the plane with the ground [10]. Therefore, they would qualify as meeting this final component of the definition as well, albeit with a minimalist capability. This gives a succinct argument that such a HUD meets, in however minimal a fashion, the academically accepted definition of AR.

We believe this history introduces two themes that are the focus of the current chapter. First is the strong connection between AR and the research and development for defense applications of the technology. Second is the importance of human factors to research and development of AR. In the remainder of this section, we review some research that supports these two themes in work for Naval and other defense-relevant applications.

25.1.1 Implementations of the Head-Up Display Concept

In work widely recognized as the first immersive display using imagery generated by computer, Ivan Sutherland [11, 12] built a head-worn immersive display system with support from the Defense Advanced Research Projects Agency (DARPA). The team needed several other hardware components in order to build a complete system. They designed and constructed a special-purpose graphics processor to determine exactly which portions of lines were visible in the perspective view of the user. This hardware was coupled with two attempts at tracking (mechanical and ultrasonic). The resulting system enabled the user to control the view of the surrounding 3D imagery by simply moving in relation to it. This work is widely cited in the literature as the beginning of AR (and virtual environments). But it is most interesting for the goals of this chapter to note the two "peculiar" phenomena noted in that work [12]. First was the challenge of conveying accurate depth relationship among wireframe objects. While hidden line removal algorithms existed, they were too expensive for early implementation. As noted, special-purpose hardware overcame this challenge. This foreshadowed one of the important perceptual challenges that will be noted for modern applications later in this chapter and elsewhere in this volume. Second was the challenge of understanding the shape of a molecular model; this was attributed to the inability to get the best viewing angle (top-down, to see the hexagonal shape of that particular molecule). That angle was not physically realistic and thus not available. But later AR work would overcome this by enabling this physically unrealistic but important viewpoint.

These challenges in understanding the geometric reality of the information presented virtually leads to an important concept for the work presented here. One general goal for AR in military contexts is providing what is known as situation awareness (SA) [13], an understanding of the environment and the entities within it. SA yields an understanding of past, current, and potential future events. This concept may be extended to multiple applications; here, we present three projects with relationships to military applications.

Aircraft Pilot

Tom Furness and others began research [14] on what eventually became known as the Visually Coupled Airborne Systems Simulator, using helmet-mounted sights shown on a "see-through display," in 1965. This work was done at the renamed US Air Force Research Laboratory at Wright-Patterson Air Force Base, which grew directly from the US Army Air Forces Aero-Medical Laboratory at Wright Field, under which Fitts did his work. Furness did cite a technical report by Fitts and others presenting data on the eye movements of pilots, but does not (to our knowledge) directly cite the above quote from Fitts as an inspiration. Birt and Furness [15] do mention the work done under the Army-Navy Instrumentation Program, which included Douglas Aircraft (though not mentioned by Birt and Furness). The latter company had hired J. Michael Naish to continue his work. So there is at best weak evidence for the connection – from either the early design concept or the early implementations of optical gunsights – reflected in a HUD implemented in the computer era of see-through displays begun by Sutherland and Furness.

Whatever the source of the inspiration, the argument for the aircraft pilot display seems to be the same. Visibility out of a cockpit is limited, and airborne tasks such as low-altitude navigation, target acquisition, and weapon delivery require pilots to reference landmarks on the terrain. However, sensors mounted on the aircraft can create visibility in areas that are occluded by the aircraft structure, or in conditions such as low light that prevent the pilot from seeing the real world.

A later system named the "Super Cockpit" superimposed flight and target data into the pilot's visual field (not unlike the early gunsights) and provided sound cues to assist localization. The key advantage of the Super Cockpit was providing spatial awareness, which is a component of SA. The pilot could understand a variety of incoming data streams and incorporate the data into his course of action. The horizon became visible through the cockpit window, rather than on an indicator in the instrument panel. Targets, navigation waypoints, and threats could similarly be registered to their 3D locations. The concept was (as Naish had earlier concluded) that such a view would improve upon using a dashboard display. The dashboard display of information required the pilot to mentally merge the virtual map with his visual field. Such merging is not an easy visualization task and would have required the pilot to take his eyes off the real environment many times in order to align virtual information. Another feature in the Super Cockpit was a rear-view mirror, analogous to the standard mechanism in a car.

Early versions of the system pointed out the need for study of the human factors [16] of such systems. Spatial reasoning is a complex task, even more so under the physical duress of flight and the emotional intensity of combat. An intuitive interface could take advantage of the natural abilities of many people to reason in three dimensions, rather than have them reason in two dimensions and try to apply that to the surrounding 3D environment. This 3D spatial reasoning is also not a trivial task, but pilots are screened for high spatial reasoning ability, so it seems natural to supply them with an inherently 3D view.

Ship Captain

The idea of a HUD is not confined to an airplane cockpit; the same concept may be applied to the bridge of a ship. Benton [17] described a system called Augmented Reality Visualization of the Common Operational Picture (ARVCOP). The system used a mobile computer and wearable display to present course deviation, a plan view, or a forward pathway. A pilot user study revealed three main sources of limitations in evaluating the system. Environmental factors included the weather (notably wind and surf), which raised the number of subjects needed for statistical evaluation to an impractical number (over 100) or would have argued for tests in simulators. Display issues (limited brightness, low field of view, and focus distance) had a negative impact on performance of the users. Training time appeared to vary widely across participants, with some estimated to need up to a full day to acclimate to the system (more than a pilot study could afford to spend).

The authors found that the registration error of real-world navigational aids to ARVCOP-generated counterparts was observed to be less than 50 yards (46 meters). Display clarity was judged to be adequate, but would benefit from a fully digital process. Differential GPS performed as expected, but a digital gyroscope sometimes failed. Image jitter was considered solved relative to previous prototypes. Icons and text became cluttered at far distances, and "a method for determining depth perception of the images" was recommended. An unresolved question of the level of detail and abstractness of the display was raised. An operational experiment was held in late 2001. The horizontal field of view in this test was 90°, but they set a goal of 360°. Finally, the system was tested and judged favorably for its ability to increase safety in low visibility conditions such as night and fog. This type of application continues to attract research interest in the US Navy. Geoghegan [18] found that experienced Naval officers, most of whom were designated as surface warfare officer, were better able to maintain close proximity to a pre-planned navigation route using a HUD; he also found a high acceptance rate of the concept among those officers.

Dismounted Personnel

Academic research also contributed to the concept of the HUD – for dismounted personnel (i.e., those not in a vehicle). Feiner et al. [19] combined differential GPS and compass tracking technology with a see-through display to create the first example of a mobile AR system. (Here, we define mobility as being in contrast to "vehicle-mounted," whether the vehicle is airborne, seaborne, or ground-based.) Later work would upgrade to real-time kinematic GPS with GLONASS, using a base station situated on a nearby building. (The mobile system was in turn built on earlier work on an AR system to guide a user through maintenance steps of a consumer laser printer. This application domain is explored further in Sects. 25.1.2 and 25.2.3.)

Assembling components onto a backpack frame and adding a stylus-enabled handheld computer allowed a user to have a campus tour guide, dubbed the "Touring Machine" and viewed through a head-worn display. A database of geocoded landmarks connected the user's position to a set of labels about the surrounding environment. Gaze-based selection was implemented using the central ray based on head pose (not using eye tracking). This interaction, through the AR display and touch pad connected to the handheld display, allowed the user to select buildings to learn more about. Later work added a database of departmental web pages and history of buildings in the geocoded locations; these "situated documentaries" enriched the campus tour by adding historical information [20]. This work was funded by the US Office of Naval Research (ONR) and coordinated with the US Naval Research Laboratory (NRL) (cf. Sect. 25.2.1). This system ultimately included a feature that if the user looked straight down (at his or her feet), the view switched to an overhead map, as if a paper map had been placed on the ground at the user's feet. This feature was also added into the application for dismounted military personnel (cf. Sect. 25.2.1), where it was very favorably received.

25.1.2 Maintenance and Repair Applications

A full review of the application of AR technology to maintenance and repair (or other industrial applications) is beyond the scope of this chapter. Readers should see relevant chapters elsewhere in this volume (cf. Chaps. 3, 18, 20, 23, and 24). We also note the existence of multiple reviews in the literature, both recent [21, 22] and older [23]. We highlight one foundational project and one project that was done for the US military.

Aircraft Manufacturing

The term "augmented reality" was coined by Thomas P. Caudell and David W. Mizell for their "application of heads-up display technology to manual manufacturing processes" [24]. This work is widely recognized for being the pioneering effort for the use of AR in a host of industrial applications. By reducing the graphical elements to "simple wire frames, template outlines, designators, and text," they were able to use inexpensive, standard computers of the time. Magnetic tracking and a low-weight head-worn display completed the hardware design.

In one well-known application of the system, an assembly worker followed a virtual analog of a wiring form board. The real form board featured a drawing of how to assemble wire bundles for an aircraft part; the virtual version used bright red lines in the display, registered to a pegboard. The next step in the manufacturing process was connector assembly; end leads of wires were to be inserted into a multi-pin connector. In the real version, a paper "map" indicated which wire was to connect to which pin; the virtual version again used red lines to virtually indicate the correct pin location. The third application required laying precut sheets of fabric in an exact orientation and location; a virtual outline of the sheet was overlaid to indicate the proper assembly. Finally, a maintenance scenario included the removal of a part for service. The part required a precise sequence of twists and turns to remove it without hitting other parts. A wireframe was superimposed on the real part, and then animated to show the 3D egress path.

Multiple hardware issues were noted. It was not truly see-through; it required fusing the real view in the left eye with the virtual image seen only by the right eye. The mount of the display was not rigid, causing registration error. The field of view was only 14°, causing users to need to explore using head movement more than eye movements, as they might normally. Latency associated with the tracking and the interface to the host computer caused further registration error. Despite these, the system was well-received and considered to be a technical success. However, user acceptance was limited [25].

Military Maintenance

Henderson and Feiner [26] developed a prototype system for supporting maintenance procedures under the field conditions experienced by USMC mechanics when working on an LAV-25A1 military vehicle. This is a light-wheeled vehicle with a revolving two-person turret in the middle. A mechanic servicing the turret sits in one of the two seats, with a resulting work area of approximately one-third of a cubic meter. To enable development, a virtual copy of the turret was created with a laser scanner. Five forms of augmented content were presented at various stages of a repair procedure. Arrows in 2D or 3D drew the user's attention to a target or direction to move. Text instructions described the task. Registered labels identified components and their context. A virtual copy of the scene at close range was shown on a 2D screen-fixed panel. Models of tools (in 3D) and components were registered to current or projected locations, as dictated by the procedure and the user's progress. The display was a custom-built stereo video display; two head-worn cameras acquired video. A computer rendered graphics on top of these images and displayed it back to a commercial, stereo head-worn display. A commercial camera-based tracking system using active (LED) markers provided head tracking.

Notably, the authors conducted a user study with a representative population of Marines. The AR display was compared against a 2D interactive electronic manual (that represented the standard technique) and against a HUD version of the software (which controlled for the effects of wearing a head-worn display). Six users completed 18 maintenance tasks. There was a significant difference in completion time between the three display conditions (AR display was fastest), but pairwise comparisons were not significant. Task localization time was also significantly different in the three display conditions, with the AR display condition faster than either of the other two. Display condition did not have a significant effect on errors made in the procedure, as recorded by an observer during the experiment. There were also significant differences observed in head motion while completing the tasks. The 2D electronic manual condition saw a significant loss of focus on the task compared to the other two displays, using a metric of distance from the desired center point of attention. Subjective feedback was generally positive, and some of the criticisms have been obviated by hardware advances in resolution and field of view (of both cameras and head-worn displays).

25.2 Research and Development Efforts

In this section, we summarize several efforts pursued by various agencies within the US Navy and USMC. Only some of these efforts were explicitly coordinated, though some sharing of ideas and results occurred. In the next section, we will draw some lessons learned across these varied projects.

25.2.1 Dismounted Infantry Training

One long-standing focus of Navy and USMC AR research and development is dismounted infantry training. This application is of potential importance to the US Marine Corps and therefore is considered to fall under the category of "Naval applications" within the US Navy and US Marine Corps. This subsection traces the history of much of this work from the research stages through development and testing phases.

Research at the Naval Research Laboratory

NRL functions as a corporate research laboratory for the Navy and USMC. Research on AR began in 1998 with the development of the Battlefield Augmented Reality System. Various summaries of the research exist in the literature [27–31]; we give a brief overview here.

It is interesting to note that BARS was premised on the idea of providing the advantages to dismounted infantry that the HUD had provided to pilots. The challenges associated with urban environments were a particular concern: a complex 3D environment, often a dynamic situation, and the loss of line-of-sight (LOS) contact of team members. Unambiguously referencing landmarks in the terrain and integrating unmanned systems into an operation can also be difficult for distributed users. All of these examples show the impairment of SA during military operations in urban terrain (MOUT). The belief was that the equivalent of a HUD would help solve these challenges. By networking the mobile users together and with a command center, BARS could assist a dispersed team in establishing collaborative SA.

This goal raised numerous issues in system configuration. BARS included an information database, which could be updated by any user. Sharing information across the area of an operation is a critical component of team SA. BARS was equipped with an information distribution system so that these updates would be sent across the network. We enabled BARS to communicate with semiautomated force (SAF) software to address the training issues discussed below. We chose to use commercially available hardware components (Fig. 25.1) so that we could easily upgrade BARS as improved hardware became available. We built many user interface components and auxiliary application suites to simulate the envisioned use of the system. In one example, routes could be drawn on the terrain in a command center application (using a map view, not a 3D perspective view) and assigned to mobile users. Mobile users could draw routes on the map view (engaged by looking down, as described above) and suggest them to commanders or directly to other mobile users. Human factors and usability were always a focus of the system and the research on it [32]. One extensive research investigation was the capability of AR to provide accurate depth perception cues to users [33].

Similar applications of head-worn AR displays [34] or near-to-eye AR displays [35] have been the focus of recent research and development efforts. Both of these projects have some intellectual ancestry in the BARS program, through funding inspired by the success of BARS (respectively) for

Fig. 25.1 Hardware implementations of the Battlefield Augmented Reality System (BARS) in 2002 (front view (**a**); back view (**b**)) and the BARS Embedded Training application in 2006 (**c**). We accepted some bulkiness in the overall design in order to make use of commercially available hardware of the time, as opposed to designing and building prototype hardware

SA and for training. We also note that the BARS application suite included a version for vehicle-mounted personnel [30], although the vehicle-related work was largely unexplored beyond a basic demonstration phase (Fig. 25.2). You et al. [36] survey some of the related research and present perspectives on potential for this line of research, and there are other groups pursuing applications for the military [37]. The BARS application aimed at providing situation awareness was further developed through a DARPA program [34]. The US Army and USMC followed up these efforts with the Integrated Visual Augmentation Systems project [38], which is currently undergoing testing and development [39] in an iterative cycle of user-centered design [40].

Around 2003, NRL was approached by the US Army Simulation, Training, and Instrumentation Command to consider developing a version of BARS to be deployed in embedded training; this meant developing a version of BARS (Fig. 25.1) that could be inserted seamlessly into existing training facilities. MOUT training requires that trainees operate in urban structures, in collaboration with or competing against other live trainees. Often the training uses simulated small-arm munitions and pit instructors against students in several scenarios. Many military bases have small "towns" for training that consist of concrete block buildings with multiple levels and architectural configurations. AR can enhance this training by providing synthetic forces (friendly or opposing) and noncombatants through SAF software.

Using AR for MOUT training is a difficult undertaking from both a technical and acceptance perspective. Many of these challenges are the same challenges encountered when AR is used to provide SA: wearable form factor, accurate tracking indoors and outdoors, and display quality. One unique challenge to using AR for military training operations is that the simulated forces need to give the illusion that they exist in the real world, including disappearing behind real buildings in a MOUT facility and occluding real buildings. One major design decision that resulted from this was to use video-based displays rather than optical see-through. Whereas SA demands that the real world be as visible as possible, training applications need behavioral realism. Commercial optical see-through displays cannot occlude the real world, so BARS was implemented with video-based displays.

One unexpected result in a test of the system was a complaint from users that they felt they could not see well through the commercial video-based display we selected [30]. This led to a research investigation on measuring basic perception through AR displays [41], which we present some of in Sect. 25.2.4. Further interest in the training capabilities came from the Office of Naval Research (ONR) on behalf

Fig. 25.2 An augmented view of the world as seen in a demonstration based on the Battlefield Augmented Reality System. This image conveys several aspects of situation awareness, assimilated (simulated) knowledge from collaborators within the AR ecosystem, and basic information about the world (e.g., compass heading shown at the top)

Fig. 25.3 A variation of the Battlefield Augmented Reality System for the training of forward observers in the task of call for fire

of the USMC; a field demonstration of a training application (Fig. 25.3) for forward observers (FOs) in the task of calling for remote (artillery and mortar) fire was conducted in October 2004 at USMC Base Quantico, with positive reviews from the trainees and instructors who saw the initial prototype, both through the trainees' (head-worn AR) view and through the instructors' control station.

Typical AR system issues like calibration were investigated [42]. Other research centered around issues with the user interface: filtering of information to mitigate the tendency toward information overload [43], object selection [44], and estimating and adapting to registration errors [45].

Development via the Office of Naval Research

Based on this research, in 2009, the Future Immersive Training Environment Joint Capabilities Technology Demonstration [46] took over the task of turning the research project into a prototype AR system for training in the USMC. Initially, ONR's project intended to develop the ability to overlay virtual objects on real terrain to train artillery and mortar forward observers, and the forward air controllers (FACs) that control close air support (CAS) missions for ground units. These tasks involve conveying coordinates for directing fire at targets (and for avoiding friendly units and civilians). Focusing on these types of tasks kept the numerical challenge manageable for the AR system, adding a relatively small number of virtual target vehicles and/or weapons' impacts to the live terrain, vegetation, and infrastructure. Additionally, these tasks generally are in stationary positions most of the time while they are adjusting fire or controlling CAS missions. As with the BARS research system, the initial AR

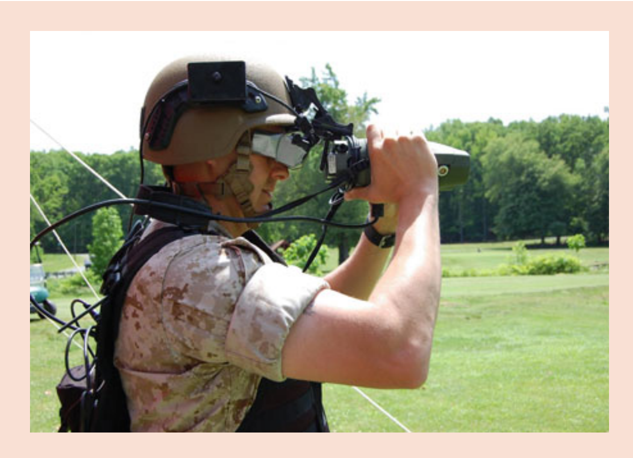


Fig. 25.4 The Augmented Immersive Team Trainer was the tested version of an AR system for training Marines in tasks that are performed by forward observers and forward air controllers. The HMD, the batteries, and the cables create issues of bulkiness, size, and weight on the head. This in turn limits how long trainees can wear the equipment and their ability to maneuver during a training scenario. Continued reduction in size of cameras, batteries, and wearable computers is vital to the practicality of mobile AR

technologies overlaid the virtual objects on an image of the real terrain using a video-based AR display.

The resulting prototype was called the Augmented Immersive Team Trainer (AITT). It was developed by ONR in a series of research programs to overlay the virtual effects of air-launched weapons on real or virtual targets in the field, using a video-based display (Fig. 25.4). During this initial technology development, testing, and experimentation, new versions of optical see-through head-mounted display (HMD) technologies became available, and ONR investigated them in conjunction with development of the AITT prototypes. However, the tracking quality was not sufficient to support see-through displays for field use. Virtual objects would appear to hover just above ground level, and there was a lot of jitter in the view even when the user was stationary. Therefore, the primary focus remained on video-based displays for the first AITT prototypes. Following multiple development cycles, tests, and feedback from a variety of Marines, ONR transitioned the first prototype using video-based displays to the USMC in 2016. At the time, commercially available optical see-through displays lacked features that were included in the requirements specified. For example, displays had a limited field of view, which could cause safety concerns or affect performance for those on the move in field training environments. In sum, the video-based, head-worn display approach made sense for the exploration of AR technologies to be used in live training. Ultimately, an optical see-through capability would be desirable for field-based military ground training, if the occlusion of the real world by virtual objects is not critical. The ability to see the real world as naturally as possible is still valued.

From a tactical and training perspective, there are additional uses for AR that could be developed and tested whether or not optical see-through technologies are ready for training. ONR continued its research into AR with their second stage of development and testing. One new feature was including AR in the Tactical Decision Kits for infantry Marines [47]. Another experiment was the Warfighter Augmented Reality (WAR) training exercise, where control measures (e.g., lateral boundaries, phase lines for synchronization, or fire coordination lines) that were formerly only drawn on maps were drawn using AR. As a result, when wearing the display, the Marine did not have to interpret from a physical map to the actual terrain to see where control measures were located. Variations of the video-based system have been demonstrated for training of specific military skills [48].

Evaluation and Testing by the Marine Corps

USMC began testing AITT to determine the suitability for two forms of training: FO learning to call for artillery and mortar fire and CAS operations. Limitations were encountered [49] with the ability for the AITT to maintain accurate terrain correlation when the trainee moved from a fixed location. The head-worn displays had limited ability to work under bright conditions (direct sunlight or even a lack of cloud cover). The virtual objects lacked accurate contrast with the environment, so that they stood out noticeably from their surroundings. Occlusion by real objects that were unknown in the database was impossible; since the model did not include vegetation and small buildings, virtual objects that should have been hidden were not.

Despite these limitations, AITT showed value in providing training. The system, in its state at the time, could not support mobile use, which the USMC ultimately desires in order enable force-on-force (FoF) training on live ranges. The combination of size, weight, and power (SWAP) requirements (from batteries, cables, and all hardware) currently precludes mobile use. However, there are training contexts for FOs and CAS operations that do not require mobility. Many tasks in these roles may be learned while the student is standing in a fixed point on a range.

Other technology issues must also be solved to make the appearance and behavior of virtual objects appropriate for training. Light levels for the real environment and the virtual objects need to be matched, and virtual objects further need to interact with the real world lighting conditions (sun/cloud, sun direction, local lights). This in particular made identification of aircraft difficult, since the simulated aircraft did not reflect sunlight as real aircraft do. While virtual objects could be occluded by the topological terrain, other real objects such as trees, vehicles, and buildings were not in the database and thus could not occlude virtual objects when the depth required it.

Discussion of Marine Training Application

From a technology perspective, until realistic and responsive optical see-through displays are available in a better form factor, it will be difficult for Marines to wear the system during training in which they are on the move. To mitigate both virtual reality (VR) simulation motion sickness and safety issues, Marines generally need to be able to see the actual terrain, vegetation, and terrain features (e.g., natural formations like cliffs and ravines, as well as man-made infrastructure).

The AR technologies in a smaller form factor could provide value by enabling users to visualize these control measures in both training and tactical operations. An added benefit was that if such displays were widely available, control measures could be drawn anywhere, not just on relatively easily identifiable terrain features such as roads or rivers. While AR hardware systems are improving, the technologies are not ready for field use, from several perspectives. Jitter of virtual objects remains too high, particularly in high-vibration environments, such as traveling cross-country in a tactical vehicle. Ruggedness and tactically usable SWAP limitations are other considerations. The potential remains, however, and the expectation is that soon a viable hardware system will become available. Furthermore, if the challenges can be overcome, then AR could become a dual-use technology: useful for both the training and SA applications described above. The benefits of dual-use hardware are compelling.

Despite this progress, the USMC has only begun to explore the potential benefits of applying AR technologies. Current technologies are still focused on small numbers of semifixed individual Marines such as FACs, FOs, and fire support teams. Certainly, AR trainers are useful for all of these, particularly considering the cost of flight hours and expenditures of live munitions. Virtual or AR training can be used for procedural and basic skills development, provide practice at the points of need, cost much less in labor and supplies, and lower the risk of accidents. At this point, moving to a head-worn AR display that does what we can already do with existing simulators is a lot closer to an engineering problem than pushing the limits of science and technology. To push these boundaries, the USMC needs AR that extends beyond the limits of training for small numbers of participants, in fixed or semifixed locations, and designed for training specialized tasks only. To leverage the potential of AR, Marines need:

- AR at the points of need so that all can easily train with AR,
- AR that enables Marines to train in actual field conditions (along with the ruggedness and SWAP that dismounted Marines need),
- better mobility in terms of being able to set the system up and use it in a wide variety of locations, without needing a significant amount of time to register sufficient known points so that real terrain and virtual objects can be properly registered.

Although Marines have training systems to support FoF training, they are laser-based and therefore limited to direct fire LOS weapons. AR has already shown the potential to increase training realism for indirect fires (artillery and mortars) and high ballistic arc weapons (grenade launchers). So AR can add more weapons to the training mix. But what all Marines in FoF training need – and that AR can potentially provide – is the ability to see and assess weapon effects, and then make decisions based on those effects. That does not mean there are not significant technology challenges remaining in order to be able to use AR for a much broader audience. The computing power needed to track all weapons' effects, from bombs to rifle bullets, is daunting. Furthermore, the ability to realistically compute and display all these weapons' effects to Marines in distributed positions, who have differing lines of sight and angles of view, is daunting. Rather than focus on the difficulty of the challenges, the following description focuses on the desired capabilities and capacities.

The first challenge is the scale of deployment of AR. Ultimately, we may provide AR capabilities for every Marine in a company-level FoF exercise, which would be approximately 360 individuals. If the system then adds attachments and supporting arms, the total number of Marines in the exercise could approach 500. Marines carry a wide variety of weapons, from rifles and machine guns through grenade and rocket launchers, artillery and mortars, weapons mounted on tanks or other vehicles (machine guns, cannon, main guns), aircraft-mounted weapons (guns, rockets, missiles, bombs), and naval guns and missiles. Many of these weapons can fire different types of munitions: from ball and tracer bullets to different forms of explosives and to illumination and obscuration (smoke) rounds. Furthermore, as robotics and autonomous technologies continue to develop, the numbers associated with human-machine teaming systems are expected to proliferate, which could raise the numbers of objects to track even higher. So it will take a lot of computing power just to simulate the variety of weapons (and their effects on varied terrain) a single Marine might use. Beyond the variety of weapons, Marines will be distributed throughout the battlefield and prepared for all domains. Marines in training need to see what they would actually be able to see in combat – and not see what they would not be able to see. As noted above, this includes blocking terrain and vegetation, temporary battlefield effects (e.g., smoke or other obscuration), and projectiles that pass through windows. The computing power needed to do this all in real time currently does not exist.

The next challenge is what is called the “fair fight” issue, which is really about synchronicity. This is not an issue intrinsic to the display, whether it is AR, VR, or desktop monitor. Given this complexity and scale of simulation, distributed computing will likely be needed to address all of the challenges. “Fair fight” means that weapons whose launch, impact, and effects are being computed in one simulation (or by one processor) must launch, impact, and inflict effects at the same time and place in all synchronous simulations (or processors). For example, if a pilot in a helicopter simulator is firing at a moving tactical vehicle, either in the field or another simulator, then the positions and timing of the vehicle and the helicopter-fired projectiles have to be the same in both simulators and, in the case of a vehicle in the field, the processor handling that vehicle’s “simulated reality.” If, as is often the case in distributed computing, asynchronous communications are used, the timing can very quickly get skewed. The helicopter could be properly aiming and firing at the vehicle’s position from 30 s ago, for example, and therefore missing it even if the pilot does everything correctly. This results in a poor training effect. While this is not inherently about AR, it needs to be solved for the AR training envisioned with AITT.

The third challenge is robustness to environments and lighting conditions. The system (or system of distributed systems) needs to be able to operate properly both day and night, in varied weather conditions. The need for fidelity of the interaction of weapon simulation with weather is a matter of debate in the training community. A related challenge is the interaction of the AR training system with other standard tools, such as night vision systems. Simulators exist for such systems, but it may be advantageous to incorporate such simulation into AR-based training.

Finally, mobility remains a significant challenge. We mentioned SWAP as an identified issue in the evaluation [49]. Marines already carry a lot of gear, including personal protective equipment (helmets and body armor), weapon systems and ammunition, food and water, radios, and also various numbers and types of batteries. USMC is trying to lighten the load. Yet current AR systems increase what Marines would have to carry. During training, some affordance can be found. For example, blank ammunition weighs less than live ammunition. So the AR system may weigh within the weight difference between live and blank ammunition. But returning to the dual-use advantages we described earlier, if the same AR system is going to be used for both training and tactical operations, it needs to be light enough that Marines can easily carry it. It also needs to be affordable both economically and in terms of power sources. If the system itself is lightweight but requires a massive number of batteries, it is not physically “affordable” to Marines who have to carry all those batteries, regardless of individual system cost.

In addition to these challenges, there are benefits beyond the combat training concepts described here. Medical AR has long been an interest to the military, and remains so (e.g., [50,51]). AR systems can actually increase the realism of focused types of training, such as by visually overlaying wounds. Corpsmen could be better trained to diagnose and treat wounds properly. Logistics support units could train having to deliver additional medical supplies and evacuate the casualties. On balance, AR technologies have a lot more potential to increase the capability to provide training to Marines that is as immersive an experience as possible. All these lead to improving Marines’ resilience and preparation, enabling Marines to safely experience things during training that they will experience in combat, without incurring injuries.

To conclude this discussion of dismounted military training, AR has the potential to increase training effectiveness and efficiency. Marines who are engaging a thinking, adaptive enemy in FoF training need to be able to realistically see in AR their virtual weapons’ effects without harming others. The initial explorations of AR for training uses have been instructive and beneficial. They have also shown that AR technologies can be useful not just for training, but may also be useful for tactical employment (as BARS was initially envisioned). We need to leverage those advances and expand them to a large number of Marines in such a way that those who are “in the fight” can see a variety of virtual objects from their individual perspectives, as they move about the battlefield. The desired AR capability needs to address the full range of weapons that Marines carry on the battlefield as well as supporting fires from artillery, mortars, ships’ guns and missiles, and air-launched weapons. Although this is currently a substantial challenge, AR technology developments can help transform military training and education environments.

This is not to say that the technology currently has no place in USMC training. Many have noted the need to have information available at low echelons of command, so that any individual Marine who must make decisions can have more information on which to base that decision [52]. The USMC in 2017 began fielding Tactical Decision Kits [53], which enable AR simulations intended to help Marines gain experience with tactical decision-making. This evolved into a project entitled Small Unit Decision Making. One implementation of the ideas uses an augmented sand table to enable a rich set of scenarios for training in tactical decision-making (Fig. 25.5). Of note in the user view shown in the figure are the colors, which come from the military standard for unit depiction [54]. The again raises the issue of color display and perception noted above, and will be discussed in Sect. 25.2.4.

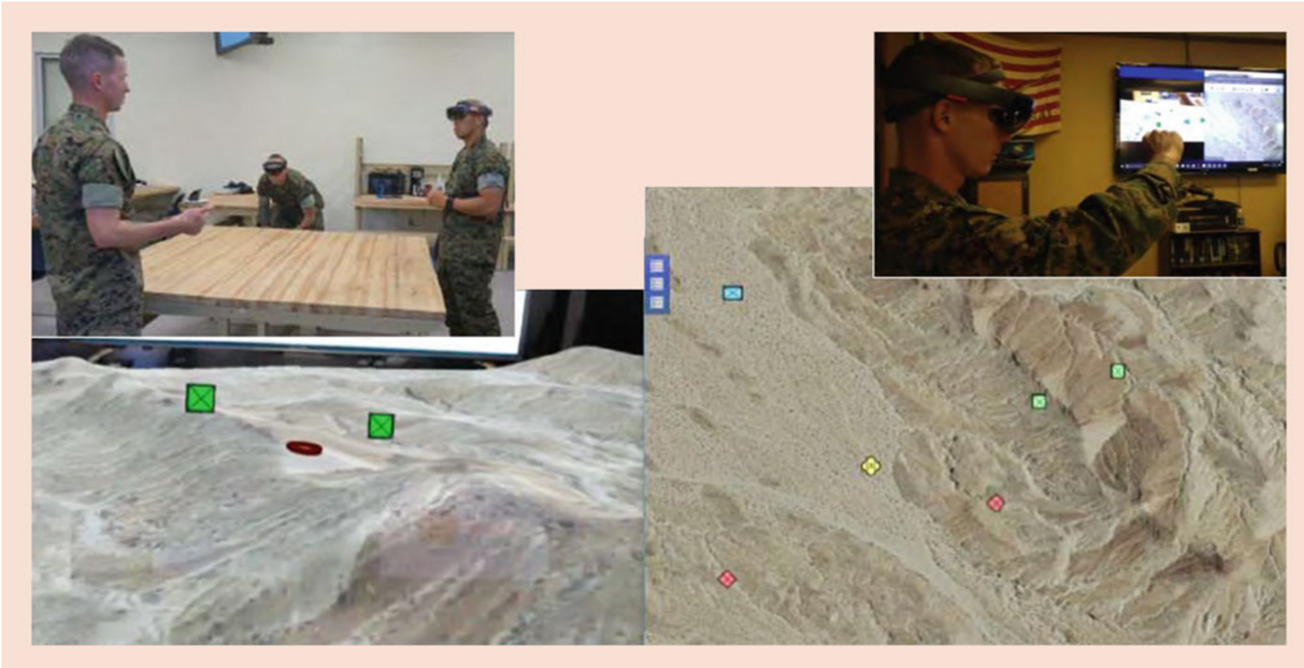


Fig. 25.5 The Small Unit Decision Making project utilized an augmented sand table to enable Marines to develop and train on a rich set of tactical scenarios. The insets show the hardware in group (top left) and individual (top right) photographs. Sample views are shown in the main images. The colors shown depict tactical information according to military specifications [54], which may or may not show well in certain AR displays (cf. Sect. 25.2.4)

25.2.2 Other Applications to Military Operations

GunnAR

In a loud environment, such as gunner positions located on various mounts of a Navy ship, it can be difficult to hear commands over the radios. A proposed solution was to provide gunners with an AR display that was capable of transmitting visual commands. This concept was developed by a Navy officer, won first prize in the Navy Innovation Jam (a competition to encourage innovative solutions to Navy problems), and was subsequently developed as a collaboration between the Naval Information Warfare Systems Command (NIWC) – then known as Space and Naval Warfare Systems Command – and DAQRI. The system included a DAQRI Smart Helmet which could transmit video to a tablet. The tablet could issue commands to the helmet such as hold fire, track, and cease fire. A compass was also included, but in the first iteration, no information was geospatially registered.

This first system was demonstrated during a Navy training exercise in the summer of 2017. In the summer of 2019, an upgraded version of the system was demonstrated during a Navy training exercise. New features included spatial registration of contacts (objects detected on radar or similar sensors) and new methods of marking points of interest, both in the tablet interface and in the display. Challenges with operational implementation of this concept included the difficulty of operating a wireless network; the displays in 2017 were attached to a wired network. Robustness and brightness of the displays were also issues.

Augmented Ship Transits for Improved Decision-Making

In 2019, NIWC developed a prototype system that aggregated automated identification system (AIS) data, radar, ship telemetry, and chart data into a single display that spatially aligned the data. In a ship transit, the navigators must steer the ship from a defined starting position to a defined ending position while avoiding dynamic obstacles such as ships and weather, avoiding static obstacles such as land or reefs, following the rules of ship navigation, and staying mission ready. The purpose of the AR system was to make it easier for sailors to get access to contextually relevant information (about potential obstacles and nearby traffic) while still maintaining SA on the ocean. The system was comprised of a laptop server, a wireless network, and a Magic Leap One AR display. The server was wired to the existing communication network for many shipboard instruments, giving the AR server access to the data listed above. A VR prototype of the system was developed, and human factors testing was conducted to optimize the visual representations to be incorporated into the AR prototype. This system was tested at sea during a Navy training exercise in the summer of 2019.

Development of this system continued into 2020, with a greater emphasis on user-centered design and human testing in a realistic bridge environment, in order to measure ship driving performance and the SA impacts of using such a system. Initial feedback suggests that an AR decision aid could reduce the likelihood of mishaps during ship transits. Some challenges of the ultimate realization of this system include difficulty of connecting to Navy data systems built by multiple contractors, registration of virtual objects with real-world objects on a moving platform, display brightness, lack of wireless infrastructure, and a less than robust hardware platform.

Virtual Scientist

NIWC Pacific and DAQRI developed a system that enabled virtual collaboration between a FO and a domain expert. The system sent 3D indexed video feed back to a scientist in a remote location via a closed proprietary network. The remote scientist could annotate the environment with virtual objects which were visible to the FO. Furthermore, the FO could scan the environment and transmit a 3D model back to the scientist via the network. These 3D models could be compared to models acquired at different times, and changes could be rapidly detected. This system was also successfully tested using a satellite communication network, showing that such a remote expert connection could be completed from any two points on the planet.

Other capabilities, such as the integration of other types of sensor information overlaid spatially on the FO's view and advanced networking schemes, were planned. Unfortunately, DAQRI went out of business in September 2019. NIWC attempted to find another company capable of providing a similar capability (remote expert assistance with true 3D registration, and 3D scanning) in a similar time frame (6 months) that was developed by DAQRI, and was unable to do so. This highlights one of the significant challenges in the emerging AR domain: companies may decide to no longer support AR, or the market may decide to no longer support the company.

Battlespace Visualization

In 2019, NIWC developed a 3D battlespace visualization application which displays terrain and units such as submarines, surface vessels, and planes. This software is connected to a common battlespace scenario generation tool called Next-Generation Threat System. The application can be networked across multiple displays so that multiple users can see the same battlespace in 3D. The advantages of using 3D battlespace visualization are hypothesized to include more complex visualization capabilities and thus improved SA. One proposed example is visualization of sensor envelopes and weather effects. This work has obvious implications for the dismounted training application described above.

25.2.3 Maintenance and Repair

As noted above, maintenance and repair applications are of interest in military scenarios, related to the larger context of industrial applications. This section highlights a collection of efforts across the Navy.

Service Maintenance Augmented Reality Tools

In 2012, the Naval Undersea Warfare Center Division, Keyport (NUWCKP), inspired by the Henderson and Feiner research [26], began looking at AR technology as a way to address various requirements in the training and performance support community. Service Maintenance Augmented Reality Tools (SMART) was a 2-year investigation of using AR as a performance support tool for shipboard maintainers [55]. The project categorized AR use cases, developed a set of candidate technology requirements for military use of AR in maintenance and maintenance training scenarios, and demonstrated the ability to share and reuse virtual content between VR familiarization training and AR performance support applications (Fig. 25.6). This project grew directly from the Navy Augmented Reality Consortium (see Sect. 25.3.1) and eventually included teams from seven Navy warfare centers.

In-Service Engineering Agent of the Future

The In-Service Engineering Agent of the Future (ISEAotF) considered AR and VR technologies a supporting thread for human-system interfaces to the end-to-end logistics information flow, with a broader scope aimed at modernizing and transforming in-service engineering (ISE) practices. The Naval Surface Warfare Center, Port Hueneme Division [56], defines ISE as "a very broad function that includes the installation of systems on ships, certification that the systems perform as designed, and the training and qualification of the Sailors that operate and maintain the systems. It also entails ensuring that the logistics support required (such as training plans, maintenance procedures, spare parts, etc.) is in place when the system is installed." The core of the ISEAotF initiative was transforming how this logistics support traveled along with the product through its life cycle. The team quickly recognized that AR and VR were themselves transformative ways for personnel to interact with that information stream.

Ocean AR

The Ocean AR system included a laptop server, a mobile phone, and two Google Glass displays connected via a wireless mesh network. This system was capable of authoring maintenance procedures including tasks and subtasks built upon an XML framework which could include texts, pictures, and videos. Maintenance procedure authoring could be performed on the Google Glass using the camera and speech

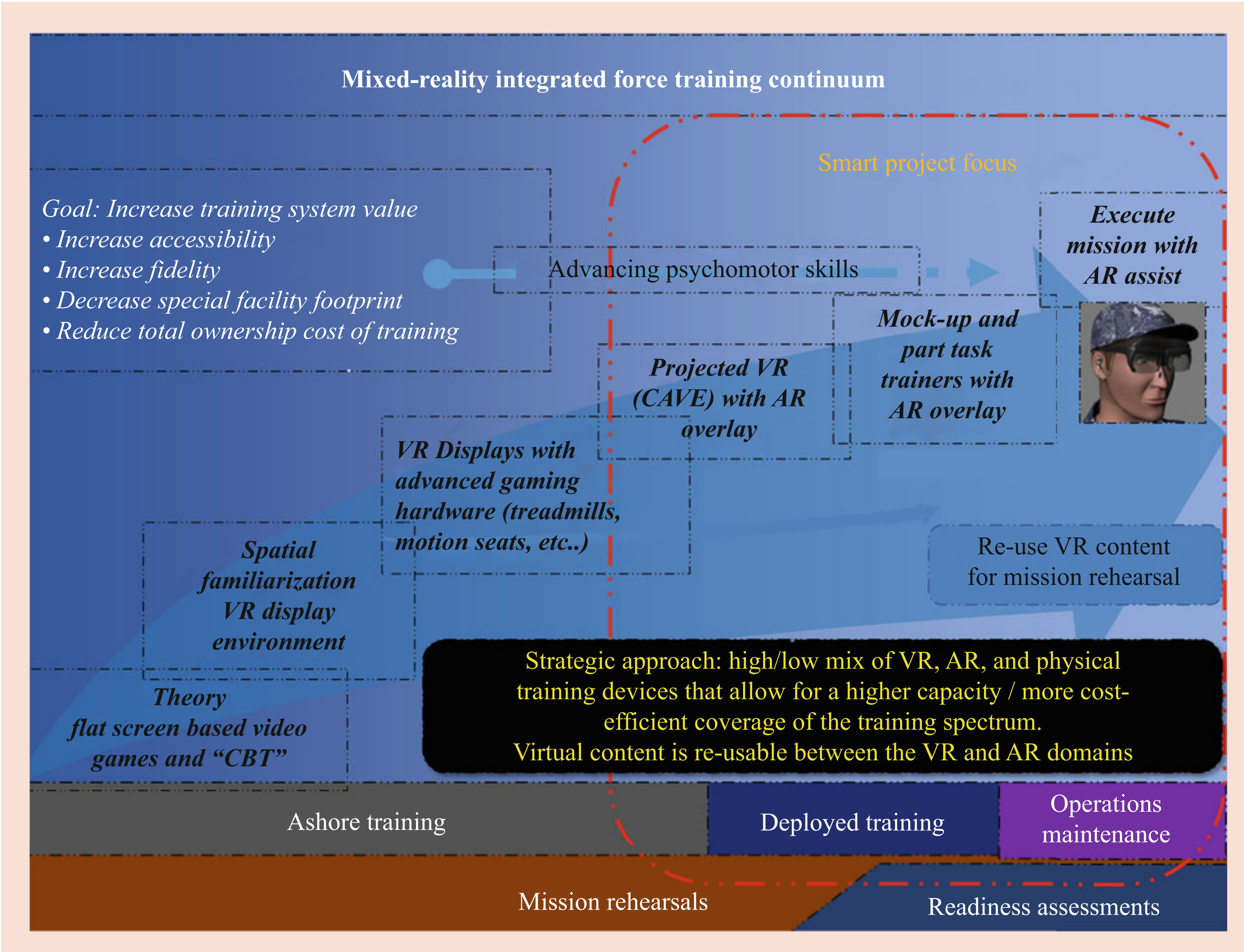


Fig. 25.6 Service Maintenance Augmented Reality Tools (SMART) sought to fit AR and VR technology into a training continuum for Navy ship maintenance

to text, by using a smart phone application designed for authoring, or by using a computer. Each of these procedures could then be uploaded to the laptop server and accessed using either of the Google Glass displays.

The existing procedure relies on planned maintenance system (PMS) Maintenance Requirement Cards (MRCs). MRCs provide detailed procedures for the accomplishment of system, subsystem, and equipment maintenance. The MRC also identifies the safety precautions, personnel, tools, parts, materials, and test equipment required to accomplish the maintenance task. Maintenance procedures based on MRCs were built and were accessible via QR code scanning capability built into the Google Glass application. As steps were completed in the procedure, the next step could be accessed via a swipe gesture or via a voice command. When a task was completed that required a supervisor's visual inspection, a photo could be taken and automatically transmitted to the supervisor for review. Once approved, the maintainer would receive notification and be able to access the next step in the procedure. A video feed from the maintainer could also be transmitted to a supervisor or an expert, and audio communications could be transmitted between the two. This system was demonstrated on a Navy carrier in September 2015.

Some challenges or limitations of implementing this technology in a Navy context included lack of wireless infrastructure, difficulty of getting approvals for new classes of hardware like Google Glass, a lack of robust physical characteristics of the display, poor battery life, a display that was a little too small to be used for more than a tweet's worth of information, no spatial registration capability, a general fear of using something that was untested and potentially unreliable, and ultimately a lack of evidence and precedent for using AR to save time and money in Navy maintenance.

Local Maintenance

Since 2016, the CREATIVE (Collaborative, Research, Engineering, Analysis, and Training in Immersive Virtual Environments) Lab at the Naval Air Warfare Center-Aircraft Division (NAWC-AD) has been conducting research with AR technologies. Their primary focus has been on the development of a real-time AR operation/maintenance support development architecture that will allow for the quick and efficient authoring of AR for maintainer support aids. The architecture aims to process XML data, identify parts information, and automatically link CAD models/schematics to traditional technical procedures. Once amalgamated, this data will be imported into a pre-constructed, generic template UI for use by various AR hardware platforms.

Upon completion, this system is hypothesized to save time and money by providing maintainers with real-time, spatially meaningful guidance while performing particularly complex maintenance tasks. Challenges associated with this effort have included an ability to reverse engineer current software applications that digest technical manual source data and construct formatted technical publications. The association of CAD models and detection software with technical documentation references is an ongoing problem. Limitations of current hardware technology and software detection systems also affect both how well components are registered in space and how fixed virtual objects remain over time.

Remote Maintenance Aid

The CREATIVE Lab spent much of 2019 conducting in-house development of an AR enhancement to remote servicing that enables two-way "live virtual" instruction and collaboration (Fig. 25.7). The AR Remote Maintenance Support Service (ARRMSS) provides the on-demand ability for a fleet maintainer to communicate directly with a remote subject-matter expert (SME) as though they were co-located in a maintenance space. This is achieved with the use of a Microsoft HoloLens 2 (maintainer) and Panasonic Toughbook Tablet (SME). Video and audio transmissions combined with spatial mapping data allow users on both ends of a network to troubleshoot problems and step through particularly challenging repair tasks/engineering investigations collaboratively, regardless of geographic location. The end product will provide "instant," virtual, and on-site support communication that allows remote field technicians and system experts to connect through live video, audio, text, image, and virtual objects (spatially geo-referenced). In August 2020 the CREATIVE Lab Team successfully conducted a Rapid Prototyping, Experimentation, and Demonstration (RPED) Advanced Naval Technology Exercise (ANTx) to demonstrate real-time use of ARRMSS between NAWC-

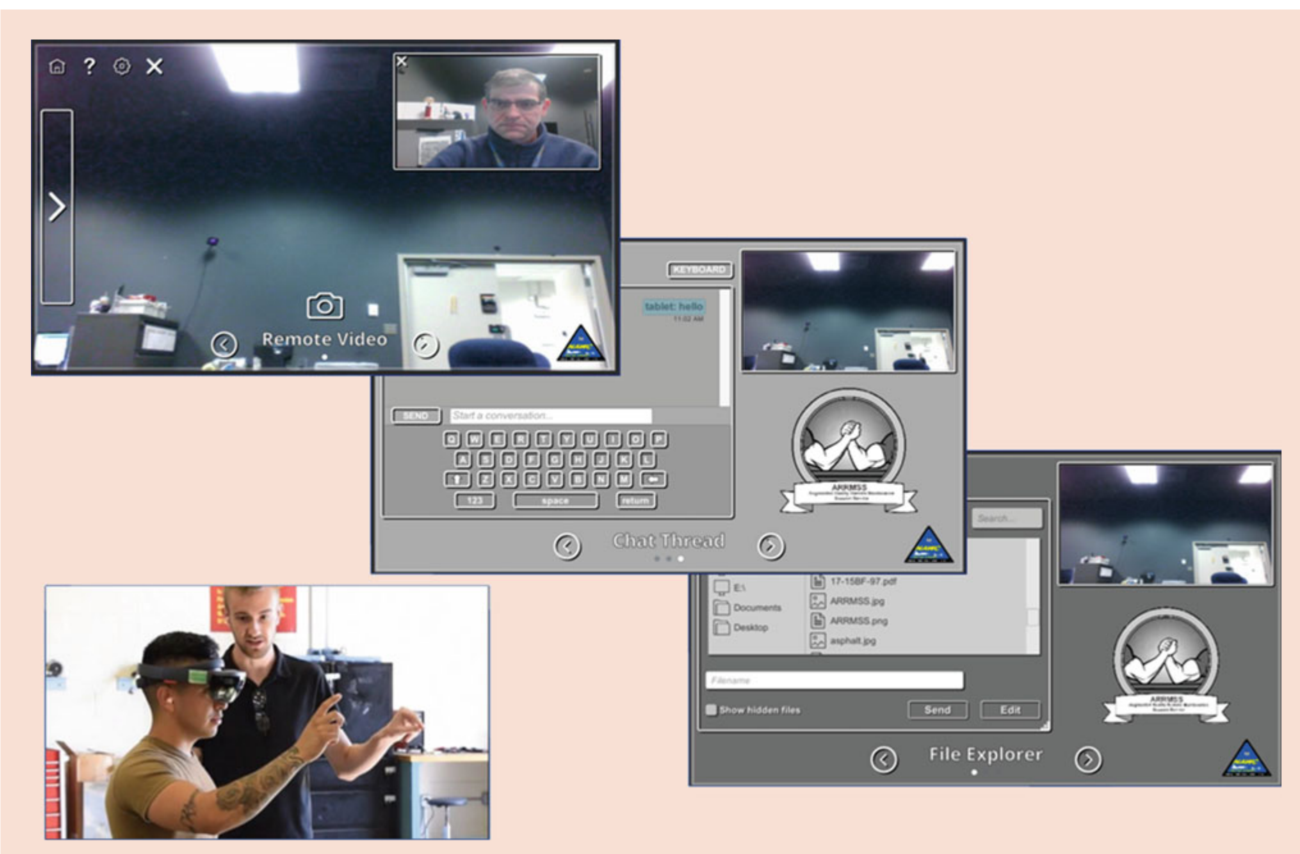


Fig. 25.7 The AR Remote Maintenance Support Service (ARRMSS) is a proposed implementation of remote servicing to provide instruction and collaboration. ARRMSS combines video and audio transmissions with spatial mapping data to promote collaborative conduct of repair tasks and engineering investigations

AD Lakehurst, NJ, and NAWC-AD Patuxent River, MD. To demonstrate usefulness the team conducted troubleshooting and maintenance on a piece of Common Ground Support Equipment (CSE). The event was conducted over the Navy Research, Development, Test and Evaluation (RDT&E) network. When fully operational, the remote support capability will save substantial funding on experts' travel to distant sites for troubleshooting and repair. The strongest limiting factor in implementing this technology immediately involves finding an appropriate network to host the application. The network must be readily accessible from many locations and support the high data rate needed to relay spatial information comfortably.

25.2.4 Other Applications and Technologies

There are many other applications and technologies of interest to the Navy. We discuss a few that do not fit in, or cut across, the categories described above.

Vestibular Therapy Using AR

One interesting application of AR is its ability to provide alternative therapy for traumatic brain injury patients. NIWC collaborated with the Navy Health Research Center (NHRC) in 2019 to develop an AR therapy system built on the Magic Leap One. Patients who have suffered a traumatic brain injury need to learn to balance again while walking, moving their arms, and focusing on objects coming from multiple directions. Simple tasks like grocery shopping can be overwhelming to these patients, so they need to practice these skills in a safe environment, like an AR one.

The system provides two AR games that patients can practice with the help of a therapist. The first game launches balls and provides a 3D audio signal indicating the direction to the patient, who then uses hand gestures to swat the balls away. The therapist has a tablet interface connected to the display, and is able to adjust the number of balls, the direction from which the ball approaches the patient, the velocity of the balls, and more. The application tracks successful hits and saves this data between sessions so that progress can be measured over time.

The second game uses the controller that comes with the Magic Leap One display. It is a target shooting game, much like a carnival game. Targets appear around the patient, and they use the trigger on the controller to shoot the target, which then explodes in a colorful fireworks display. In both games, the color of the ball or the color of the target can be changed, requiring more focus of the patient when they are asked to only shoot the green targets, and avoid all the others. To further challenge patients, a "psychedelic" mode was also developed, which overlays continuously morphing white-striped lines conforming to the flat surfaces in the environment. This capability was transferred to NHRC and Walter Reed National Military Medical Center in 2019.

Rapid Prototypes

In 2016, NIWC developed a few rapid prototypes to showcase how one might use the Microsoft HoloLens display and computer system. Of particular interest were the HoloLens' environmental sensors, allowing for very stable display of 3D objects registered to the real world. It also incorporated gesture interactions that offered a new way of interacting with 3D content. One rapid prototype that NIWC Pacific developed showcased a simple maintenance procedure of a P100 pump. The pump appeared life-size sitting on the ground in front of the user. A person could walk around and inspect any side of the pump model, and use gestures to cycle through different steps in the training procedure. Another prototype was a multi-view visualization of a drone flying through a canyon. The user could observe the drone from a third person perspective and switch to a mode where they were seeing what the drone would be seeing. Finally, a visualization was developed showing satellite constellations surrounding the earth. Each satellite could be selected using gaze and hand gestures, and more information could be obtained about that particular satellite.

Human Factors

NUWCKP has also done some research into optimizing the human factors and user interface aspects of AR (and VR). For example, menu and interactive components of the immersive 3D environment do not carry over well from the traditional flat-screen computer user interface domain. The group has looked at media selection criteria, for determining when factors such as safety, fidelity, comprehension, and cost will drive the choice of developing elements of training systems in either virtual or physical media delivery methods. There is an ongoing review of published literature and industry trends (especially in the gaming industry) to address shortcomings in the technology that can lead to adverse physiological effects like cybersickness or VR sickness.

Ongoing projects at NUWCKP concern advancing the technology for virtual force-feedback (tactile) elements of a virtual experience, as well as using AR and VR technologies for data visualization, especially for the day-to-day operations of the cybersecurity community. The research leverages the capability of virtual content to convey important conceptual information in a much more intuitively understandable form, allowing operators to respond to cyberspace anomalies more quickly and with a better understanding of the possible outcomes of their responses. The NUWCKP research team is also investigating other human-assistive technologies such as robotics and exoskeletons, as well as artificial intelligence for decision support systems. There is a recognized intersection between these technologies and the use of AR or VR presentation methodologies to provide cognitive elements of their user interface.

Two issues raised by the user interface are natural manipulation of virtual objects and tactile feedback, which can lead to the ability to convince the users they are touching or acting on something of substance that is represented by a virtual object. There is ongoing work in both of these areas using tactile gloves. Latency in the data from the gloves leads to a sort of "uncanny valley" experience; latency or loss of tracking of the hands can completely disrupt the AR experience. In vision-based tracking, this may result from poor or harsh lighting conditions. The tactile issues are a concern for training and virtual design scenarios.

Color Perception

Twice in this chapter, we have noted the issue of proper display and perception of color for military applications. This is a particular concern when the application must adhere to the military standard [54] that mandates particular colors depicting units within a battlespace (cf. Fig. 25.5). When the problem first arose in the BARS training application, NRL launched a sub-project aimed at measuring the distortion of color perception (along with other basic perceptual properties, such as the intertwined measures of visual acuity and contrast sensitivity and stereoacuity) [41]. In this section, we summarize some results and present a new analysis of some of the existing data [57] using a more perceptually uniform color space [58], in keeping with more recent work on this issue in AR displays [59].

The first way to investigate the challenge of color in AR is whether the display is presenting something that should be perceptually different from the intended representation. We measured this by using a color and luminance meter (Konica Minolta CS-200 [60]). This gave XYZ tristimulus specifications of the color and luminance at a point. The study design called for volunteers to perceptually match color patches in three conditions:

- a baseline condition of unmediated (but physically separated) visual inspection of the reference monitor and target to be modified to match,
- unmediated reference monitor, but target on the AR display,
- unmediated reference monitor, but real-world target mediated by the optics of the AR display (see-through condition);

therefore, we took measurements with the meter in these three conditions as well. There are two sources of distortion of color in each AR case; the above conditions can introduce distortion, but on top of these individual conditions, the reference monitor's display gamut may also introduce a distortion. Thus, the data is presented as four series (Fig. 25.8), the three noted above, and the specification transformed into the CAM16 color space [58]. Each series has 12 data points, corresponding to 12 color targets that were tested with users, as described below.

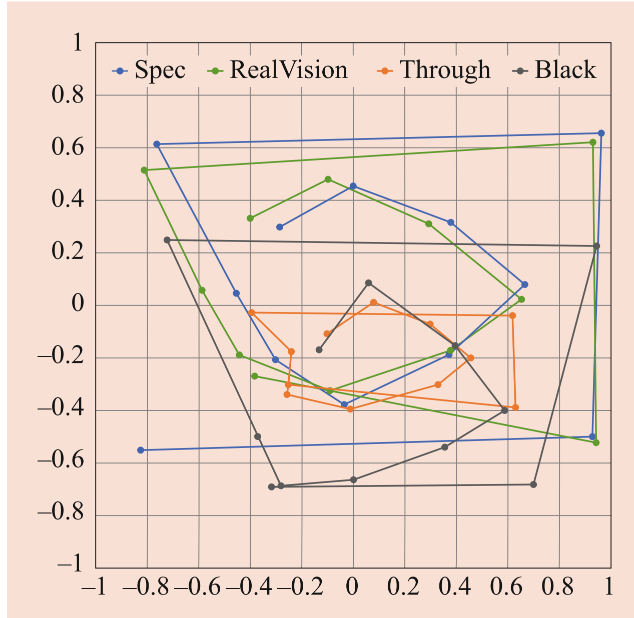


Fig. 25.8 Four data series corresponding to the specification of colors and the three display conditions, as measured by the Konica Minolta CS-200 Color and Luminance Meter. The difference between the baseline `RealVision` condition and the `Spec` (specification) shows the distortion created by the monitor gamut. The additional differences between the see-through (`Through`) condition and the `RealVision` condition, as well as between the graphics-on-black background (`Black`) condition and the `RealVision` condition, indicate distortion caused by the AR display conditions, as measured by the color meter. Note the distortion in the lower left (blue-green) corner of the monitor gamut (`RealVision`), which otherwise closely follows the specification (`Spec`). The graphics-on-black background (`Black`) data shifts notably toward the lower (blue) portion of the color space. On the other hand, the `Through` condition compresses toward the center (achromatic) point. This graph predicts that users will have a challenge when it comes to accurate color perception in some AR display conditions

The data shows that the reference monitor had a minor distortion from the specification, mostly due to one point with the specification nearest the blue-green (lower left in Fig. 25.8) corner of the chroma coordinate plane. These differences reflect the gamut of the reference monitor. What is most important to note is the additional distortion of the objective color measurements in both the optical see-through condition (series denoted `Through` in Fig. 25.8) and the view of graphics on a black background condition (denoted `Black`). The see-through data shows a significant compression of the data toward the center of the chroma plane (which is the achromatic point), shifted slightly toward the blue portion of the space. This is presumed to come from the optics of the display, although it remains to be determined whether this is an inherent challenge presented by the optics or merely reflects the challenge of aiming a color meter through the AR display optics at a reference patch beyond the display in the real world. The graphics-on-black data is shifted toward the bottom of the plot, which is the blue portion of the chroma plane. It also shows a slight compression toward the centroid of the data series. This data would lead us to predict that those wearing a see-through AR display will have a challenge when it comes to accurate perception of color.

We measured the distance from the projection of the color specification for all 12 color patches. We computed the ΔE_{Jab} color distance metric from the specification for each color sample and averaged over each display condition. The results show that the `Through` condition was not statistically different from the `RealVision` condition; this may be due in part to the high variance associated with this portion of the data. This would indicate that future work ideally should gather a larger data set for a more definitive result. The `Black` condition was statistically different than the `RealVision` condition; the meter's measures of color are significantly affected by the black background (Table 25.1). This leads us to hypothesize that users will have a difficult time matching colors on the st1080 display (viewed over a black background) against a reference monitor.

Eight volunteers consented and contributed data to the user study, with each session lasting approximately 30–45 min. All users had normal color perception. Each user matched 12 color targets against the reference monitor, in the 3 conditions listed above and a fourth condition in which the background for the see-through display graphics was a white field (graphics-on-white or `White`). We analyzed the results with the dependent metric of color difference ΔE_{Jab} in CAM16-UCS, an adjustment of the color space to a more perceptually uniform distance metric [58]. The results are partly predicted by the results with the color meter. There was a significant difference between the display conditions, $F(3,21) = 68.185$, $p < 0.001$, $\eta^2 = 0.864$. The `Black` condition was the most difficult, but post hoc testing (Table 25.2) shows that the `Black` condition led to significantly greater error only compared to the `RealVision` condition, with the difference between `Black` and `Through` being marginally significant. With only eight users, all these results should be considered preliminary, but it does raise the concern that over a dark background, users may not perceive the colors that the AR designer intended to display. Figure 25.9 illustrates the variation in the color map for the four display conditions. Each bubble map shows the bubbles plotted at the average location across the eight users, with the bubble radius indicating the mean error associated with that color patch. The bubbles overall are larger for the `Black` condition (upper right of Fig. 25.9), mostly because the minimum error (size of any individual bubble) is larger than the other conditions. All display conditions had color patches which users had a difficult time matching accurately.

This issue is not specific to a single display; we conducted a similar user study with the nVisorST display [61], an early, high-resolution, stereo, optical see-through display. In this study [41], 19 volunteers consented to match 22 target colors in 5 display conditions. The baseline `RealVision` condition was the same (albeit with a different monitor). Four graphics-on-color-background conditions were studied; the background colors were a saturated red, green, blue, and yellow. The results showed that there was a significant difference between the display conditions, $F(4,72) = 93.044$,

Table 25.1 Statistical evaluation of the color differences shows that the effect of the black background was significant in changing the color values measured by the color meter. The Student t-statistic is calculated for the two AR conditions versus the `RealVision` condition, and the mean distance (average ΔE_{Jab}) in color space is computed compared to the specification's projection into CAM16 color space. DOF stands for degrees of freedom, calculated with the Welch-Satterthwaite equation for quantities with unequal variances. The `Through` condition shows a different value, but with its high variance, we cannot declare it to be statistically different, whereas the `Black` condition does meet the standard with its low p-value

Display condition	Mean distance	St. Dev.	t-statistic	DOF	p-value
RealVision	12.11	13.58			
Through	39.32	23.62	1.364	17	0.190
Black	42.89	9.00	7.192	19	<0.001

Table 25.2 Statistical evaluation of the color differences between perceptual prediction and user-matched targets shows that the effect of the black background was significant in changing the color values perceived to be matched by the study participants. The Student t-statistic is calculated for the first three conditions versus the `Black` condition, and the mean distance (average ΔE_{Jab}) in color space is computed compared to the predicted projection into CAM16-UCS color space [58]. DOF stands for degrees of freedom, calculated with the Welch-Satterthwaite equation for quantities with unequal variances. The difference between the `Black` condition and the `RealVision` condition is significant, and the difference between the `Black` and `Through` conditions is marginally significant; more data would perhaps clarify this result

Display condition	Mean distance	St. Dev.	t-statistic	DOF	p-value
RealVision	7.19	5.01	3.325	9	0.009
Through	12.46	7.57	1.909	12	0.080
White	18.95	9.85	0.495	13	0.629
Black	21.55	11.14			

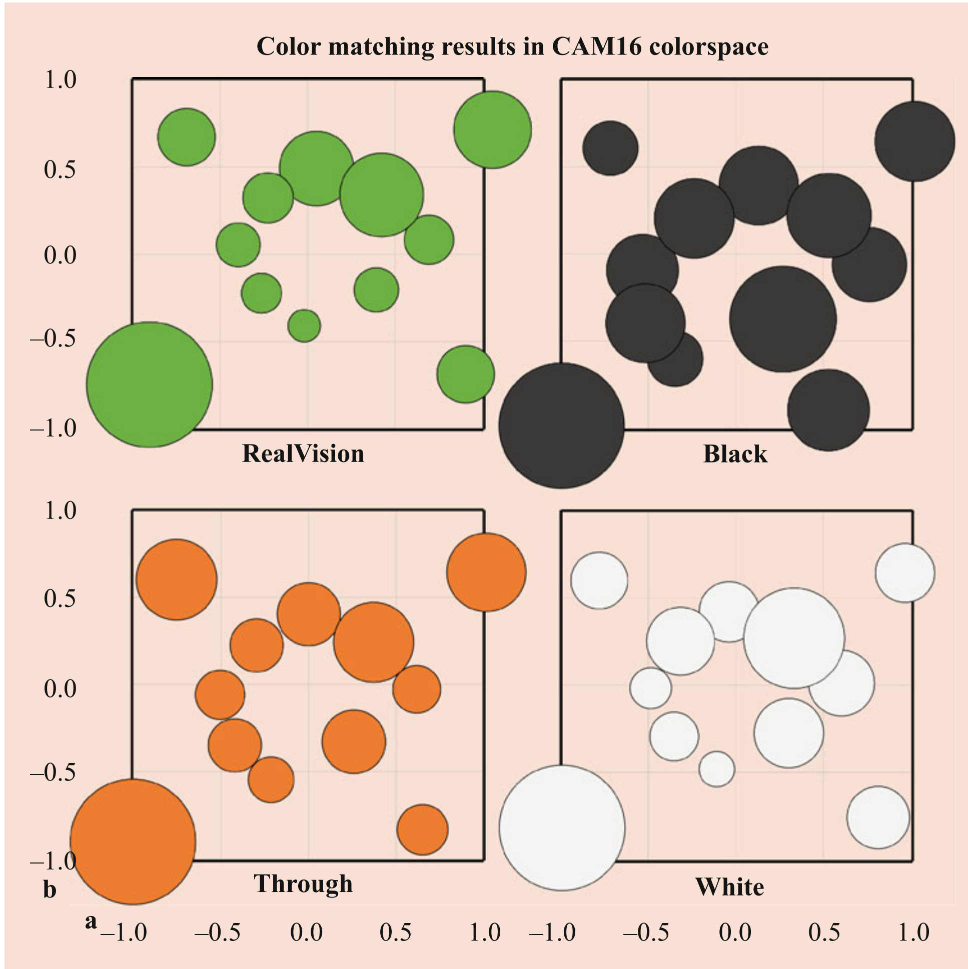


Fig. 25.9 Four graphs are shown, corresponding to the four display conditions in which users matched colors. The difference between the graphics-on-black-background `Black` condition and the baseline `RealVision` condition is statistically significant. The difference between the `Black` condition and the see-through (`Through`) condition is marginally significant. Note the large bubble in the lower left (blue-green) corner of the each graph, echoing the error in the monitor gamut (cf. Fig. 25.8). These graphs further indicate the challenge that users will have in properly perceiving colors

$p < 0.001$, $\eta^2 = 0.742$. Post hoc testing (Table 25.3) showed that two (green and yellow) color backgrounds were statistically more difficult than the baseline `RealVision` condition, with the blue background marginally significant for being more difficult. No color backgrounds were significantly different from each other. We conducted this analysis with the white point correction for see-through recommended by Hassani and Murdoch [59]; however, we saw no overall improvement with this white point correction. Examining the data for each color sample under each display condition, we note several curious observations. First, the direction of the shift from the specified color (transformed into the CAM16-UCS color space) to the location matched in the baseline condition is not consistent (Fig. 25.10). Most points were pulled toward the achromatic center, but some were not. Whether this is due to error (including noise) in the measurement, the color space conversions, or the perceptual matching is unclear. We note that some of the points with the atypical shift direction also have large errors, so perhaps outlier analysis would be helpful.

Figure 25.11 shows four bubble graphs that correspond to the one in Fig. 25.10. These bubbles indicate the color match error in CAM16-UCS color space for each target color over the background specified in the title of each graph. One can see that, while not immune to some gross errors, the graph for the red background (upper left) has mostly smaller bubbles; it was not significantly higher error than the `RealVision` condition. The blue background caused a marginally significant increase in error, and more of the bubbles are larger. This trend grows for the yellow and green backgrounds, both of which had a significant increase in error compared to the `RealVision` condition. We also note that many of the error bubbles overlap (sometimes completely) in all four graphs. This leads us to consider whether at least some users may not have been able to tell some of the target colors apart. While we attempted to distribute them evenly in the color space, we did not reposition them in response to the distortion caused by the monitor gamut, let alone the background conditions. So some overlap is to be expected, but this still is a curious result. Some color targets saw increased error on the matching task when we used the white point correction to predict the location in color space that would be matched. This is also a curious result, with similar possible sources to the atypical direction of error noted above.

To summarize this line of research into color perception in AR, we return to the issues that instigated this work. We

Table 25.3 Statistical evaluation of the color differences between perceptual predictions and user-matched targets shows that three of the color backgrounds tested made a significant or marginally significant difference in the error on the matching task. This data was collected using an nVisorST display (different than the data in Table 25.2), but a similar general trend emerges. Certain background colors make it hard to perceive the intended color of the graphics on the AR display

Display condition	Mean distance	St. Dev.	t-statistic	DOF	p-value
RealVision	17.53	20.45			
Red	25.83	21.31	1.193	35	0.229
Blue	32.64	26.06	1.989	34	0.054
Yellow	34.85	23.08	2.384	35	0.019
Green	37.16	23.30	2.688	35	0.009

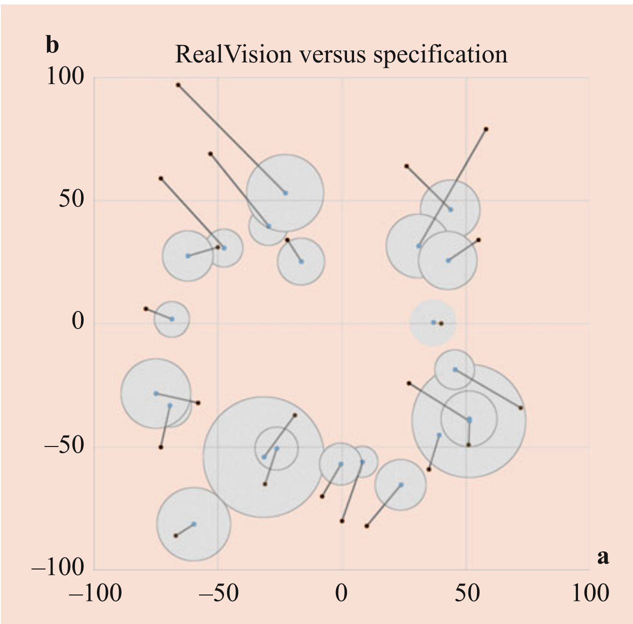


Fig. 25.10 This graph over CAM16-UCS color space shows the matched locations in the baseline `RealVision` condition (blue dots) along with the specification (red dots) of the target color (connected by a line to the center of the bubble for each target location). The bubble size indicates the mean error across users for that target. We note that there is some inconsistency in the direction of the error, though many points were pulled toward the achromatic center point of the color plane. We also note that two of the larger bubbles (indicating larger matching error) in the bottom (blue) half of the color space are on the points with atypical direction of shift from the specification

noted that users had remarked that they were unable to see in one of our early displays, and the recent augmented sand table application uses standard military colors [54] to represent the identity of entities in a battlespace. There are light, dark, and medium shades of blue (friendly), red (hostile), green (neutral), yellow (unknown), and purple (civilian). We converted the RGB color space definitions of the tones into CAM16-UCS for distance computations. Looking at the pairwise differences (Table 25.4), we found that certain color pairs were separated by values of ΔE_{Jab} that are low enough to be near the *average* errors we found for color matching in the AR display conditions. This means that at least some of our users were making errors with some colors that were greater than the tolerances implied by Table 25.4. This raises an issue of the accuracy of color perception that we can expect in such applications implemented with optical see-through AR displays.

25

25.3 Discussion

As a preamble to a discussion of potential synergy among the various applications described in this chapter, we note two mechanisms by which Naval and USMC collaboration already takes place on AR concepts, technologies, and applications.

25.3.1 Augmented Reality in the Navy: A Roadmap

In September 2015, as part of an effort to explore applicability of AR to Navy problems, a group of Navy technologists formed the Navy Augmented Reality Consortium. The group wrote a roadmap for applying AR technology in the Navy [62]. Contributors included participants from multiple Navy warfare centers, DARPA, and commercial partners of some of those groups at the time. It includes an overview of AR, a predicted timeline for AR development, AR use cases aligned with the timeline, opportunities and challenges, and a stakeholder map. This organization is analogous to industry consortia like the Augmented Reality for Enterprise Alliance and (primarily) academic organizations like the community that participates in the IEEE International Symposium on Mixed and Augmented Reality.

25.3.2 Human Factors Engineering Technical Advisory Group

This chapter explored the US Navy's and Marine Corps' interests in AR as a tool for increasing warfighter and civilian support staff capabilities. While there are clear distinctions between immersive VR and AR, there are common technolo-

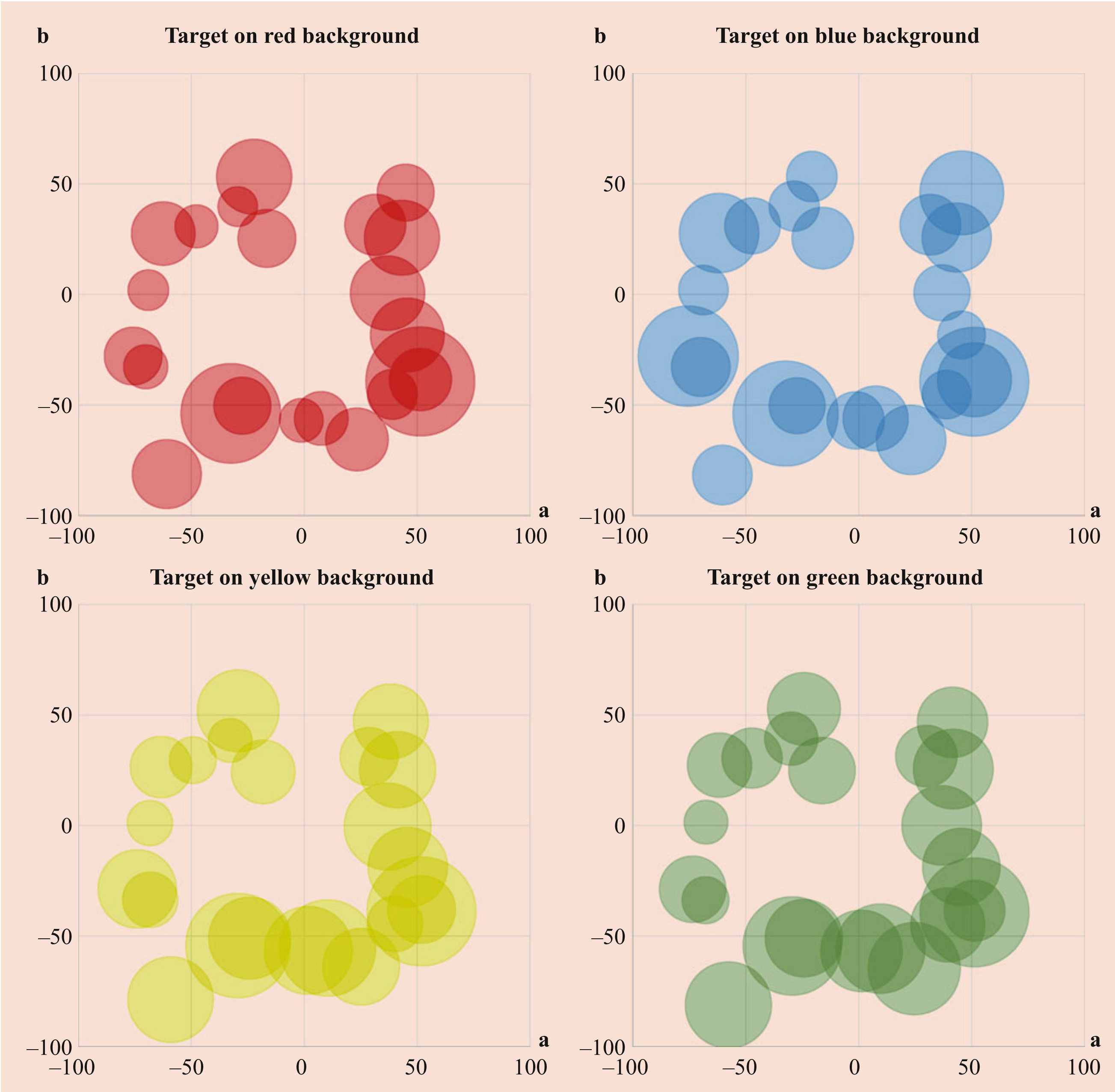


Fig. 25.11 Four graphs are shown, corresponding to the four display backgrounds over which users matched colors in the nVisorST display. These graphs further indicate the challenge that users will have in properly perceiving colors. Each graph shows the errors for one background color, plotted in CAM16-UCS color space. The bubble size indicates the mean error. Note that many bubbles overlap, leading to concern that some users may actually not have been able to tell target colors apart, had they been asked to directly compare them (which was not a task in the study)

gies and content production methodologies between them. As such, the technologies are viewed as part of a human-system interface continuum. Currently the Office of the Undersecretary of Defense Human Factors Engineering Technical Advisory Group (HFE-TAG) has an ongoing effort to rewrite the US DoD standard on human-systems integration to address these relatively untapped (within DoD) human interface technologies.

25.3.3 Lessons Learned

The introduction and the trace of the research and development of the training application highlighted the first theme. AR is a natural tool to provide SA in whatever context may be envisioned by the developers of an application. However, across the many domains and applications, we have noted multiple issues with the various technologies:

Table 25.4 The matrix above the line (which is symmetric and thus shown only above the diagonal) shows the range of ΔE_{Jab} values for the light, medium, and dark shades of the standard military colors (whose specifications are below the double line). These colors are used to represent entities in battlespace maps or other symbol systems. The entries in the matrix indicate how perceptually close the colors might be, and thus what distances of the distortion of color perception might be problematic for the accurate interpretation of these colors. Comparing these with the average errors found in some of the AR display conditions shown in Tables 25.2 and 25.3, we see that while most distances between colors are comfortably greater than the errors we noted, some are rather close. Noting that the values reported in Tables 25.2 and 25.3 are averages across many users, we must raise the issue of whether all users will accurately recognize colors in an AR application that relies on these standard colors

Entity type	Hostile	Friend	Neutral	Unknown	Civilian
Hostile		(75,115)	(83,127)	(71,96)	(41,71)
Friend			(40,74)	(58,80)	(51,77)
Neutral				(27,35)	(78,130)
Unknown					(80,115)
Civilian					

RGB specification				
Entity type	Nominal color	Dark	Medium	Light
Hostile	Red	(200,0,0)	(255,48,49)	(255,128,128)
Friend	Blue	(0,107,140)	(0,168,220)	(128,224,255)
Neutral	Green	(0,160,0)	(0,226,0)	(170,255,170)
Unknown	Yellow	(255,220,0)	(255,255,0)	(255,255,128)
Civilian	Purple	(80,0,80)	(128,0,128)	(255,161,255)

- Accurate modeling of the environment appears to be a requirement for many of the application contexts. Certainly, if the application needs to align virtual objects with the terrain and have both real and virtual objects properly occluded by each other, then an accurate model of the environment seems to be a prerequisite. Perhaps certain approximations and real-time data may alleviate the requirements, but this is certainly a direction for future work. Models are also critical for many of the maintenance and repair applications (in the military, as we note, and in more general maintenance applications of AR [22] (see also Chaps. 3, 18, 20, 23, and 24 in this volume). However, this occurs at very different scales in different applications. An open question in each application is the fidelity of the models needed to support the application. In procedural training contexts, greater compromise on the accuracy may be acceptable, whereas for training ultimately aimed at improving performance in the context of military operations or large-scale simulation (exercises), the fidelity may have a significant impact on the quality of the training.
- Many application contexts require users to assimilate and integrate data from multiple sources. This is a challenge for several applications, but especially those in which providing SA is the primary goal for the AR system to support.
- Mobility is a desirable goal or critical feature shared at one of these levels across multiple applications. Current AR hardware has certainly improved in permitting mobility for an AR user, but as noted above, the capability in military applications is insufficient to support some applications. While communications technology has been demonstrated to support a wide range for mobility, guaranteeing such communications is a challenge. Networking requirements and the local computation load are a larger issue facing mobile applications of AR (see ▸ Chap. 11, "Networking and Cyber Foraging for Mobile Augmented Reality").
- Related to mobility is the robustness of the system to the operating environment, interaction with ambient light, sound, and weather, a problem also noted for mobile AR games [63].
- Some military training is inherently dangerous; training aircrews to handle in-flight fires is costly and hazardous. AR offers the potential to reduce both the risk and the cost of such training [64]. This argument is also supported above with respect to the training of FACs and FOs. This is a frequently cited argument in favor of investigating AR as a technology in military contexts.
- Another theme that led to challenges for multiple applications was how to fit into existing processes or procedures and how to tap into existing data sources. While AR and VR technologies have long been noted as revolutionary approaches, there are cultural barriers to radical changes [25], whereas evolutionary improvements in the context of existing approaches are more likely to find support.
- We observe two visual styles of AR in the applications described in this chapter. Information overlaid directly on an object of interest may communicate more effectively, but it can also obscure the user's critical field of view. (This issue was important to the SMART program described in Sect. 25.2.3, as well as the research on which it was based [26]. It was also noted in the SA application of BARS.) Information placed at the periphery requires a cognitive integration step, but should not obscure the critical real-world elements. Appropriate use of each style depends heavily on the scenario and type of information being displayed to the user. At present, most commercially available displays are unable to support both simultaneously, due to limitations of the field of view. AR developers must make a deliberate choice as to which type of information display is more important in their application, and select the appropriate display device. It may even be that different users will prefer different styles. The issue of cognitive improvement, especially during training, is also of interest in other military contexts [65].
- A related issue with many AR applications is that the amount of virtual content can quickly become overwhelming for the user, a problem which has also been noted in

manufacturing contexts [66] and entertainment applications [67]. This results in degraded user performance, a condition known as "information overload." While filtering methods mitigated this condition in the BARS application for providing SA [30] and design choices sought to avoid this in some maintenance and repair applications, designers must be very conscious about how much and how visually stimulating or startling the virtual content on AR displays will be.
- While the focus of AR technologies is primarily on the visual display, the audio portion of AR contributes more than expected. Having cues "read aloud" to the user or directional sounds that accompany directional cues can greatly enhance the AR effectiveness, and in some circumstances compensate for user disabilities.
- Natural manipulation of virtual objects, perhaps with tactile feedback, led multiple applications to implement gesture-based controls as part of the UI. Certain applications (such as the operational and training contexts for BARS) have a defined set of gestures that are already in use for face-to-face collaboration. Other applications, such as the therapy games and object selection/inspection applications (Sect. 25.2.4), might have more freedom in the design space. But these interfaces require significant improvement to be accepted in Naval domains. Latency and error in tracking hand motions form one challenge for these technologies, again subject to environmental factors such as lighting. But these issues may be most important for the maintenance applications, where the development of muscle memory from learning tasks to be repeated can be a potential benefit of virtual training.

There are certainly other lessons that are demonstrated by this body of work, many of which are also not unique to the Naval work documented here. We have seen the value of basic research and the importance of studying the human factors of the intrinsically interactive concept of AR. We have seen the adoption of user-centered design on a wider scale in military applications of the technology. While this was long a component of the development of BARS, we are encouraged by its wider adoption. We have demonstrated the importance of hardware advances to the success of AR applications, and contributed to the body of requirements recommended to AR hardware developers.

25.4 Conclusions and Future Directions

Multiple Navy agencies continue to conduct basic research, develop and demonstrate new concepts, and test prototype AR systems. While many concepts have been demonstrated, we need to continue this work in order to produce robust implementations and data-based evidence that these conceptual applications reduce risk, reduce cost, or improve capability in such a way that clearly demonstrates the return on investment. We believe AR developers within government and industry would be wise to focus on measurable improvements in capability, reductions in mistakes, reductions in time to complete tasks, and reductions in cost associated with the application of AR technology. The Navy and USMC are most likely to incorporate AR into training first, then maintenance and repair, and eventually into operations as AR becomes more capable, reliable, and robust. Industrial applications of AR on a wide scale will most likely precede the Navy's wide acceptance of this new and promising technology.

Supplementary Video Material
Examples of the applications described in this chapter can be seen in the following video(s):
https://www.youtube.com/watch?v=kpUHslpU1sI
https://www.facebook.com/1843698365865999/videos/1121203451636360/?__so__=channel_tab
https://cdn.dvidshub.net/media/video/1510/DOD_102814876/DOD_102814876-1024x576-1769k.mp4
https://www.dvidshub.net/video/421182/augmented-immersive-team-trainer-aitt
https://www.youtube.com/watch?v=iO4aLSbzC-o
https://www.youtube.com/watch?v=Bp36MelcjfQ

Acknowledgments The authors wish to thank Steve Ellis, Steve Feiner, Henry Fuchs, Lawrence Rosenblum, Peter Squire, and Natalie Steinhauser. Interested parties may contact Mark A. Livingston regarding information in this chapter. The opinions offered in this chapter are those of the authors; they do not represent official positions of the Navy, USMC, DoD, or US government. Distribution Statement A: Approved for public release, distribution is unlimited.

References

1. Newman, R.L.: *Operational Problems Associated with Head-Up Displays During Instrument Flight*. Air Force Aerospace Medical Research Laboratory, TR–80–116 (1980)
2. Weintraub, D.J., Ensing, M.: *Human Factors Issues in Head-Up Display Design: The Book of HUD*, 92–102. University of Dayton Research Institute SOAR, Dayton (1992)
3. Fitts, P.M.: Psychological requirements in aviation equipment design. J. Aviat. Med. **17**, 270–275 (1946)
4. Kim, J.: The origin of the see-through graphical interface: World War II aircraft gunsights and the status of the material in early computer interface design. Convergence: The International Journal of Research into New Media Technologies **21**(2), 213–227 (2015)
5. Ellis, S.R.: Towards determination of visual requirements for augmented reality displays and virtual environments for the airport tower. In: *NATO Human Factors and Medicine Panel Workshop on Virtual Media for Military Applications* (2006)
6. Fischer, E., Haines, R.F., Price, T.A.: Cognitive issues in Head-Up displays. *National Aeronautics and Space Administration*, Technical Paper 1711 (1980)
7. Ellis, S.R.: Personal communication (2020)
8. Naish, J.: A review of some Head-Up display formats. In: *National Aeronautics and Space Administration*, Technical Paper 1499 (1979)

9. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. Environments **6**(4), 355–385 (1997)
10. Baxter, J., Workman, J.: *Review of Project Displays of Flight Information and Recommendations for Further Development Department of Supply, Australian Defence Scientific Service, Aeronautical Research Laboratories, Human Engineering Report 2* (1962)
11. Sutherland, I.E.: The ultimate display. In: *Information Processing 1965, Proceedings of IFIP Congress 65*, pp. 506–508 (1965)
12. Sutherland, I.E.: A head-mounted three-dimensional display. In: *Proceedings of the 1968 Fall Joint Computer Conference*, vol. 33, pp. 757–764. Thompson Book Co., London (1968)
13. Salas, E. (ed.): *Situational Awareness*. Routledge, London (2011)
14. Furness III, T.A.: The application of head-mounted displays to Airborne Reconnaissance and weapon delivery. In: *Proceedings of Symposium for Image Display and Recording*. Also available as U.S. Air Force Avionics Laboratory Technical Report TR–69–241 (1969)
15. Birt, J.A., Furness III, T.A.: Visually coupled systems. Air University Review **25**(3), 28–40 (1974)
16. Furness III, T.A.: The super cockpit and its human factors challenges. In: *Proceedings of the Human Factors Society—30th Annual Meeting* (1986)
17. Benton, C., Koonce, J.: Augmented reality: a new tool to increase safety in maritime navigation. https://pdfs.semanticscholar.org/06f6/8fee97f291236cb5aaee8fc14442a707f2b1.pdf (No date)
18. Geoghegan, B.J.: *Navigational Heads-Up Display: Will a Shipboard Augmented Electronic Navigation System Sink or Swim?*. M.S. Thesis, Naval Postgraduate School, California (2015)
19. Feiner, S., MacIntyre, B., Höllerer, T., Webster, A.: A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In: *International Symposium on Wearable Computing (ISWC)*, pp. 74–81 (1997)
20. Höllerer, T., Feiner, S., Pavlik, J.: Situated documentaries: embedding multimedia presentations in the real world. In: *International Symposium on Wearable Computing (ISWC)*, pp. 79–86 (1999)
21. Wang, X., Ong, S., Nee, A.: A comprehensive survey of augmented reality assembly research. Adv. Manuf. **4**, 1–22 (2016)
22. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. Robot. Comput. Integr. Manuf. **49**, 215–228 (2018)
23. Ong, S., Yuan, M., Nee, A.: Augmented reality applications in manufacturing: a survey. Int. J. Prod. Res. **46**, 2707–2742 (2008)
24. Caudell, T.P., Mizell, D.W.: Augmented reality: an application of heads up display technology to manual manufacturing processes. In: *Proceedings of Hawaii International Conference on System Sciences*, vol. II, pp. 659–669 (1992)
25. Curtis, D., Mizell, D., Gruenbaum, P., Janin, A.: Several devils in the details: making an AR application work in the airplane factory. In: *Proceedings of the First International Workshop on Augmented Reality*, pp. 47–60 (1999)
26. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE Trans. Vis. Comput. Graph. **17**(10), 1355–1368 (2011)
27. Julier, S.J., Baillot, Y., Lanzagorta, M., Brown, D.G., Rosenblum, L.J.: BARS: battlefield augmented reality system. In: *NATO Symposium on Information Processing Techniques for Military Systems* (2000)
28. Livingston, M.A., Rosenblum, L.J., Julier, S.J., Brown, D.G., Baillot, Y., Swan II, J.E., Gabbard, J.L., Hix, D.: An augmented reality system for military operations in Urban Terrain. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (2002)
29. Livingston, M.A., Brown, D., Julier, S.J., Schmidt, G.S.: Military applications of augmented reality. In: *NATO Human Factors and Medicine Panel Workshop on Virtual Media for Military Applications* (2006)
30. Livingston, M.A., Rosenblum, L.J., Brown, D.G., Schmidt, G.S., Julier, S.J., Baillot, Y., Swan II, J.E., Ai, Z., Maassel, P.: Military applications of augmented reality. In: *Handbook of Augmented Reality*. Springer, Berlin (2011). Chap. 31
31. Rosenblum, L.J., Feiner, S., Julier, S.J., Swan II, J.E., Livingston, M.A.: The development of mobile augmented reality. In: *Expanding the Frontiers of Visual Analytics and Visualization*, pp. 431–448. Springer, Berlin (2012)
32. Livingston, M.A., Swan II, J.E., Julier, S.J., Baillot, Y., Brown, D.G., Rosenblum, L.J., Gabbard, J.L., Höllerer, T.H., Hix, D.: Evaluating system capabilities and user performance in the battlefield augmented reality system. In: *Performance Metrics for Intelligent Systems Workshop* (2004)
33. Livingston, M.A., Dey, A., Sandor, C., Thomas, B.H.: Pursuit of 'X-ray Vision' for augmented reality. In: *Human Factors in Augmented Reality Environments*. Springer, Berlin (2012). Chap. 4
34. Gans, E., Roberts, D., Bennett, M., Towles, H., Menozzi, A., Cook, J., Sherrill, T.: Augmented reality technology for day/night situational awareness for the dismounted soldier. In: *Proceedings of SPIE Display Technologies and Applications for Defense, Security, and Avionics IX; and Head- and Helmet-Mounted Displays XX*, vol. 9470 (2015)
35. Oskiper, T., Sizintsev, M., Branzoi, V., Samarasekera, S., Kumar, R.: Augmented reality scout: joint unaided-eye and telescopic-zoom system for immersive team training. In: *IEEE International Symposium on Mixed and Augmented Reality*, pp. 25–30 (2015)
36. You, X., Zhang, W., Ma, M., Deng, C., Yang, J.: Survey on Urban warfare augmented reality. Int. J. Geo Inf. **7**(2), 46 (2018)
37. Kase, S., Roy, H., Bowman, E., Patton, D.: Augmented reality enabling intelligence exploitation at the edge. In: *Proceedings of SPIE Display Technologies and Applications for Defense, Security, and Avionics IX; and Head- and Helmet-Mounted Displays XX*, vol. 9470 (2015)
38. South, T.: Army's Do-it-all Goggles to Reach Soldier's Hands in 2021. *DefenseNews.com* (2020); last accessed 12 February 2021. https://www.defensenews.com/news/your-army/2020/10/11/armys-do-it-all-goggles-to-reach-soldiers-hands-in-2021/
39. Mittal, V.: Army's Augmented Reality Goggles Show It May Have Finally Cracked Code for Developing Electronics. *Forbes.com* (2020); last accessed 12 February 2021. https://www.forbes.com/sites/vikrammittal/2020/12/20/ivas-program-shows-that-army-can-move-at-speed-of-technology/?sh=c2b645b4ab69
40. Norman, D.: *User Centered System Design*. CRC Press, New York (1986)
41. Livingston, M.A., Gabbard, J.L., Swan II, J.E., Sibley, C.M., Barrow, J.H.: Basic perception in head-worn augmented reality displays. In: *Human Factors in Augmented Reality Environments*. Springer, Berlin (2012). Chap. 3
42. Baillot, Y., Julier, S.J., Brown, D.G., Livingston, M.A.: A tracker alignment framework for augmented reality. In: *International Symposium on Mixed and Augmented Reality (ISMAR '03)* (2003)
43. Julier, S.J., Lanzagorta, M., Baillot, Y., Brown, D.G.: Information filtering for mobile augmented reality. IEEE Comput. Graph. Appl. **22**(5), 12–15 (2002)
44. Schmidt, G.S., Brown, D.G., Tomlin, E.B., Swan II, J.E., Baillot, Y.: Toward disambiguating multiple selections for frustum-based pointing. In: *IEEE Symposium on 3D User Interfaces* (2006)
45. Coelho, E.M., Julier, S.J., MacIntyre, B.: OSGAR: a scene graph with uncertain transformations. In: *IEEE International Symposium on Mixed and Augmented Reality* (2004)
46. Muller, P.: The future immersive training environment (FITE) JTCD: improving readiness through innovation. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (2010)

47. Military.com: *New Grunt-Invented Kit Could Help Prepare Marine Corps for War* (2017). https://www.military.com/defensetech/2017/10/03/new-grunt-invented-kit-help-prepare-marine-corps-war, last accessed 20 Feb 2020
48. Zhu, Z., Branzoi, V., Sizintsev, M., Vitovitch, N., Oskiper, T., Villamil, R., Chaudhry, A., Samarasekera, S., Kumar, R.: AR-Weapon: live augmented reality based first-person shooting system. In: *Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 618–625 (2015)
49. Jones, N., Odermann, K., Squire, P., Read, A., Steinhauser, N.: Creating effective LVC training with augmented reality. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (2017)
50. Wilson, K.L., Doswell, J.T., Fashola, O.S., Debeatham, W., Darko, N., Walker, T.M., Danner, O.K., Matthews, L.R., Weaver, W.L.: Using augmented reality as a clinical support tool to assist combat medics in the treatment of tension pneumothoraces. Mil. Med. **178**(9), 981–985 (2013)
51. Roussel, M.: *Army Researchers Developing Augmented Reality Software for Field Care* (2019). https://www.army.mil/article/223561/, last accessed 28 February 2020
52. National research council committee on improving the decision making abilities of small unit leaders. In: Popp, R.L., Williams, M.J. (ed.). *Co-chairs: Improving the Decision Making Abilities of Small Unit Leaders*. National Academies Press, New York (2012)
53. Marine Corps Rapid Capabilities Office et al.: Tactical Decision Kit (TDK) System Overview. In: *marines.mil* (2017), last accessed 12 February 2021. https://www.mcwl.marines.mil/Portals/34/Tactical%20Decision%20Kit%20-%20Overview.pdf?ver=2017-06-15-134634-927
54. *U. S. Department of Defense Interface Standard: Joint Military Symbology (Revision D)*. U.S. Department of Defense, New York (2014). https://quicksearch.dla.mil/Transient/27F99A139271477EA221C408C728BE81.pdf
55. Lam, J., Navarro, P.: Secured mobile maintenance and operator system including wearable augmented reality interface, voice command interface, and visual recognition systems and related methods, U.S. Patent 9,746,913 (2017)
56. Naval Surface Warfare Center, Port Hueneme Division. https://www.navsea.navy.mil/Home/Warfare-Centers/NSWC-Port-Hueneme/What-We-Do/In-Service-Engineering/, last accessed 28 February 2020
57. Livingston, M.A., Ai, Z., Decker, J.W.: Human factors for military applications of head-worn augmented reality displays. In: *Advances in Human Factors in Simulation and Modeling, Proceedings of Applied Human Factors and Ergonomics 2018* (2018)
58. Li, C., Li, Z., Wang, Z., Xu, Y., Luo, M.R., Cui, G., Melgosa, M., Brill, M.H., Pointer, M.: Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. Color. Res. Appl. **42**, 703–718 (2017)
59. Hassani, N., Murdoch, M.J.: Investigating color appearance in optical see-through augmented reality. Color. Res. Appl. **44**(4), 492–507 (2019)
60. Konica M.: CS-200 Color and Luminance Meter. In: *konicaminolta.us*, last accessed 12 February 2021. https://sensing.konicaminolta.us/us/products/cs-200-color-and-luminance-meter/
61. nVis, Inc.: nVisor ST Head-Mounted Display. In: *nvisinc.com*, last accessed 12 February 2021. https://www.nvisinc.com/product/products.html
62. Kvavle, J., Wright, C., Calhoun, E., Countryman, C., White, D., Dang, T., Deputy, T., Ha, I., Rozovski, D., Leppen, J., McPherson, R., Mannion, J., Navarro, P., Struab, E., Vukelic, J., Violante, P., Walker, D.: Augmented reality in the U.S. Navy: a roadmap. In: *SPAWAR Systems Center Pacific Technical Report 3014* (2016)
63. Wetzel, R., Blum, L., Broll, W., Oppermann, L.: Designing mobile augmented reality games. In: *Handbook of Augmented Reality*. Springer, Berlin (2011). Chap. 25
64. Mayberry, C.R., Jaszlics, S., Stottlemyer, G., Fritz, G.: Augmented reality training application for C-130 Aircrew training system. M&S J. (2014)
65. Brunyé, T.T., Brou, R., Doty, T.J., Gregory, F.D., Hausey, E.K., Lieberman, H.R., Loverro, K.L., Mezzacappa, E.S., Neumeier, W.H., Patton, D.J., Soares, J.W., Thomas, T.P., Yu, A.B.: A review of US army research contributing to cognitive enhancement in military contexts. J. Cogn. Enhanc. **4**(4), 453–468 (2020)
66. Ramalho, F., Soares, A.L.: Augmented reality in complex manufacturing systems as an informational problem: a human-centered approach. In: *iConference 2020 Proceedings* (2020)
67. Giering, S.v.d.A.L., Schneegass, C., Ludwig, M.: The *SpaceStation App*: design and evaluation of an AR application for education television. In: *ACM International Conference on Interactive Media Experiences*, pp. 24–33 (2020)

Mark A. Livingston is a computer scientist at the Naval Research Laboratory. He earned his Ph.D. while researching augmented reality at the University of North Carolina at Chapel Hill, served as Program Chair of the International Symposium on Mixed and Augmented Reality, helped develop the Battlefield Augmented Reality System (BARS), and led the study of human factors in BARS.

Josh M. Kvavle is an electrical engineer at Naval Information Warfare Center Pacific. He earned his Ph.D. with a focus on phonics systems from Brigham Young University, and has worked on applications of augmented reality in the Navy with an emphasis on end user engagement throughout the development process.

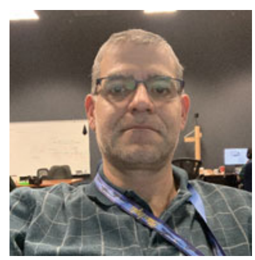
Corey Countryman is a systems engineer and CISSP within the Human Performance Engineering Branch at the Naval Undersea Warfare Center in Keyport, Washington. Most recent work includes developing Augmented and Virtual Reality Training Systems and Integration and Interoperability (I&I) guidance for DASN(RDT&E). He has a BSEE from the University of Idaho, and Systems Engineering certificate from the Naval Postgraduate School.

Dr. Kendy Vierling is a human performance scientist and the Director of the Naval Education Division at the United States Department of the Navy. She has led learning modernization, human performance, science and technology, strategic planning, policy, and assessment initiatives to enhance military learning, performance, and resilience. She has served as a Defense Advanced Research Projects Agency (DARPA) Service Chiefs Fellow.

Nathan Jones was the Manpower, Personnel, and Training Technical Support Lead at Marine Corps Systems Command Program Management in Training Systems. Mr Jones has over 16 years of experience in human performance, human-systems integration, and assessments. He oversaw training front-end analyses, verification, validation, and accreditations, training effectiveness evaluations, and training domain expertise for PM-TRASYS.

Michael Confessore is a computer scientist and mechanical engineer at the Naval Air Warfare Center-Aircraft Division (NAWC-AD) in Lakehurst, NJ. He earned his bachelor's degree from NYU Tandon and his masters from Monmouth University. He serves as the training lead for the NAWC-AD Creative Lab where he has managed numerous R&D efforts in the fields of virtual, mixed, and augmented reality.

Ryan Bowersock is the founder of the CREATIVE Lab at NAWC-AD Lakehurst. He earned a bachelor's degree for Mechanical Engineering from the University of Cincinnati and a master's degree for Engineering Management from Drexel University. He's been with NAWC-AD for 18 years. He manages the core CREATIVE Lab Team and oversees all research and SE/ALRE trainer development/deployment in the lab.



Aristides Staikos is the Lead Augmented/Virtual/Mixed Reality Researcher for the CREATIVE Lab at the Naval Air Warfare Center-Aircraft Division (NAWC-AD) in Lakehurst, NJ. He has led the research and development of ARRMSS for the last 2 years. He has worked as a Department of Defense Civilian engineer for the last 21 years with experience in Army Tactical Wireless Communications Networks and Naval Aviation Support Equipment.

Augmented and Mixed Reality for Shipbuilding

26

Tiago M. Fernández-Caramés and Paula Fraga-Lamas

Contents

26.1 Introduction 644

26.2 State of the Art 644
26.2.1 Shipbuilding Welding 644
26.2.2 Shipbuilding Painting 645
26.2.3 Shared Information 645
26.2.4 Step-by-Step Guidance 645
26.2.5 Design and Construction Assistance 645
26.2.6 Commercial Developments 645

26.3 Potential Shipbuilding Tasks to Be Enhanced with AR/MR 646
26.3.1 Quality Control 646
26.3.2 Guided Manufacturing 646
26.3.3 Product and Tool Tracking 646
26.3.4 Warehouse Management 647
26.3.5 Predictive Maintenance 647
26.3.6 Augmented and Mixed Reality Communications 647
26.3.7 Hidden Area Visualization 647
26.3.8 Monitoring and Interaction with IIoT Devices 648
26.3.9 Easy Interaction with Advanced Industrial Software 648
26.3.10 Structure Visualization 648
26.3.11 Training 648
26.3.12 Product and Tool Maintenance 649

26.4 AR/MR Architectures for Shipbuilding Applications 649
26.4.1 Traditional AR/MR Communications Architectures 649
26.4.2 Advanced Communications Architectures 651

26.5 AR/MR Hardware and Software for Shipbuilding Applications 651
26.5.1 Ideal AR/MR Device Characteristics 651
26.5.2 AR/MR Devices for Shipbuilding Applications 653
26.5.3 Software AR/MR Frameworks for Shipbuilding 658
26.5.4 Lessons Learned from the Shipbuilding State of the Art 661

26.6 Main Challenges for the Development of AR/MR Shipbuilding Applications 661

26.7 Conclusions 664

References 664

Abstract

Industry 4.0 is paving the way for the automation of many industrial fields that can improve the efficiency of manufacturing processes. Such an improvement can be remarkable on industries that involve a relevant number of complex processes, which need to be optimized individually in order to increase their overall performance. Shipbuilding is one of such industrial fields that can benefit from the application of the principles of Industry 4.0 and from the use of the latest technologies. Among the different Industry 4.0 technologies, augmented reality (AR) and mixed reality (MR) can be especially helpful for enhancing shipbuilding tasks, since they can be applied to many of them, while providing useful and attractive visual interfaces that enable shipyard operators to receive information on the tasks they are working on and that allow them to interact with physical and virtual elements. This chapter first reviews the state of the art on the application of commercial and academic AR/MR solutions to shipbuilding. Moreover, the most relevant shipbuilding tasks to be enhanced with AR/MR and their challenges are analyzed. Furthermore, this chapter also details the latest and most promising AR and MR hardware, software, and communication architectures aimed at being deployed in shipyard workshops and on ships under construction. As a result, this chapter provides a thorough review of the most recent developments on the application of AR and MR to shipbuilding and includes useful guidelines for future developers.

Keywords

Augmented reality · Mixed reality · Shipbuilding · Industry 4.0 · Human-machine interaction · IIoT · SDK · Microsoft HoloLens · Fog computing · Edge computing

T. M. Fernández-Caramés (✉) · P. Fraga-Lamas (✉)
Dept. of Computer Engineering, CITIC Research Center, Universidade da Coruña, A Coruña, Spain
e-mail: tiago.fernandez@udc.es

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_26

26.1 Introduction

Shipbuilding companies provide solutions that allow for designing and constructing vessels. Specifically, shipbuilders work in fields like construction technologies, naval maneuvering systems, combat solutions, vessel repair techniques, or turbine manufacturing. Such fields can be enhanced by Industry 4.0 [1, 2], thus enabling the creation of Shipyard 4.0, which can make use of the latest technologies related to the Industrial Internet of Things (IIoT), artificial intelligence, robotics, additive manufacturing, Big Data techniques, and, recently, blockchain [3] or post-quantum IoT [4], in order to improve their process efficiency.

The concepts behind the Industry 4.0 paradigm are expected to optimize the performance of shipyard operators, which will be required to make use of human-machine interfaces with human-centered design [5] to carry out their daily tasks. An AR/MR device is one of such interfaces that will help operators by providing guidance in assembly processes, useful contextual information, or easy human-machine interaction [6]. The advantages of the use of AR/MR will be exclusive not only to the shipbuilding industry but to other sectors that, jointly with Virtual Reality (VR), are expected to create in 2025 a market of US $80 bn [7].

Nonetheless, there has been a long way for AR/MR to reach its current development stage. The first properly documented industrial AR/MR developments were presented by Boeing in the 1990s with the objective of helping factory workers with manufacturing and assembly processes [8]. However, AR/MR technologies were not pushed firmly until the late 1990s, when the German government funded the ARVIKA project [9, 10], which was aimed at developing mobile AR interfaces for relevant industrial companies like Audi, BMW, Airbus, or Chrysler. After ARVIKA, AR/MR systems progressed at a relatively slow pace but allowed for providing useful industrial applications related to the design [11, 12] or assembly [13–16] of products.

This chapter is aimed at analyzing the most relevant applications, challenges, and components of successful AR/MR shipbuilding applications. Specifically, the rest of this chapter is structured as follows. Section 26.2 reviews the latest and most relevant AR/MR applications for the shipbuilding industry, including those related to welding or painting tasks or those used for guiding operators during assembly and maintenance procedures. Section 26.3 describes potential shipbuilding tasks that can be enhanced through AR/MR, while Sect. 26.4 details the components that conform to traditional and advanced architectures that are able to fulfill the communications requirements of the latest AR/MR applications deployed in shipyards and ships under construction. Section 26.5 reviews the most recent and relevant hardware and software that can be used for developing AR/MR shipbuilding applications. Section 26.6 enumerates the main challenges related to the adoption of AR/MR solutions by the shipbuilding industry. Finally, Sect. 26.7 is devoted to the conclusions of the chapter.

26.2 State of the Art

During the last years, diverse authors and companies have proposed AR/MR systems with the objective of helping shipyard operators and shipbuilding companies in their daily tasks. Since most of the best documented solutions come from academia, the following subsections, except the last one, are related to noncommercial systems, while the last one refers to the initiatives carried out by IT or shipbuilding companies.

Some recent academic works provide a general vision of AR/MR in relation to the shipbuilding industry. For example, a literature review focused on future research opportunities is presented in [17]. Other authors focus on the role of emerging technologies like AI, IoT, AR/MR, or VR as enablers of the digital twin of the ship [18]. For instance, there are also country-specific works. The authors of [19] provide a thorough overview of potential applications in Norwegian shipyards. Finally, other authors provide a general overview of possible applications and briefly introduce the implementation of an application, e.g., sheet metal forming [20].

26.2.1 Shipbuilding Welding

Welding is a task that is essential in shipbuilding and is involved in numerous processes related to the construction of the blocks that conform a ship. AR and MR can help ease the welding tasks performed in a shipyard. For instance, the system detailed in [21] allows for replacing current welding screens with an AR helmet able to display relevant contextual information on welding processes (e.g., potential corrections, welding errors). In the work described in [22], a robot for onboard welding is presented. In such a system, the robot is controlled by the operator through a wireless controller, and visual information is sent using a projection system carried by the robot. Thus, the operator focuses on the welding task and avoids distractions related to the use of traditional screens. A third AR-based welding solution that is worth mentioning is detailed in [23]. In such a paper, the authors present a training system that makes use of AR glasses, a motion tracking system, external speakers, and a torch in order to simulate welding processes in real time. The presented system is also able to quantify the quality of the welding through the use of the inputs from the torch and an artificial neural network.

26.2.2 Shipbuilding Painting

Painting is also an activity that is usually performed in a shipyard. To train operators in such an activity, some authors suggested a solution for AR glasses that makes use of a paint gun to simulate the interaction with virtual steel structures during the painting process [24]. The paint gun is actually quite sophisticated: it includes force feedback and emits painting sound depending on the trainee interactions. Thanks to the proposed system, no real paint or steel are wasted, and it is possible to qualify objectively the work of the trainee immediately after the completion of the painting exercise.

26.2.3 Shared Information

Operators need to communicate among them and share information in order to perform the different processes involved in manufacturing and maintenance procedures. AR and MR can help to provide a virtual communication channel where shipyard operators can share the required information. For instance, in [25], the authors describe an AR-based intercommunication solution that can be easily applied to the processes involved in a shipyard. Specifically, such an AR system enables shipyard workers to place virtual notes in specific locations or equipment. The proposed solution stores note information on a central server, and then the deployed mobile devices (e.g., tablets) retrieve the necessary data when performing maintenance procedures or during certain plant processes.

26.2.4 Step-by-Step Guidance

Most shipbuilding manufacturing processes that shipyard operators need to follow carefully are detailed in paper or electronic documents. Such documents need to be read when needed or are memorized for performing assembly or maintenance operations. Unfortunately, the mentioned operations are prone to human errors, so different researchers have suggested AR/MR-based systems to support them. For instance, in [26], the authors propose an AR guidance solution for tablets that provides step-by-step instructions. Another example is detailed in [27], where it is presented an AR assistance system for mechanics that need to maintain or repair armored vehicles on the field. Specifically, the proposed system consists of a wrist control interface and an AR display that provides contextual information (e.g., text notes, arrows, animations) to locate parts or to illustrate the operations to be performed.

26.2.5 Design and Construction Assistance

Although naval engineers do their best when designing a ship, discrepancies between the designed and the real ship are frequent, especially when building ships for the first time. In addition, in order to shorten ship production times, the design and construction phases can overlap in time. Such an overlapping may derive into discrepancies that are detected by the shipyard operators. Consequently, they need to be fixed by the designers in the CAD models so as not to repeat the same errors again. For example, to avoid these issues in the specific area of ship pipe construction, the authors of [28] propose to visualize virtual pipes in situ through an AR system for tablets. The characteristics of each virtual pipe can be adapted to the requirements of the real environment (with the help of an optical measuring tool), and then the adjusted characteristics can be sent to the pipe workshop to build pipes with the exact dimensions and features.

The interpretation of the documentation is also a common and relevant problem in shipbuilding, since it is essential for operators and engineers when understanding how to build a part or the involved technical requirements. To simplify such an interpretation, in [29], the authors describe an AR-based application for mobile devices that is able to show the 3D models created from scanned designs and whose characteristic points are identified. In the field, the developed application takes pictures of the actual parts with a mobile device camera and compares them with the stored designs in order to identify the part. If it is identified, the stored 3D model and its characteristic points are retrieved and visualized on the mobile device.

26.2.6 Commercial Developments

In the last years, several companies have led initiatives for the commercial developments of AR/MR applications for the shipbuilding industry. For instance, Newport News Shipbuilding [30] worked on the development of AR/MR applications aimed at decreasing the cost of different shipbuilding tasks related to training, operation, or maintenance procedures. Similarly, Index AR Solutions [31] has created AR/MR applications that can be used in both shipyards and ships in order to ease the construction, handling, or maintenance of certain products, tools, or machines.

It is also worth mentioning the work carried out in the field of AR and MR by two of the largest European shipbuilders: BAE Systems [32] and Navantia [33]. In the case of BAE systems, it has used AR devices during the construction of offshore patrol vessels and during the design of the Type 26 frigate. In the case of Navantia, it is immersed into an ambitious program called “Shipyard 4.0” that is aimed at transferring the principles of Industry 4.0 to its shipyards, including the use of AR and MR solutions for the construction of its F-110 frigates [34, 35].

26.3 Potential Shipbuilding Tasks to Be Enhanced with AR/MR

AR and MR applications can be deployed in industrial environments in order to ease the work of shipbuilding operators. To illustrate how such a work can benefit from the use of AR and MR devices, the next subsections provide different examples of tasks carried out in almost every modern shipyard.

26.3.1 Quality Control

Quality control is essential in shipbuilding and other hi-tech industries that build complex products. Currently, human quality control supervisors inspect industrial products visually and determine whether they pass the required quality requirements. AR and MR can help quality control processes by automating them through the use of computer vision techniques. For carrying out such tasks, product models need to be first obtained, which can be performed, for instance, through depth sensors or 3D cameras and reconstruction software (Fig. 26.1 illustrates an example of the extraction of the 3D model of a pipe for a quality control process). After obtaining the model, then it is possible to determine how accurate it is respect to its CAD model.

26.3.2 Guided Manufacturing

Many of the products that will become components of a ship are manufactured in the shipyard workshops. Manufacturing involves multiple stages and very precise steps that need to be followed to meet the expected quality requirements. AR and MR can be useful by guiding shipyard operators through the different manufacturing steps. Guidance is usually performed through the visualization of the manufacturing process through 3D models of the different stages of the product together with relevant virtual contextual information (e.g., animations, text documentation). In addition, it is possible to ease the AR/MR user experience of the operator through tangible interfaces placed on his/her workbench so that no computers or keyboards are necessary to interact with the virtual content.

As an example, Fig. 26.2 shows a screenshot of an AR application based on Vuforia [36] developed by the Joint Research Unit (JRU) Navantia-UDC (University of A Coruña) that helps shipyard operators in the assembly of a reduction gear of a turbine.

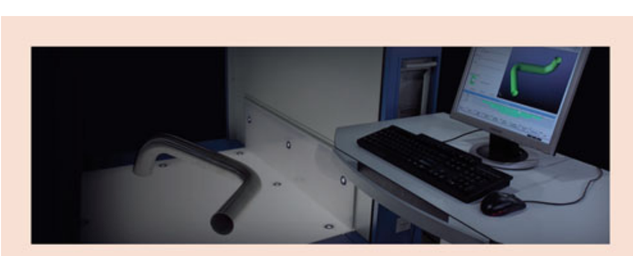


Fig. 26.1 Example of 3D model extraction for pipe quality control

26.3.3 Product and Tool Tracking

If AR/MR technologies are used in conjunction with real-time location systems (RTLs) based in technologies like RFID [37–39], it is possible to locate in three dimensions the surrounding tagged products and tools. Moreover, the described AR/MR systems may be used to visualize the values or the states of the sensors and actuators deployed throughout a shipyard. As an example, Fig. 26.3 illustrates an AR application for locating pipes inside a shipyard workshop.

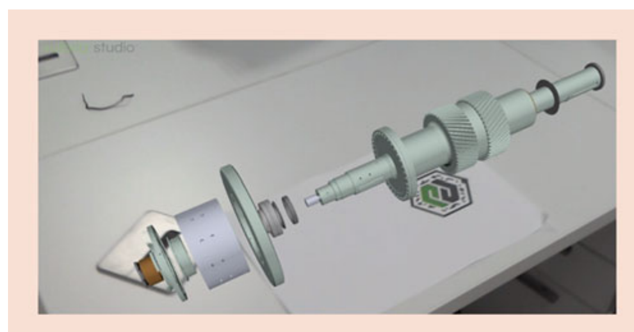


Fig. 26.2 AR application developed by the JRU Navantia-UDC for the assembly of a reduction gear

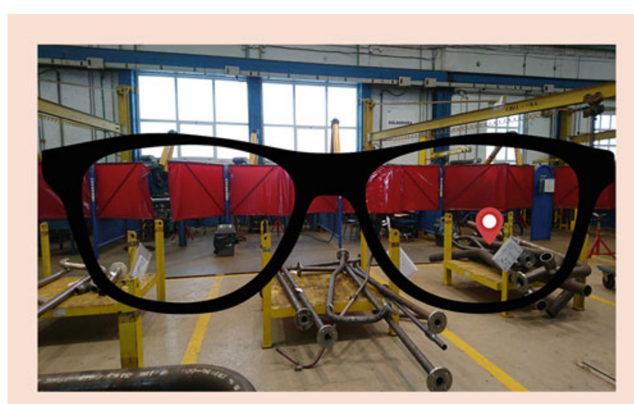


Fig. 26.3 AR application for locating pipes in a shipyard workshop

26.3.4 Warehouse Management

The components used by the multiple shipyard workshops are stored in warehouses, which are either located inside the workshops or in a remote central warehouse. AR/MR devices are able to assist shipyard operators when storing, locating, or relocating certain items. Such a guidance during warehouse processes allows for reducing human errors and for decreasing the required management time.

An example of AR/MR-based warehouse management application is shown in Fig. 26.4. Specifically, the figure illustrates an AR guiding application that shows in green the position of an item and the path that needs to be followed by the warehouse operator to collect it.

26.3.5 Predictive Maintenance

Predictive maintenance helps to prevent industrial problems with hardware and software components used in either the shipyard or onboard. Essentially, predictive maintenance collects information from human operators and sensors installed on industrial equipment and analyze it (e.g., by using Big Data or data mining techniques) in order to detect imminent or future faults. In such a field, AR and MR devices can be useful for predictive maintenance tasks when showing visually the potential faults to the operators. Furthermore, operators can even mitigate them through virtual interfaces that are linked to certain actions (e.g., to stop a machine that is going to fail soon and needs to fixed).

26.3.6 Augmented and Mixed Reality Communications

Shipyards usually occupy large areas that make it difficult to communicate shipyard operators and supervisors in a fast way. Such a communication frequently requires the personnel to walk or to drive significant distances or to make use of mobile phones. However, remote voice communications through mobile phones are prone to misunderstandings, and mobile data communications are mostly based on the exchange of pictures and simple messages.

AR and MR devices are able to augment personal communications by overlapping virtual elements with reality to provide accurate guidance. Moreover, AR/MR devices are also usually able to share the operator's point of view to show to the remote assistance personnel a precise view of what the operator is seeing and what exists around him/her. Furthermore, AR/MR devices enable recording the whole communication exchange (i.e., audio, video, contextual information) and store it securely for documenting the situation.

Figure 26.5 shows a screenshot of an application (Skype for Microsoft HoloLens) that allows for communicating an operator with a remote support worker. The figure actually shows what the operator is watching through the Microsoft HoloLens. At the bottom, on the right, it is shown the remote worker who can guide, send pictures, and include visual indications for the local operator. Specifically, the screenshot shows one of the instants when an operator wanted to carry out the maintenance of a pipe bending machine. During such a process, a remote worker was guiding the local operator through voice commands and by making use of a picture that illustrated the internal components of the bending machine.

26.3.7 Hidden Area Visualization

In certain areas of a workshop and in a ship, it is usually necessary to visualize structures (e.g., pipes, electrical wiring) that are located behind walls, ceilings, or bulkheads. An AR/MR system can overlap virtual elements to reality with the objective of revealing the internal structural components. Such a visualization can help during maintenance procedures and repairs. For instance, an AR/MR solution may help during the detection of an electrical failure in a vessel, since it is necessary for the operators to locate the places where the wiring is installed behind bulkheads before proceeding to their disassembly.

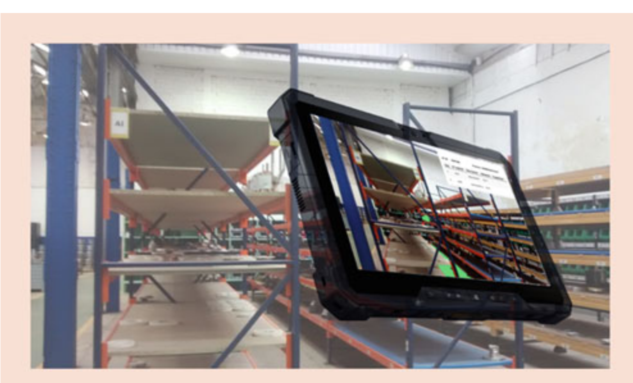


Fig. 26.4 AR application for guiding warehouse operators

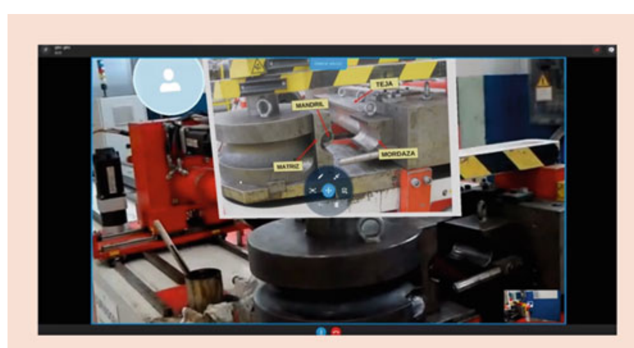


Fig. 26.5 Augmented communication application for Microsoft HoloLens developed by the JRU Navantia-UDC

26.3.8 Monitoring and Interaction with IIoT Devices

AR and MR are not really different in terms of the way they interact with IIoT devices (i.e., they both request data and send remote commands to IIoT devices), but the actions derived from such interactions differ: strictly speaking, AR only collects and displays information on IIoT devices, while MR also allows for interacting with them, resulting in physical changes to the user's environment (e.g., to start or stop an industrial machine). Such an interaction is usually carried out through virtual switches or control panels that can be operated through hand gestures, bottom presses, or voice commands.

It must be noted that AR and IIoT devices often use heterogeneous and legacy technologies that make them difficult to intercommunicate in an easy, flexible, and interoperable way. To solve the aforementioned issue, the authors of [40] propose an open-source framework to ease the integration of AR/MR systems and IoT devices as well as the transfer of information among them. The proposed framework makes use of widely used standard communication protocols and open-source tools.

26.3.9 Easy Interaction with Advanced Industrial Software

Industry 4.0 has led to the modernization of traditional industrial management software and has led to the deployment of advanced software like Product-Lifecycle Management (PLM), Enterprise Resource Planning (ERP), or Manufacturing Execution System (MES) software. Such a software has been traditionally accessed through computers or mobile devices (e.g., tablets, smartphone), but it can also be accessed through AR/MR devices, which allow for showing attractive visual interfaces, reducing the need for using keyboards, monitors, or fixed computers.

As an example, Fig. 26.6 shows a screenshot of an MR application for Microsoft HoloLens developed by the JRU Navantia-UDC that retrieves and shows the blueprints of a specific area of a ship. In such an application, the operator can watch different 2D plans of the ship and then show them in 3D at a real scale.

26.3.10 Structure Visualization

Like in other industries, in shipbuilding, it is useful to visualize structures in a virtual way prior to their deployment in order to evaluate how they will fit in a real environment. In addition, the visualization of the whole shipyard is interesting for marketing purposes or when making use of certain software like a digital twin, whose data can be fed into a virtual AR model.

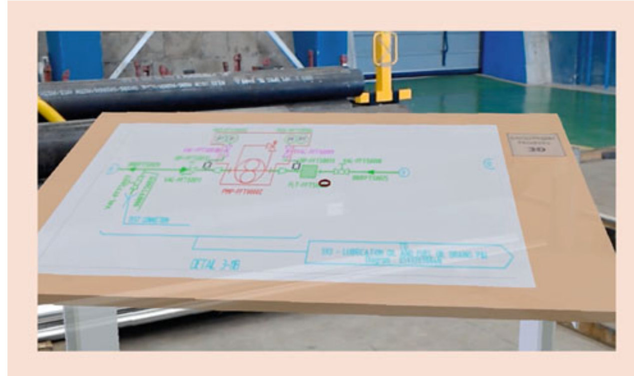


Fig. 26.6 Microsoft HoloLens application developed by the JRU Navantia-UDC for visualizing ship blueprints

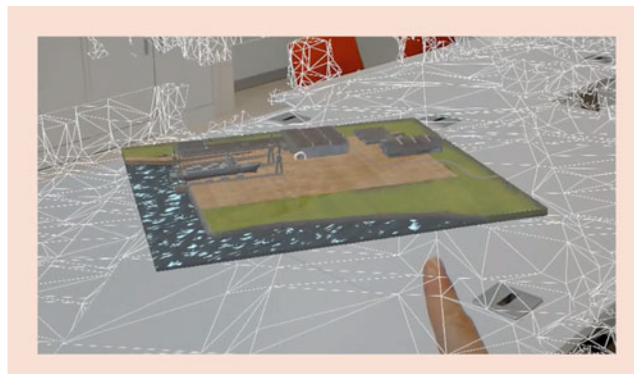


Fig. 26.7 Visualization of the plant of a shipyard through an MR application developed by the JRU Navantia-UDC

Figure 26.7 shows an example of an application for Microsoft HoloLens that enables visualizing the plant of a shipyard (in this case, of a shipyard of Navantia in Ferrol) and which can also be used as a kind of augmented reality digital twin.

26.3.11 Training

Shipbuilding involves a significant number of processes where operators need training to acquire the skills needed to create high-quality components and products. The development of part of such skills usually requires years of practical experience and the supervision of the most experienced operators. In fact, when the oldest operators retire, their experience and skills, most of which are not documented (or are difficult to document), are lost.

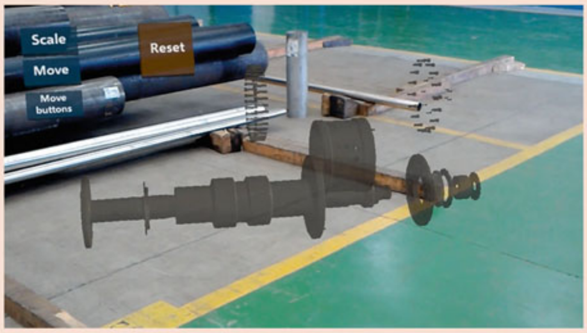


Fig. 26.8 AR application developed by the JRU Navantia-UDC for the maintenance of a hydraulic clutch

AR and MR can help during the training of new operators by avoiding the need for constant supervision and by providing an attractive learning interface. In addition, by monitoring the interaction of the learning operators, it is possible to evaluate their progress through the training. Furthermore, if the AR/MR is designed and implemented properly, the knowledge from the most skilled workers can be preserved (e.g., by including contextual expert tips or multimedia information).

26.3.12 Product and Tool Maintenance

Most of the machines and tools used in a shipyard need to be maintained with a specific frequency that may depend on their age or usage level. Moreover, the products built in a shipyard and that are installed in a ship also need to be revised periodically in order to guarantee their optimal performance. AR and MR can help during the maintenance process by providing a virtual interface that shows step-by-step instructions on how to disassemble a product and how to perform the maintenance procedure. As an example, Fig. 26.8 shows a screenshot of an AR application for Microsoft HoloLens [41] developed for Navantia for the maintenance of a hydraulic clutch.

26.4 AR/MR Architectures for Shipbuilding Applications

This section shows, for the data handling perspective, the limitations of traditional AR/MR communications architectures and describes the inner workings of both traditional and advanced AR/MR communications architectures together with some recommendations for their implementation and deployment.

26.4.1 Traditional AR/MR Communications Architectures

Traditional AR/MR architectures are essentially divided into three layers, which are depicted in Fig. 26.9:

- AR/MR Device Layer: it essentially consists of Head-Mounted Display (HMD) and Hand-Held Device (HHD) systems that provide visualization and interaction interfaces to the operators. Thus, such devices embed displays, microcontrollers, and other computational resources that allow for tracking physical or environmental markers and for rendering and showing to the user certain virtual graphical content. Two main operating scenarios are distinguished in the shipbuilding industry due to the significant differences in their working conditions: the shipyard (i.e., workshops and external infrastructures like storage areas or docks) and ships under construction [35]. For a shipyard and its workshops, the traditional architecture also considers the presence of spatial display devices. Such devices usually require to deploy projectors that, due to space constraints, are not easy to use in ships under construction.
- Data Routing Layer: it collects data from the AR/MR device layer and sends them to the shipbuilder's cloud. Thus, this layer is basically composed by network infrastructure, which has to deal with the problems that arise when transmitting wirelessly in an environment like a shipyard, where there are numerous large metallic objects and large structures. However, the most challenging environment is a ship under construction where there is a significant density of metallic objects that usually act as Faraday cages. Such an abundance of metals blocks wireless transmissions and impedes carrying out broadband communications. In addition, it must be noted that it is not common to deploy wired communications infrastructure in the ship until the last construction stages, so AR/MR devices commonly need to make use of wireless interfaces.
- Shipbuilder's cloud: it is a traditional cloud infrastructure that is usually hosted in the shipbuilder's premises. Such a cloud is able to run the essential software that is shared among the shipyard designers, operators, and the front office. Thus, the cloud usually runs MES, PLM, or ERP software besides design software (e.g., CAD-based software), industrial control and monitoring systems (e.g., SCADA systems), or IIoT platforms that may also provide digital twin features.

Although the traditional architecture has been already deployed by using different centralized computing solutions [42–44], its use may lead to certain response latency

26

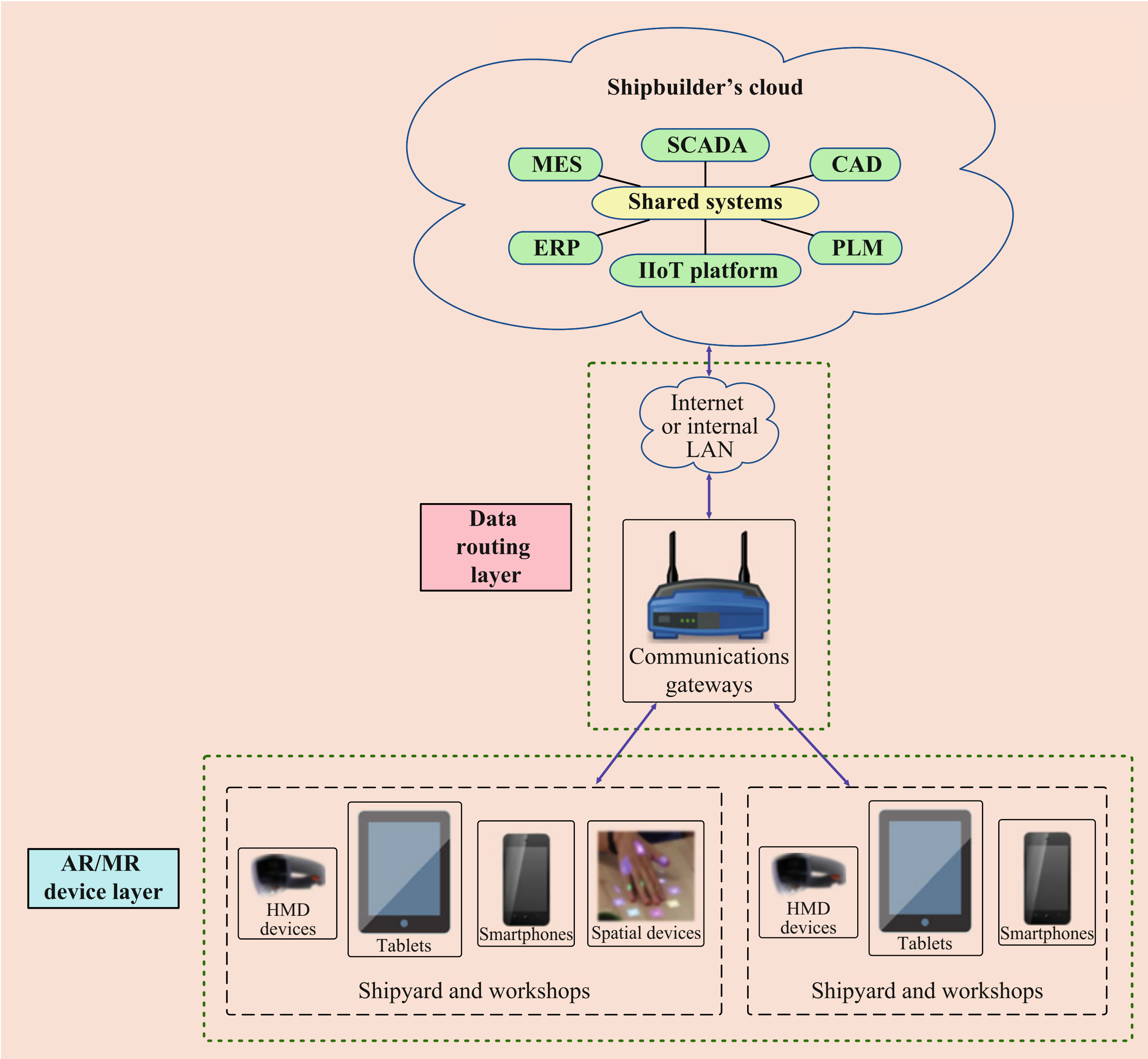


Fig. 26.9 Traditional AR/MR architecture

limitations. Note that AR/MR devices need to retrieve and show visually data as fast as possible in order to guarantee an attractive user experience. Due to this reason, AR/MR usually store most of the multimedia content locally, thus providing low-latency responses. However, this local storage strategy is only practical in applications that demand visualizing only very specific static information. On the contrary, many Industry 4.0 applications require AR/MR devices to exchange information dynamically with remote servers (e.g., for receiving data from an ERP) or IIoT networks (e.g., when it is necessary to show in real time the values of certain sensors deployed on the working environment [45–47] or carried by the operators [48]).

It is also worth mentioning that a traditional cloud-based architecture can include an AR/MR content caching server that stores the necessary information and then provides it to the AR/MR devices when needed. Such an approach requires the caching server to predict which content is more likely to be demanded by the deployed AR/MR devices and then to download it from the cloud before it is requested by the devices. It is possible to deploy the caching server on an AR/MR device, but it may impact energy consumption and computational power due to the currently existing hardware constraints. Therefore, it is recommended to enhance traditional architectures with additional layers in order to provide such caching services or to implement similar low-latency approaches.

26.4.2 Advanced Communications Architectures

Different authors have suggested alternative approaches to decrease response latency and improve rendering performance [49] in order to overcome the limitations of traditional AR/MR cloud-based architectures.

The most popular alternative consists in using local powerful computers that store AR/MR content and that perform video processing tasks [50, 51]. Other authors have proposed implementing edge computing-based solutions [52, 53], which make use of the design principles related to fog computing, cloudlets, and mobile edge computing [54] with a clear target: to move the processing power and the services provided by the cloud to the edge of the network, next to the AR/MR devices.

Regarding the application of fog computing in shipbuilding AR/MR applications, it consists in the deployment of processing and storage nodes (usually low-power devices like microcontroller-based nodes or single-board computers (SBCs) like Raspberry Pi or Orange Pi PC) throughout a shipyard [34]. Such nodes collect AR/MR device requests and respond to them when they are able to do so (otherwise they forward the request to the cloud or to other fog computing nodes). Moreover, the AR/MR data retrieved from the cloud at a certain time instant can be cached locally and then delivered to the AR/MR devices when they demand it. Mobile edge computing and cloudlets act in a similar way to fog computing, but the former deploys the processing nodes in cellular network-based stations, while the latter uses local high-end computers that are usually as powerful as cloud computing servers.

The existing literature describes different AR/MR edge computing systems, being one of the most relevant the one detailed in [53], which suggests offloading the most demanding AR/MR algorithms to a high-end computer. Thus, the authors are able to decrease significantly video transmission latency. Nonetheless, the researchers conclude that, although their solution works for HHD systems, it still needs to be optimized to support high data rates and small network latencies like the ones required by many HMD solutions.

Figure 26.10 illustrates an example of AR/MR edge computing-based architecture aimed at reducing response latency and at accelerating AR/MR content rendering. The architecture is similar to the traditional AR/MR architecture but includes a more complex data routing layer that makes use of fog computing and cloudlets.

Fog computing nodes are really helpful in AR/MR systems deployed in shipyards, since they are distributed physically throughout it (i.e., they are able to provide location-dependent services). Such nodes are able to run low-latency and QoS-aware applications, thus reducing the amount of network traffic that is commonly forwarded to the cloud. In addition, thanks to the processing capabilities of the fog computing nodes, they reduce the computational load of the cloud. Cloudlets help fog computing nodes when dealing with the most complex processing tasks and, therefore, decrease the cloud computational load significantly.

It is also worth mentioning that Fig. 26.10 also illustrates the ability of the architecture for interconnecting fog computing nodes and cloudlets between them, so they can collaborate when handling AR/MR device requests. Moreover, fog computing nodes can interact with IIoT networks in order to provide access to them and to collect sensor values that can be later processed and shown as contextual information by AR/MR devices. Furthermore, the architecture can be adapted to use other Industry 4.0 enabling technologies like Distributed Ledger Technologies (DLTs) such as blockchain [55–57], which have been already tested in shipbuilding scenarios [58].

Finally, it must be noted that, although it is not illustrated in Fig. 26.10, any modern AR/MR architecture for the shipbuilding industry should take cybersecurity seriously. For instance, a shipyard is often considered a critical infrastructure, especially in the case of shipyards where military vessels are built. Unfortunately, cybersecurity is usually neglected in AR/MR systems despite the efforts carried out by some authors [52]. Other authors emphasize the importance of preserving energy efficiency [59, 60] and of avoiding potential cyber-attacks that have already been performed against other Industry 4.0 technologies [61, 62].

26.5 AR/MR Hardware and Software for Shipbuilding Applications

This section reviews the AR/MR hardware device characteristics and AR/MR software framework features that should be provided in order to implement successful AR/MR shipbuilding applications.

26.5.1 Ideal AR/MR Device Characteristics

AR/MR hardware has evolved remarkably since the first industrial prototypes were released in the early 1990s [63]. Currently, commercial AR/MR devices have become less expensive and embed more computing power [64]. Moreover, the latest devices include several advanced features that are among the ideal characteristics that an AR/MR device should have when developing shipbuilding or similar smart manufacturing applications [65].

For instance, common commercial AR/MR devices support basic features like:

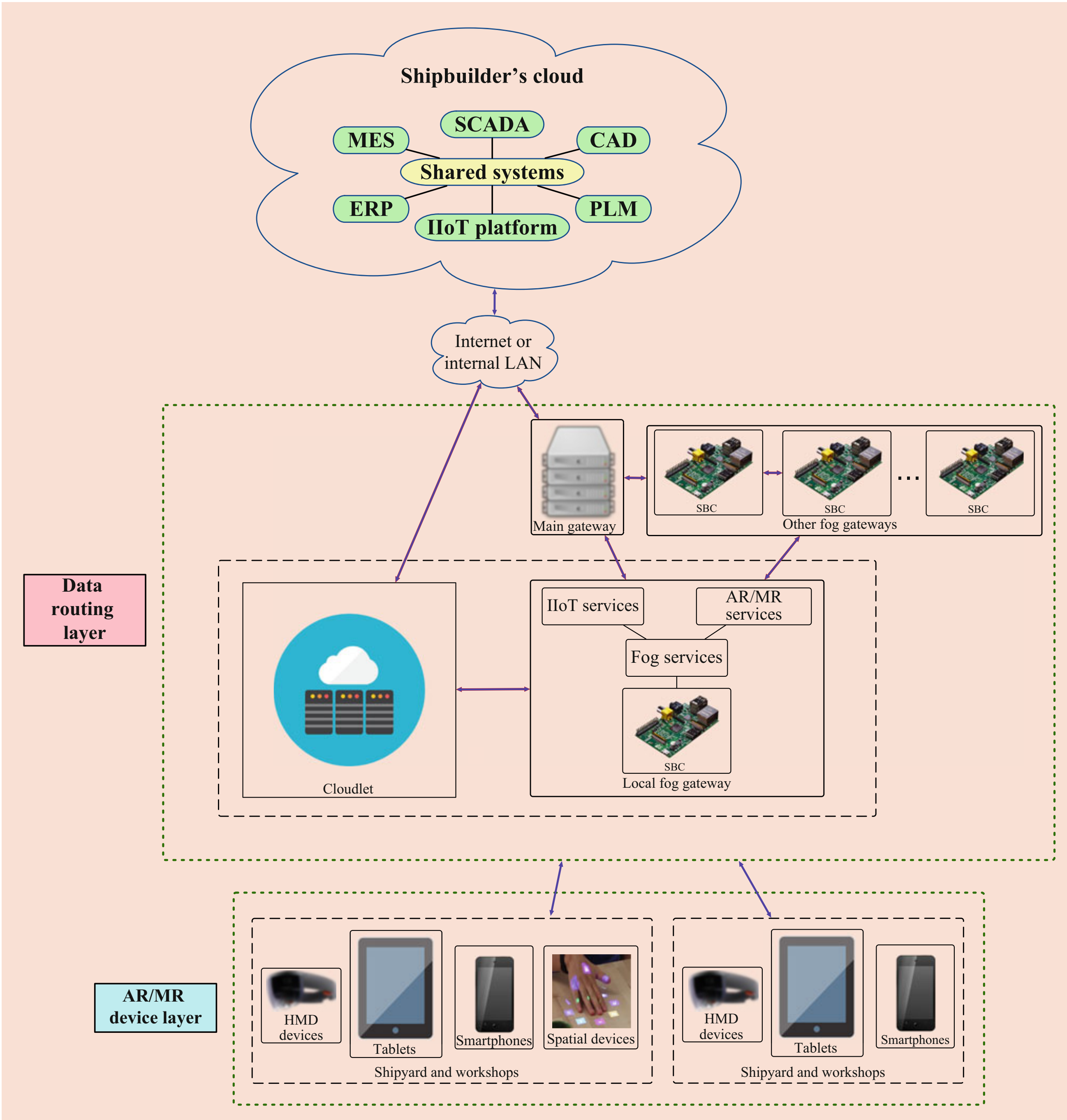


Fig. 26.10 Advanced edge computing-based AR/MR architecture

- Hands-free operation.
- Ability to show animations and multimedia content to indicate the exact point where a process needs to be performed on a product or machine (this is called "cueing").
- Direct interaction with the environment: the operator only needs to look at the desired element to make it possible to interact with it.

In contrast, more sophisticated devices are required in order to provide more advanced features like:

- Visualization of context-aware information for enriching the execution of certain tasks.
- It is possible to position the operator through different techniques (e.g., by using an embedded GPS or other

indoor location techniques [66]) and then guide him/her through the shipyard.

- Visualization of videos, animations, 3D models, and 2D schematics in order to enrich the operator user experience.
- Improved immersive experience that allows for increasing emotional engagement in the performed processes. To enhance such an experience, the ideal AR/MR device should provide:
  - A wide field of view. For a good experience, it is usually recommended to provide at least 30° horizontally.
  - A lightweight hardware interface that allows for wearing it through the whole working shift.
  - Long-lasting batteries able to power the AR/MR device for at least 8 h.
  - Optical or retinal projection, since a common problem is that video-based displays frequently incur in delays during image projection that harm user experience.
  - Voice-based interaction. This feature is especially useful when carrying out tasks that require operators to use both hands. Nonetheless, it should be evaluated properly the use of voice recognition in noisy environments like the ones typically found in industrial scenarios.

In addition, certain constraints should be taken into account when selecting an AR/MR hardware for developing shipbuilding applications:

- Current AR/MR smart glasses embed a limited amount of memory, which limits the physical area that they can map and store in real time. As a consequence, developers have to think of ways to load content dynamically, as soon as a shipyard operator enters a non-mapped area.
- Lighting is essential for providing a good user experience. This is due to the fact that many AR/MR techniques make use of computer vision algorithms that rely on thresholds and color measurements that are impacted by the brightness of the scenario, by the type of light source, or by the light temperature. This factor is especially critical for AR/MR devices that run marker-based software.
- Precise object and feature detection and tracking require a significant amount of computational power. Such a requirement can be difficult to implement in mobile and light devices, so new techniques and architectures will need to be designed and adapted to shipbuilding scenarios.
- It is also worth mentioning that the shipyard industrial machinery can produce electrical interference that may impact the reliability and performance of the sensors embedded into an AR/MR device. This may become a problem for smart glasses that enhance their pattern recognition algorithms with sensor measurements. For instance, smart glasses that rely on GPS receivers or WiFi/Bluetooth signals to position a shipyard worker usually have accuracy problems indoors, especially when there is a relevant presence of metallic objects.
- Certain AR/MR platforms make use of object tracking techniques that fuse information collected from sensors and from a camera. However, such tracking techniques are executed periodically, at specific time instants, but a shipyard and a ship under construction change continuously. Thus, such changes may mislead the AR/MR object positioning system.

26.5.2 AR/MR Devices for Shipbuilding Applications

Many shipbuilding tasks depend nowadays on the use of computers to retrieve and visualize data. However, in industrial scenarios, the execution of such tasks was in many cases constrained due to the use of ruggedized but cumbersome, non-mobile, and computationally limited hardware devices. Fortunately, tablets and smartphones evolved significantly in the last years, which increased operator mobility during the execution of certain tasks. However, tablets and smartphones require users to switch between the performed task and the operation of the device, which usually cause distractions. AR/MR HMD devices avoid most of such distractions by enabling hands-free operation and by overlapping reality and virtual content.

The mentioned benefits are some of the drivers of the estimated growth of AR/MR device sales: it is expected that 5.4 million devices will be sold in 2020 [78]. This section analyzes the characteristics of the latest and most popular hardware devices able to run shipbuilding AR/MR applications, which are compared in Table 26.1. The analysis of such a table allows for drawing the following conclusions:

- Device prices have decreased in the last years, reaching a point where they can compete with ruggedized tablets and smartphones. As it can be observed in the table, the prices of the selected devices range between $ 499 and $ 3500.
- The differences in terms of features among the compared smart AR/MR glasses are relevant and are usually dependent on their price. For example, the cheapest smart glasses of the comparison (the EPSON Moverio BT-30C) require no batteries and have been devised for visualizing content from external devices (Windows 10 PCs and Android 8.0 or later devices), which are connected to the glasses through a USB type-C connector. In contrast, the current most expensive glasses (Microsoft HoloLens 2) embed relatively powerful processors, which enable them to provide their own operating system and fast recognition and visualization features.

26

Table 26.1 Main features of potential AR/MR devices for shipbuilding applications (first part)

Product	Weight	Price	Hardware	Sensors	Connectivity	Battery	Display	Input/Output	Accessories	Software
ATHEER ThirdEye Gen's X2 MR Glasses [67]	~170 g	$ 1950	Integrated CPU/GPU, 4 GB of RAM, 64 GB of storage	9 Degrees of Freedom (DoF) Inertial Measurement Unit (IMU) with accelerometer, gyroscope and magnetometer, ambient light sensor, thermal sensor	WiFi, Bluetooth, GPS	1750 mAh battery	Stereoscopic dual 720p see-through displays at 60 fps with a field of view of 42°	13 MP HD camera, noise-cancelling microphone, 2 additional wide angle cameras, diverse control mechanisms (i.e., gaze, voice and gesture control, wireless controller)	Attachable corrective lens (to avoid wearing eye glasses), it can be attached to a hard hat for ANSI Z87.1 certification, optional sensor can be attached via the USB-C port	Based on Android 8.1, VisionEye SDK, support for Unity
EPSON Moverio BT-200 [68]	~88 g Headset and ~124 g Controller	~€ 700	TI OMAP 4460 1.2 Ghz Dual Core, 1 GB of RAM, 8 GB of internal memory, MicroSD (max. 2 GB)/MicroSDHC (max. 32 GB)	GPS in the controller, 3-axis compass, gyroscope and accelerometer in both the headset and the controller	IEEE 802.11 b/g/n with WiFi Miracast, Bluetooth 3.0, USB 2.0 (On-The-Go)	2720 mAh Li-Po battery, Battery life: ~6 h.	Poly-silicon TFT active matrix, 0.42 inch wide panel (16:9), 518,400 dots [(960× 540) × 3] 60 Hz. Field of view: 23° (diagonal), Screen size (projected distance): 80 inch at 5 m - 320 inch at 20 m, 24-bit color (16.77 million colors)	Camera (VGA), microphone, capacitive multi-touch touch-pad, Sound Output with Dolby Digital Plus	3.5 mm headphone jack, microphone	Android 4.0.4
EPSON Moverio BT-300 [69]	~69 g (without cable)	€ 849	Intel Atom x5 (1.44 GHz, quad-core), 16-GB internal storage, MicroSDHC (up to 32 GB).	GPS, magnetometer, accelerometer, gyroscope and ambient light sensor.	WiFi (IEEE 802.11a/b/g/n/ Bluetooth 4.1, micro-USB	Battery life: 6 h	Optical see-through, 24-bit HD color display. Field of view: 23° (diagonal)	Camera: 5 MP, 720 p. 4-pin mini-jack	Trackpad	Android 5.1
EPSON Moverio BT-350 [70]	~119 g (without shade and cables)	$ 1020	Intel Atom x5 (1.44 GHz, quad-core), 16-GB internal storage, MicroSDHC (up to 32 GB)	9-DoF IMU (magnetometer, accelerometer and gyroscope), GPS, ambient light sensor	WiFi (IEEE 802.11a/b/g/n/ Bluetooth 4.1, micro-USB	2950 mAh Li-Po battery, Battery life: ~6 h	Optical see-through, 24-bit HD color display. Field of view: 23° (diagonal)	5 MP camera, 4-pin mini-jack	n/a	Android, Moverio SDK

Table 26.1 Main features of potential AR/MR devices for shipbuilding applications (second part)

Product	Weight	Price	Hardware	Sensors	Connectivity	Battery	Display	Input/Output	Accessories	Software
EPSON Moverio BT-30C [71]	~95 g (without controller, shade and harness)	$ 499	n/a	9-DoF IMU with magnetometer, accelerometer and gyroscope	n/a	No batteries	Optical see-through, 24-bit HD color display. Field of view: 23° (diagonal)	4-pin mini-jack, USB type-C, no camera	n/a	Compatible with Windows 10 PCs and Android 8.0 or later devices with USB type-C connector
EPSON Moverio BT-35E [72]	~119 g (without shade and harness)	$ 899	n/a	9-DoF IMU (magnetometer, accelerometer and gyroscope), GPS, ambient light sensor	n/a	No batteries	Optical see-through, 24-bit HD color display. Field of view: 23° (diagonal)	5 MP camera, HDMI 1.4, 4-pin mini-jack, USB type-C	n/a	Compatible with Windows 10 and Android 7.0 or later devices with USB type-C connector
EPSON Moverio Pro BT-2000 [73]	~290 g (without cables)	~$ 1999	TI OMAP 4460 (1.2 GHz, dual-core), 1 GB of RAM, 8 GB of internal memory, MicroSD (max. 2GB)/MicroSD (max. 32GB)	GPS, magnetometer, gyroscope and ambient light sensor	WiFi (IEEE 802.11 b/g/n), Bluetooth 3.0 and BLE, USB 2.0 (On-The-Go), 4-pin mini jack	1240 mAh (×2) Lithium-ion batteries, Battery life: ~4 h	Optical see-through, 24-bit color display, Field of view: 23°	5 MP camera (image and video recording), 0.3 million pixels (depth sensing)	Earphones with microphone	Android 4.0.4, 3D support
Microsoft HoloLens [41]	579 g	~$3000	Custom Microsoft Holographic Processing Unit HPU 1.0, Intel 32-bit architecture, 2 GB of RAM, 64 GB of storage	IMU, ambient light sensor	WiFi (IEEE 802.11ac), Bluetooth 4.1 LE, Micro-USB 2.0	Battery life: 2–3 h with active use, 2 weeks in standby	See-through holographic lenses (waveguides), 2x HD 16:9 light engines, automatic pupillary distance calibration, 2.3 M total light points holographic resolution, 2.5 k light points per radian, Field of view: 34°	2 MP photos, HD video, external speakers, 3.5 mm audio jack, four environment understanding cameras, mixed reality capture, four microphones	n/a	Windows 10 with Windows Store. Human Understanding (spatial sound, gaze tracking, gesture input, voice support)
Microsoft HoloLens 2 [41]	566 g	~$3500	Second-generation custom-built holographic processing unit, Qualcomm Snapdragon 850 compute platform, 4 GB of RAM, 64 GB of storage	IMU (accelerometer, gyroscope, magnetometer), 4 visible light cameras (for head tracking), 2 infrared cameras (for eye tracking), 1 MP time-of-flight depth sensor	WiFi (IEEE 802.11ac 2 × 2), Bluetooth 5, USB type-C	Battery life: 2–3 h with active use	See-through holographic lenses (waveguides), 2k 3:2 light engines, >2.5k radiants (light points per radian), Field of view: 52°	8 MP photos, 1080p30 video, built-in spatial sound, 5-channel microphone array	n/a	Windows 10 with Windows Store, human tracking: hand, eye, and voice tracking (with iris recognition-based security)

(continued)

Table 26.1 (continued)

Product	Weight	Price	Hardware	Sensors	Connectivity	Battery	Display	Input/Output	Accessories	Software
Optinvent ORA-2 [74]	90 g	€ 699	OMAP 4460 (1.2 GHz, dual-core, 32-bit), 4 GB of flash, MicroSDHC (up to 32 GB), 1 GB of RAM	GPS, 3-axis accelerometer, gyroscope, magnetometer, ambient light sensor	WiFi (IEEE 802.11b/g/n), Bluetooth, micro-USB	Battery life: 3 h, external battery pack	Optical see-through, 23-bit full color display, 16:9 aspect ratio. Field of view: 23° (diagonal)	Camera: 5 MP pictures, 720p30 video	Touchpad, ear speaker, microphone	Android 4.4.2
Penny C Wear 30 Extended [75]	65 g (tethered computer box), 115 g (battery weight)	Unknow	Intel Core m5-6Y30, 4 GB of RAM, 256 GB of storage	3-axis gyroscope	WiFi, Bluetooth, micro-USB	Unknown characteris-tics, allows for using external power banks	Retinal see-through, OLED 854×480 resolution. Field of view: 42° × 25°	Audio and camera not integrated	Jawbone click sensor	Windows and Linux are supported
VUZIX Blade [76]	~90 g	$ 999	ARM Cortex-A53, 1 GB of RAM, 8 GB of flash, MicroSD	3-axis accelerometer, gyroscope and magnetometer, proximity sensor (for head detection), ambient light sensor, pressure sensor	Micro USB (for con-trol/power/upg WiFi (IEEE 802.11 b/g/n), Bluetooth 4.1	470 mAh lithium battery	480 × 853 1:1 display, Field of view: 19°, Brightness: >1000 nits, 24-bit color	8 MP camera, 720p30 or 1080p24 video, dual noise-canceling microphone.	It only supports right-eye mounting	Android 5.1.1
VUZIX M400 [77]	68 g (smart viewer), 62 g (battery), 60 g (USB-C cable and frame over glasses)	$ 1799	Octa-core 2.52 GHz Qualcomm XR1, 6 GB of RAM, 64 GB of flash memory	3-axis accelerometer, gyroscope and magnetometer, GPS/GLONASS, touchpad with multi-finger support	WiFi (IEEE 802.11 b/g/n/ac Bluetooth 5	1000 mAh Li-Po battery	Occluded OLED 640 × 360 16:9 display, Field of view: 16.8°, Brightness: 2000 nits, 24-bit color	12.8 MP camera, 4K30 video, triple noise-cancelling microphones, 3 control buttons	Mounting options: hard hat, headband, or glasses; Eyeglass frames with or without lens; Use with left or right eye; Available safety glasses	Android 8.1

- The battery life of the compared devices ranges between 2 and 6 h of continuous use.
- The majority of the compared devices embed an IMU with an accelerometer, a gyroscope, and a magnetometer.
- It is also very common to embed an ambient light sensor and, in the case of devices designed to be used outdoors, a GPS.
- Most of the latest devices embed a significant amount of RAM (4 or 6 GB), while the oldest ones only 1 GB. Regarding static storage, the compared devices can manage up to 256 GB of flash memory, which can be expanded in some cases through MicroSD cards.
- All devices except for EPSON Moverio BT-30C/BT-35E embed WiFi and Bluetooth transceivers. Specifically, the latest smart glasses support IEEE 802.11 a/b/g/n/ac and Bluetooth 5. Moreover, most models include some sort of USB connector to communicate with the glasses.
- Most devices embed optical see-through displays with a field of view that has increased in the last years and that currently ranges between 16.8° and 52°. The resolution of the displays varies from one model to another, but most of them provide 24-bit color and good brightness (up to 2000 nits).
- The resolution and quality of the embedded cameras differ significantly among the compared smart glasses:
  - They provide between 2 MP and 13 MP.
  - Video resolution varies from only VGA to 4K. The latest cameras can record 4K video at 30 frames per second.
  - Some glasses include different cameras and additional features. For instance, EPSON Moverio BT-2000 smart glasses embed a deep-focus camera, while Microsoft HoloLens 2 include, for head and eye tracking, four visible light cameras and two infrared cameras. Moreover, certain smart glasses embed noise-canceling microphone arrays or speakers.

It is worth mentioning that the authors of this chapter have tested extensively two of the compared models in a shipyard: EPSON Moverio BT-2000 and Microsoft HoloLens. Different instants of the tests are shown in Figs. 26.11, 26.12, 26.13, and 26.14. Specifically, Figs. 26.11 and 26.12 show two moments during marker reading tests carried out inside a block of a ship under construction (the results of such tests are documented in [35]). Regarding Figs. 26.13 and 26.14, they show two instants of the tests performed with the Microsoft HoloLens in two areas of a pipe workshop.

Finally, it must be indicated that, despite last years' remarkable AR/MR evolution, many of the currently available AR/MR devices should be considered as experimental developments. Therefore, developers have to be cautious when

Fig. 26.11 EPSON Moverio BT-2000 tests inside a ship block

Fig. 26.12 EPSON Moverio BT-2000 marker reading test in a ship under construction

Fig. 26.13 HoloLens test in a pipe workshop

Fig. 26.14 HoloLens test for bending machine maintenance

selecting a hardware platform, since many companies have not survived. For instance, in 2019, promising companies like Daqri [79] and ODG [80] went out of business.

26.5.3 Software AR/MR Frameworks for Shipbuilding

There are currently available a wide array of AR/MR SDKs (Software Development Kits), libraries, and frameworks that could be used to develop applications for the shipbuilding industry. Such software components differ in their implementation performance and in certain features. The following are the main features that should be provided by such software in order to implement successful AR/MR shipbuilding applications:

- It should provide 2D/3D graphic libraries to allow for visualizing and rendering multimedia content in real time.
- The implemented image/marker detection techniques should also be able to track objects and to overlap content on them (with or without making use of AR/MR markers).

Table 26.2 SDKs for developing AR/MR shipbuilding applications (first part)

Framework	Availability	Platform	AR/MR Detection and Tracking Features	Relevant Characteristics
ALVAR [81]	Free, Commercial SDK	Android, iOS, Windows Phone, Windows, Linux, OSX	Marker, Natural Feature Tracking (NFT)	AR/MR software library developed by the VTT Technical Research Center of Finland
ARKit 3 [82]	Commercial SDK	iOS, OSX	People occlusion detection, real-time people motion detection, multiple face tracking, robust 3D object detection, machine learning for detecting the planes in the environment	AR/MR framework developed by Apple. It allows for using simultaneously the front and back cameras, it enables creating collaborative experiences where multiple users share the same map or tracking up to three faces at once using the TrueDepth camera on iPhone X, iPhone XS, iPhone XS Max, iPhone XR, or iPad Pro
ARCore [83]	Free, Commercial SDK	Android, iOS, Unity, Unreal	Motion tracking, environmental understanding	ARCore is developed by Google and is currently supported for a limited group of mobile devices [84]
ARmedia [85]	Free, Commercial SDK	Android, iOS	Marker, NFT, GPS, IMU	SDK aimed at developing applications to track objects, images, locations, and device movements in real time to build AR/MR applications
ARToolkitX [86]	Open Source, Commercial SDK	Android, iOS, Linux, OSX, Windows	Marker, NFT	Open-source AR/MR framework that is a fork of the original ARToolKit, which was acquired by Daqri [79] in 2015
ArUco [87]	Open Source	Linux, OSX, Windows, Android	Marker	AR/MR library based on OpenCV
Augumenta Interaction Platform [88]	Commercial SDK	Android, Linux, Windows	NFT	SDK that provides gesture control and virtual surfaces. It has been already tested on smart glasses like ODG R-7, EPSON Moverio, Google Glass, or Microsoft HoloLens
Beyond Reality Face v5 [89]	Commercial SDK	HTML5 browser with JavaScript (it works in Chrome, Firefox, Edge 16, Opera, and Safari 11)	Face tracking	Lightweight multi-platform SDK for real-time face detection and face tracking

Table 26.2 Most relevant SDKs for developing IAR applications (second part)

Framework	Availability	Platform	AR/MR Detection and Tracking Features	Relevant characteristics
Catchoom [90]	Free, Commercial SDK	Android, iOS	Visual Search	Marker-based AR framework aimed at developing mobile applications
EasyAR [91]	Free, Commercial SDK	Android, iOS, OSX, Windows	Marker, NFT, SLAM	SDK for detecting and tracking surfaces, images, and 3D objects
HP Reveal (formerly Aurasma) [92]	Free, Commercial SDK	Android, iOS	NFT, Visual Search	It makes use of a proprietary image recognition solution and other patented technologies
Instant Reality [93]	Free, Commercial SDK	Android, iOS, Linux, OSX, Windows	Markers, NFT, GPS, IMU sensors, Facial Tracking, Visual Search, ContentAPI, SLAM, TrackerInterface	Developed by Fraunhofer IGD and ZGDV in cooperation with other industrial partners. This framework for mixed reality systems presents interfaces for developers to access components for AR/VR applications
MAXST [94]	Free, Commercial SDK	Android, Windows, iOS, OSX	Markers, NFT, SLAM	Cross-platform SDK for developing AR/MR applications
OpenSpace3D [95]	Open Source	Multi-platform (Android, iOS, Linux, OSX, Windows)	Markers (ArUco fiducial marker detection)	Framework that provides supports for Google Cardboard, HTC Vive, Oculus, or Leap motion
Onirix [96]	Free, Commercial SDK	Android, iOS	Markers, NFT,	Framework for detecting and tracking scenes, surfaces, and markers
Pikkart [97]	Free, Commercial SDK	Android, iOS	NFT,	SDK for detecting and tracking images. It also support geolocated augmented markers
Reality Kit [98]	Free	iOS, OSX	The same as ARKit	High-level framework based on Apple's ARKit to accelerate the creation of AR experiences
SSTT (Simplified Spatial Target Tracker) [99]	Proprietary	Android, iOS, Linux, OSX, Windows, Windows Mobile, Blackberry	Markers, NFT	Tracking library for developing AR/MR applications that is built on OpenCV
Vuforia [100]	Free, Commercial SDK	Android, iOS	Markers, NFT, Visual Search	AR/MR framework currently owned by PTC that is able to detect and track planar images and 3D objects in real time
Wikitude [101]	Free, Commercial SDK	Android, iOS, BlackBerry OS	GPS, IMU sensors, ContentAPI	AR SDK able to detect and track images. It is supported by EPSON Moverio, Microsoft HoloLens, and VUZIX smart glasses
ZapWorks Studio/Designer [102]	Commercial SDK	Android, iOS	Markers	Set of tools for creating AR experiences

26

- Voice-based interaction should be available for applications where shipyard operators cannot make use of a physical interface. In scenarios where noise may interfere the voice recognition system, it would be useful other hands-free operation techniques, like the ones based on eye tracking and on gesture detection.

The previous features have to be provided by an AR/MR SDK, library, or framework, which act as a middleman between the user/developer and the underlying hardware. The latest and most popular AR/MR development software that can be used for creating shipbuilding applications is summarized in Table 26.2. The first column of the table indicates the way a developer can access the software: most of the solutions provide a free version and an extended paid version, but there are a few that are open source (i.e., ARToolKitX, ArUco, OpenSpace3D). The second column enumerates the supported platforms, where it can be observed that, except for the Apple-specific SDKs, the rest is multi-platform and mainly aimed at supporting mobile devices. Finally, the third column refers to the techniques to detect and track objects, while the fourth column indicates the most relevant characteristics of the analyzed AR/MR software.

Among the compared software, the authors of this chapter evaluated thoroughly two of them: Vuforia and ARToolKit [35]. As an example, Fig. 26.15 shows a screenshot of a Vuforia application for assembling a clutch. Examples of ARToolKit applications are shown in Figs. 26.16 and 26.17: the former is aimed at identifying shipyard pallets, while the latter is for identifying pipes. Moreover, the authors have contributed to the development of the Microsoft HoloLens shipbuilding applications previously shown in Figs. 26.2, 26.5, 26.6, 26.7, and 26.8, as well as to the testing of third-party software developed with ARCore (some of such tests are illustrated in Figs. 26.18 and 26.19, where two Android applications called AR Industry and AR Machine were evaluated).

It is important to note that the development software should be compatible with the selected hardware platform.

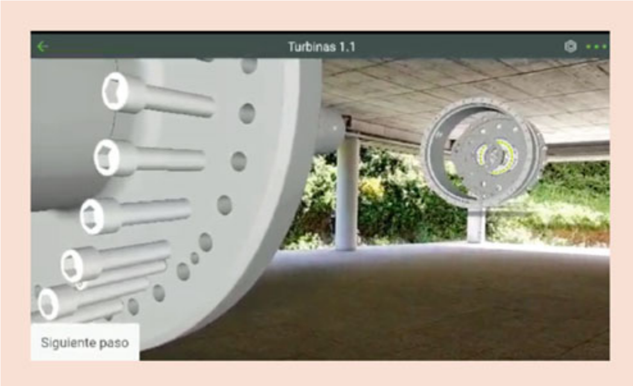


Fig. 26.15 Vuforia application developed by the JRU Navantia-UDC for clutch assembly

Fig. 26.16 ARToolKit application developed by the JRU Navantia-UDC for pallet identification

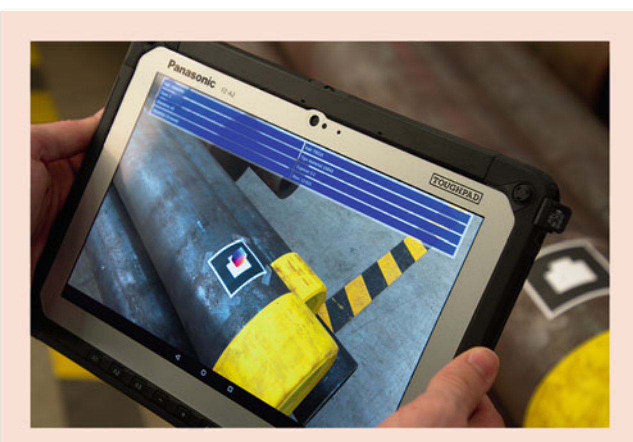


Fig. 26.17 ARToolKit application developed by the JRU Navantia-UDC for pipe identification

For instance, ARKit and ARCore are supported by a reduced set of mobile platforms. Other platforms like Augumenta have been used to develop applications for smart glasses like ODG R-7, EPSON Moverio, or Microsoft HoloLens.

Fig. 26.18 ARCore test of the application AR Industry

Fig. 26.19 ARCore test of the application AR Machine

Moreover, it is worth pointing out that the selected software should support Unity or Unreal, which are two of the most popular game engines but which have also been used for developing AR/MR applications. For instance, ARCore has native support both for Unity and Unreal.

It is also essential to select the appropriate computer graphics software. Such a software has to provide the necessary tools for creating attractive visual content, and it has to be able to export such a content with an appropriate file format in order to be further processed with engines like Unity or Unreal. From the experience of the authors of this chapter, the open-source graphics software Blender [103] is usually a good choice (as an example, Fig. 26.20 shows a screenshot of Blender during the design of a shipyard before exporting it to Unity).

Finally, it should be mentioned that the AR/MR software scene changes continuously and in the last years many software companies have stopped maintaining certain AR/MR applications. Therefore, although Table 26.2 reflects the features of the software at the time of writing (December 2019), it is possible that such features will change over time due to

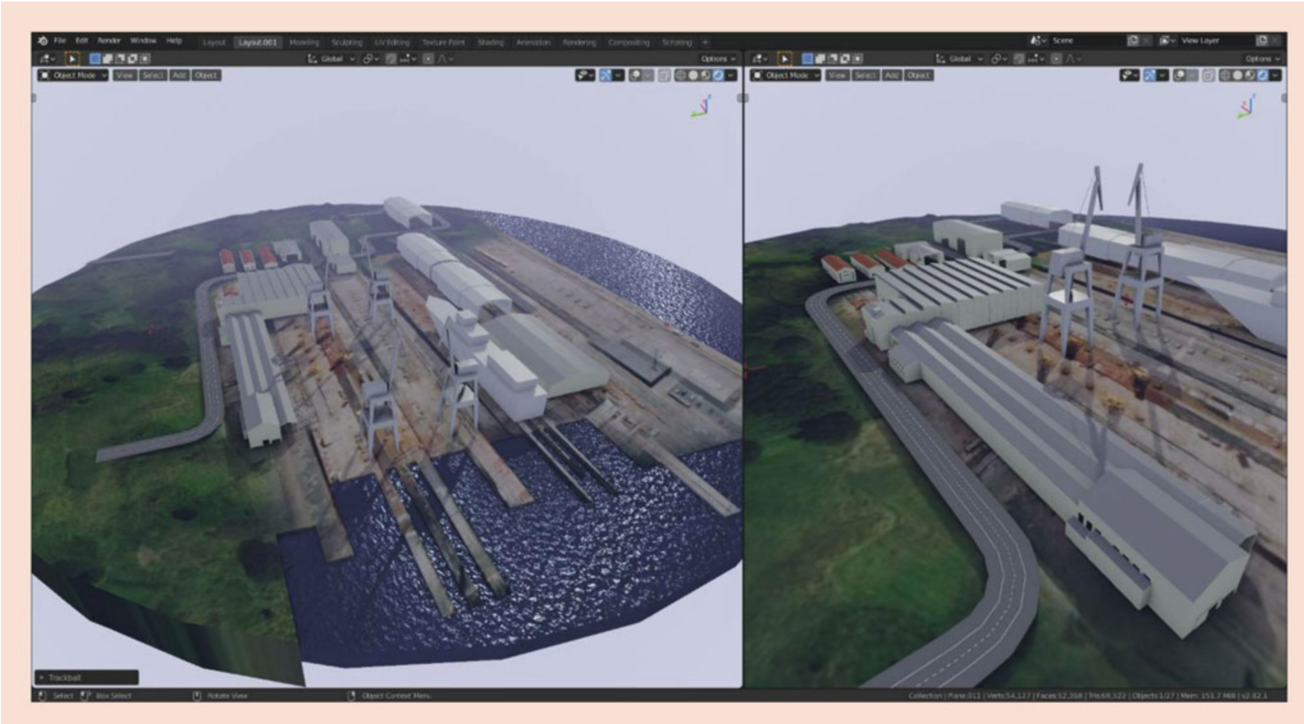


Fig. 26.20 Blender design of a shipyard

commercial strategies (e.g., the type of license may change) or to company mergers, acquisitions, or even bankruptcy.

26.5.4 Lessons Learned from the Shipbuilding State of the Art

This subsection provides additional insights of the state of the art presented in Sect. 26.2 by reviewing such works in terms of used hardware devices, software implemented, the proposed architecture, and the main advantages and limitations that such solutions provide to the shipyard tasks. The main aspects of such works are summarized in Table 26.3.

26.6 Main Challenges for the Development of AR/MR Shipbuilding Applications

Although AR/MR systems have evolved significantly in the last years, they still struggle with different generic industrial requirements and with specific necessities of the shipbuilding industry. The following are the most relevant:

- User experience has been improved noticeably in some of the latest AR/MR HMD systems like Microsoft HoloLens, but it needs to be further enhanced from the technical point of view. For instance, the field of view needs to be increased in order to cover as much human field view as possible so as to provide a safe and immersive experience in industrial environments.
- AR/MR hardware (both HMD and HHD devices) needs to be polished to be used in shipbuilding environments. For example, AR/MR devices have to be ruggedized to withstand the use in industrial scenarios, which can be really tough in the shipbuilding industry, requiring to resist the exposition to water, acids, or low/high temperatures. Moreover, HMD and HHD devices need to be optimized in terms of comfort, since one of the common operators' complaints is the excess of weight of certain devices.
- AR/MR devices also need to be adapted to meet the legal safety requirements that exist in shipyards and ships. For example, HMD devices need to be adapted to be worn with helmets or welding screens, which are compulsory in certain areas of the shipyard and during the construction of a ship.
- Battery life has also to be improved in many AR/MR devices so that they can work continuously during the whole operator working shift (roughly 8 h).
- The infrastructure that supports AR/MR devices and applications has to be adapted to the specific characteristics of shipbuilding scenarios. In the case of shipyards, it is necessary to provide fast wireless communications coverage to large areas. A ship under construction supposes a major challenge, since in practice it acts similarly to a Faraday cage that blocks wireless communications

Table 26.3 Comparison of state of the art shipbuilding applications

Reference (year)	Task	Hardware	Software	Comm. Architecture	Advantages	Limitations
Matsuo et al. [20]	Sheet Metal Forming	Markers	ARToolKit, FARO Laser scanner Focus3D	–	Interesting remarks about technical problems in the shipyard	Limited demo, article focused on reviewing the state of the art
Aiteanu et al. [21]	Shipbuilding Welding	See-through helmet. The visual system contains two stereo high dynamic range CMOS (HDRC) cameras and a HMD connected to a portable computer system	–	The portable computer system includes three standard PCs that communicates with a middleware CORBA via a gigabit Ethernet interface	Images at low frame rate can be saved together with welding parameters. Substitution of expensive X-ray quality inspection	Software details are not described; the communications architecture is limited.
Andersen et al. [22]	Shipbuilding Stud Welding	Human-robot interaction with projection mapping and IMU device (Wii remote)	Native OpenGL	–	The system is designed to instruct a robot to perform high precision and rapid welding tasks. Usability tests were satisfactory	Users requested additional information, e.g., next step in the process as well as current state of the system
Fast et al. [23]	Shipbuilding Gas Metal Arc Welding (Simulation)	Welding torch attached to a feedback device, a HMD, a 6-DOF tracking system, and external audio speakers	ENDEAVR framework (neural network software) to control the simulation	–	The process is recorded for review. The tool was used by over 100 people with different experience during 3 weeks	The system was graded as good (scale: poor, fair, good, great, excellent). Deficiencies were reported with respect to the brightness of the display and the fidelity of the audio feedback.
Lee et al. [24]	Shipbuilding Spray Painting (Simulation)	Two display configurations: a projection-based wall-type display and an HMD, a motion tracking system (six infrared cameras, six DOF), on/off buttons in the gun trigger, and air compressor to provide repulsive force feedback	Heuristic approach, OpenSG, and OpenCV libraries	Four PCs (one as tracking server, two for visualizing four display channels, and one for master control) connected through a local Gigabit Ethernet switch	Trainees can practice repeatedly for longer time; preparation for practice is optimized (no need for masks, environmental friendly)	Shortcomings reported regarding tracking failures, occlusion problems, and repulsive force feedback. Further work: virtual motion guides and a layered multiple display technique to improve depth perception.
Flatt et al. [25]	Shared Information (sticky notes)	Tablets (Samsung Galaxy Tab S, Samsung Nexus 10)	Ontology-based context-aware framework	Modular system, WiFi network	Object recognition is stable; the system is scalable without affecting the stability of the tracking process	Further work: a shared history log related to its components location
Havard et al. [26]	Step-by-step guidance, maintenance	Asus transformer pad tablet, Tronsmart plug (Miracast)	CAD model, UML entity model	–	Simple solution	No implementation details, only available for tablet. Epson Moverio BT-200 smart glasses (pending work)

Henderson et al. [27]	Step-by-step guidance, maintenance, and repair	Custom-built see-through head-worn display, NaturalPoint OptiTrack Tracking system, wrist-worn controllers	Valve Source Engine SDK, game engine models	Gigabit Ethernet connection and 802.11g link	Locate tasks in a maintenance sequence quicker	Bulky low-resolution hardware prototype, problems with occlusions.
Olbrich et al. [28]	Design and Construction Assistance, pipe layout planning	Tablet PC with a consumer webcam as a video device	X3D standard, an optical measurement tool	–	Pipe geometry can be inspected, modified, and stored	Further work: a precise evaluation of the tracking accuracy
Oh et al. [29]	Design and Construction Assistance, 3D model visualization	Android smartphone, Server (Intel i7 3.4 GHz)	Catia modeling tool	–	Recognition time of 3D model visualization is reduced compared with existing methods, improvements in the recognition rate	Initial development. Further work: 3D model to display it efficiently.

not only from the outside of the ship but also from the inside the ship, when it is necessary to communicate nearby rooms. Therefore, more research is required on new interfaces and architectures that provide broadband communications in scenarios that are tough in terms of electromagnetic propagation.

- Marker-based AR/MR solutions need to be optimized to work in low-light environments like the ones found in certain shipyard workshops and in ships under construction. In addition, such kind of systems needs to be enhanced to increase their recognition and tracking distances, since they may be under 1 meter for certain markers and scenarios [35].
- Since AR/MR systems are probably going to be essential for carrying out everyday shipbuilding tasks, it is necessary to increase the research of AR/MR cybersecurity in order to prevent potential cyber-attacks on both AR/MR devices and the infrastructure that supports them.

26.7 Conclusions

This chapter reviewed the latest developments of the application of AR and MR to the shipbuilding industry. After analyzing the most relevant and currently available AR/MR applications, the main potential shipbuilding tasks to be enhanced by AR/MR were studied. Then, the chapter detailed the internal components of the communications architectures that support many of the current AR/MR shipbuilding applications and described how researchers are enhancing them in order to provide fast responses and advanced services in the shipyard and in ships under construction. Finally, the chapter reviewed the most promising hardware and software and enumerated the main challenges faced by current AR/MR-based shipbuilding applications, thus providing clear guidelines for future researchers and developers.

Acknowledgments This work was supported by the Plant Information and Augmented Reality research line of the Joint Research Unit (JRU) Navantia-UDC. The authors would also like to acknowledge the support of Miguel A. Vilar-Montesinos and O. Blanco-Novoa and the support received from the Centro de Investigación de Galicia "CITIC," funded by Xunta de Galicia and the European Union (European Regional Development Fund-Galicia 2014–2020 Program), by grant ED431G 2019/01.

References

1. Xu, H., Yu, W., Griffith, D., Golmie, N.: A survey on industrial internet of things: a cyber-physical systems perspective. IEEE Access **6**, 78238–78259 (2018). https://doi.org/10.1109/ACCESS.2018.2884906
2. Aceto, G., Persico, V., Pescapé, A.: A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies, perspectives, and challenges. IEEE Commun. Surv. Tutorials **21**(4), 3467–3501 (2019). https://doi.org/10.1109/COMST.2019.2938259
3. Fernández-Caramés, T.M., Fraga-Lamas, P.: Towards post-quantum blockchain: a review on blockchain cryptography resistant to quantum computing attacks. IEEE Access (2020). https://doi.org/10.1109/ACCESS.2020.2968985
4. Fernández-Caramés, T.M.: From pre-quantum to post-quantum IoT security: a survey on quantum-resistant cryptosystems for the internet of things. IEEE Internet Things J. (2019). https://doi.org/10.1109/JIOT.2019.2958788
5. Fernández-Caramés, T.M., Fraga-Lamas, P.: A review on human-centered IoT-connected smart labels for the industry 4.0. IEEE Access **6**, 25939–25957 (2018). https://doi.org/10.1109/ACCESS.2018.2833501
6. Fraga-Lamas, P., Fernández-Caramés, T.M., Blanco-Novoa, O., Vilar-Montesinos, M.A.: A review on industrial augmented reality systems for the industry 4.0 shipyard. IEEE Access **6**, 13358–13375 (2018)
7. Goldman Sachs Goldman: Sachs Global Investment Research Technical Report: Virtual and Augmented Reality—Understanding the Race for the Next Computing Platform (2016)
8. Mizell, D.W.: Virtual reality and augmented reality in aircraft design and manufacturing. In: Proceedings of WESCON/94. IEEE, Anaheim (1994)
9. Wohlgemuth, W., Triebfürst, G.: Arvika: Augmented reality for development, production and service. In: Proceedings of DARE 2000, pp. 151–152. ACM, Elsinore (2000)
10. Friedrich, W.: Arvika-augmented reality for development, production and service. In: Proceedings of the International Symposium on Mixed and Augmented Reality, pp. 3–4. IEEE/ACM, Darmstadt (2002)
11. Fruend, J., Matysczok, C., Radkowski, R.: AR-based product design in automobile industry. In: Proceedings of the IEEE International Workshop Augmented Reality Toolkit, pp. 1–2. IEEE, Darmstadt (2002)
12. Lee, W., Park, J.: Augmented foam: a tangible augmented reality for product design. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 106–109. IEEE/ACM, Vienna (2005)
13. Novak-Marcincin, J., Barna, J., Janak, M., Novakova-Marcincinova, L., Torok, J.: Visualization of intelligent assembling process by augmented reality tools application. In: Proceedings of the IEEE International Symposium on Logistics and Industrial Informatics, pp. 33–36. IEEE, Smolenice (2012)
14. Park, H.S., Choi, H.W., Park, J.W.: Augmented reality based cockpit module assembly system. In: Proceedings of the International Conference on Smart Manufacturing Application, pp. 130–135. IEEE, Gyeonggi-do (2008)
15. Li, G., Xi, N., Chen, H., Saeed, A.: Augmented reality enhanced "top-down" nano-manufacturing. In: Proceedings of the IEEE Conference on Nanotechnology, pp. 352–354. IEEE, Munich (2004)
16. Kocisko, M., Teliskova, M., Baron, P., Zajac, J.: An integrated working environment using advanced augmented reality techniques. In: Proceedings of the International Conference on Industrial Engineering and Applications, pp. 279–283. IEEE, St. Petersburg (2017)
17. Molina Vargas, D.G., Vijayan, K.K., Mork, O.J.: Augmented reality for future research opportunities and challenges in the shipbuilding industry: a literature review. Proc. Manuf. **45**, 497 – 503 (2020). https://doi.org/10.1016/j.promfg.2020.04.063
18. Morais, D., Waldie, M., et al.: How to implement tech in shipbuilding: charting the course to success. In: SNAME Maritime

Convention. The Society of Naval Architects and Marine Engineers (2018)
19. Hjartholm, B.T.: Augmented reality for operator support in Norwegian shipyards: a study of applications, benefits and challenges (2019)
20. Matsuo, K., Rothenburg, U., Stark, R.: Application of ar technologies to sheet metal forming in shipbuilding. In: Abramovici, M., Stark, R. (eds.) Smart Product Engineering, pp. 937–945. Springer, Berlin (2013)
21. Aiteanu, D., Hillers, B., Graser, A.: A step forward in manual welding: demonstration of augmented reality helmet. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 309–310. IEEE/ACM, Tokyo (2003)
22. Andersen, R.S., Bøgh, S., Moeslund, T.B., Madsen, O.: Task space HRI for cooperative mobile robots in fit-out operations inside ship superstructures. In: Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication, pp. 880–887. IEEE, New York (2016)
23. Fast, K., Gifford, T., Yancey, R.: Virtual training for welding. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 298–299. IEEE/ACM, Arlington (2004)
24. Lee, G.A., Yang, U., Son, W., Kim, Y., Jo, D., Kim, K.H., Choi, J.S.: Virtual reality content-based training for spray painting tasks in the shipbuilding industry. ETRI J. **32**(5), 695–703 (2010)
25. Flatt, H., Koch, N., Röcker, C., Günter, A., Jasperneite, J.: A context-aware assistance system for maintenance applications in smart factories based on augmented reality and indoor localization. In: Proceedings of the IEEE Conference on Emerging Technologies Factory Automation, pp. 1–4. IEEE, Luxembourg (2015)
26. Havard, V., Baudry, D., Louis, A., Mazari, B.: Augmented reality maintenance demonstrator and associated modelling. In: Proceedings of IEEE Virtual Reality, pp. 329–330. IEEE, Arles (2015)
27. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE Trans. Vis. Comput. Graph. **17**(10), 1355–1368 (2011)
28. Olbrich, M., Wuest, H., Riess, P., Bockholt, U.: Augmented reality pipe layout planning in the shipbuilding industry. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, pp. 269–270. IEEE, Basel (2011)
29. Oh, Y.J., Park, K.Y., Kim, E.K.: Mobile augmented reality system for design drawing visualization. In: Proceedings of the International Conference on Advanced Communication Technology, pp. 1296–1300. IEEE, Pyeongchang (2014)
30. Newport News Shipbuilding: http://nns.huntingtoningalls.com/ar/
31. Index AR Solutions: https://www.indexarsolutions.com/about-us/
32. Virtual Reality Technology Transforms Design of UK Warships: http://www.baesystems.com/en/article/virtual-reality-technology-transforms-design-of-uk-warships
33. Navantia official web page: https://www.navantia.es
34. Fernández-Caramés, T.M., Fraga-Lamas, P., Suárez-Albela, M., Vilar-Montesinos, M.A.: A fog computing and cloudlet based augmented reality system for the industry 4.0 shipyard. Sensors **18**(6), 1798 (2018)
35. Blanco-Novoa, O., Fernández-Caramés, T.M., Fraga-Lamas, P., Vilar-Montesinos, M.A.: A practical evaluation of commercial industrial augmented reality systems in an industry 4.0 shipyard. IEEE Access **6**, 8201–8218 (2018)
36. PTC's Vuforia official web page: http://www.vuforia.com
37. Fraga-Lamas, P., Noceda-Davila, D., Fernández-Caramés, T.M., Diaz-Bouza, M.A., Vilar-Montesinos, M.A.: Smart pipe system for a shipyard 4.0. Sensors **16**(12), 2186 (2016)
38. Fraga-Lamas, P., Fernández-Caramés, T.M., Noceda-Davila, D., Vilar-Montesinos, M.A.: RSS stabilization techniques for a real-time passive UHF RFID pipe monitoring system for smart shipyards. In: Proceedings of the IEEE International Conference on RFID, pp. 161–166. IEEE, Phoenix (2017)
39. Fraga-Lamas, P., Fernández-Caramés, T.M., Noceda-Davila, D., Diaz-Bouza, M.A., Vilar-Montesinos, M.A., Pena-Agras, J.D., Castedo, L.: Enabling automatic event detection for the pipe workshop of the shipyard 4.0. In: Proceedings of FITCE, pp. 20–27. IEEE, Madrid (2017)
40. Blanco-Novoa, O., Fraga-Lamas, P., Vilar-Montesinos, M.A., Fernández-Caramés, T.: Towards the internet of augmented things: an open-source framework to interconnect IoT devices and augmented reality systems. In: Proceedings of the 6th International Electronic Conference on Sensors and Applications. MDPI (2019)
41. Microsoft HoloLens official web page: http://www.hololens.com
42. Schroeder, G., Steinmetz, C., Pereira, C.E.: Visualising the digital twin using web services and augmented reality. In: Proceedings of the IEEE International Conference on Industrial Informatics. IEEE, Poitiers (2016)
43. Smparounis, K., Mavrikios, D., Pappas, M., Xanthakis, V., Viganò, G.P., Pentenrieder, K.: A virtual and augmented reality approach to collaborative product design and demonstration. In: Proceedings of the IEEE International Technology Management Conference. IEEE, Lisbon (2008)
44. Matysczok, C., Wojdala, A.: Rendering of highly polygonal augmented reality applications on a scalable PC-cluster architecture. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE/ACM, Arlington (2004)
45. Hernández-Rojas, D., Fernández-Caramés, T.M., Fraga-Lamas, P., Escudero, C.J.: Design and practical evaluation of a family of lightweight protocols for heterogeneous sensing through BLE beacons in IoT telemetry applications. Sensors **18**, 57 (2017)
46. Hernández-Rojas, D.L., Fernández-Caramés, T.M., Fraga-Lamas, P., Escudero, C.J.: A plug-and-play human-centered virtual teds architecture for the web of things. Sensors **18**(7), 2052 (2018)
47. Blanco-Novoa, O., Fernández-Caramés, T.M., Fraga-Lamas, P., Castedo, L.: A cost-effective IoT system for monitoring indoor radon gas concentration. Sensors **18**(7), 2198 (2018)
48. Fernández-Caramés, T.M., Fraga-Lamas, P.: Towards the internet of smart clothing: a review on IoT wearables and garments for creating intelligent connected e-textiles. Electronics **7**(12), 405 (2018)
49. Fernández-Caramés, T.M., Fraga-Lamas, P., Suárez-Albela, M., Diaz-Bouza, M.A.: A fog computing based cyber-physical system for the automation of pipe-related tasks in the industry 4.0 shipyard. Sensors **18**(6), 1961 (2018)
50. Zhang, J., Ong, S.K., Nee, A.Y.C.: A volumetric model-based CNC simulation and monitoring system in augmented environments. In: Proceedings of the International Conference on Cyberworlds, pp. 33–42. IEEE, Lausanne (2006)
51. Rao, Q., Grünler, C., Hammori, M., Chakrabort, S.: Design methods for augmented reality in-vehicle infotainment systems. In: Proceedings of the ACM/EDAC/IEEE Design Automation Conference, pp. 1–6. IEEE, San Francisco (2014)
52. Langfinger, M., Schneider, M., Stricker, D., Schotten, H.D.: Addressing security challenges in industrial augmented reality systems. In: Proceedings of IEEE International Conference on Industrial Informatics. IEEE, Emdem (2017)
53. Schneider, M., Rambach, J., Stricker, D.: Augmented reality based on edge computing using the example of remote live support. In: Proceedings of the Annual International Conference on Industrial Technology. IEEE (2017)
54. Dolui, K., Datta, S.K.: Comparison of edge computing implementations: fog computing, cloudlet and mobile edge computing. In: Proceedings of the Global Internet of Things Summit, pp. 1–6. IEEE, Geneva (2017)

55. Fraga-Lamas, P., Fernández-Caramés, T.M.: A review on blockchain technologies for an advanced and cyber-resilient automotive industry. IEEE Access **2019**(7), 17578–17598 (2019)
56. Fernández-Caramés, T.M., Fraga-Lamas, P.: A review on the application of blockchain for the next generation of cybersecure industry 4.0 smart factories. IEEE Access **7**, 45201–45218 (2019)
57. Fernández-Caramés, T.M., Froiz-Miguez, I., Blanco-Novoa, O., Fraga-Lamas, P.: Enabling the internet of mobile crowdsourcing health things: a mobile fog computing, blockchain and IoT based continuous glucose monitoring system for diabetes mellitus research and care. Sensors **19**(15), 3319 (2019)
58. Fernández-Caramés, T.M., Blanco-Novoa, O., Froiz-Miguez, I., Fraga-Lamas, P.: Towards an autonomous industry 4.0 warehouse: a UAV and blockchain-based system for inventory and traceability applications in big data-driven supply chain management. Sensors **19**(10), 2394 (2019)
59. Suárez-Albela, M., Fernández-Caramés, T.M., Fraga-Lamas, P., Castedo, L.: A practical evaluation of a high-security energy-efficient gateway for IoT fog computing applications. Sensors **17**(9), 1978 (2017)
60. Suárez-Albela, M., Fraga-Lamas, P., Castedo, L., Fernández-Caramés, T.M.: Clock frequency impact on the performance of high-security cryptographic cipher suites for energy-efficient resource-constrained IoT devices. Sensors **19**(1), 15 (2019)
61. Fernández-Caramés, T.M., Fraga-Lamas, P., Suárez-Albela, M., Castedo, L.: A Methodology for Evaluating Security in Commercial RFID Systems, chap. Radio Frequency Identification, Prof. Paulo Crepaldi (Ed.). InTech (2017)
62. Fernández-Caramés, T.M., Fraga-Lamas, P., Suárez-Albela, M., Castedo, L.: Reverse engineering and security evaluation of commercial tags for RFID-based IoT applications. Sensors **17**(1), 28 (2016)
63. Caudell, T.P., Mizell, D.W.: Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: Proceedings of the International Conference on System Sciences, vol. ii, pp. 659–669. IEEE, Kauai (1992)
64. Syberfeldt, A., Danielsson, O., Gustavsson, P.: Augmented reality smart glasses in the smart factory: Product evaluation guidelines and review of available products. IEEE Access **5**, 9118–9130 (2017)
65. Thomas, B.H., Sandor, C.: What wearable augmented reality can do for you. IEEE Pervasive Comput. **8**(2), 8–11 (2009)
66. Fernandez, T.M., Rodas, J., Escudero, C.J., Iglesia, D.I.: Bluetooth sensor network positioning system with dynamic calibration. In: Proceedings of the International Symposium on Wireless Communication Systems. IEEE, Trondheim (2007)
67. ATHEER ThirdEye Mixed Reality Smart Glasses and AR software. ThirdEye official web page: https://www.thirdeyegen.com
68. EPSON Moverio BT-200 (Developer Version Only) official web page: https://epson.com/Clearance-Center/Wearables/Moverio-BT-200-Smart-Glasses-%28Developer-Version-Only%29/p/V11H560020
69. EPSON Moverio BT-300 (AR/Developer Edition) official web page: https://epson.com/For-Work/Wearables/Smart-Glasses/Moverio-BT-300-Smart-Glasses-%28AR-Developer-Edition%29-/p/V11H756020
70. EPSON Moverio BT-350 Technical Information: https://tech.moverio.epson.com/en/bt-350/
71. EPSON Moverio BT-30C Technical Information: https://tech.moverio.epson.com/en/bt-30c/
72. EPSON Moverio BT-35E Technical Information: https://tech.moverio.epson.com/en/bt-35e/
73. EPSON Moverio BT-2000 Technical Information: https://tech.moverio.epson.com/en/bt-2000/
74. Optinvent ORA-2 glasses official web page: http://www.optinvent.com/
75. Penny C Wear glasses official web page (currently only available through Internet Archive): https://bit.ly/3kgINA4
76. Vuzix Blade official web page. See-Through AR Smart Glasses Powered by Industry-Leading Waveguide Optics: https://www.vuzix.com/products/blade-smart-glasses
77. Vuzix M400 official web page: https://www.vuzix.com/products/m400-smart-glasses
78. Tractica. Smart Augmented Reality Glasses. Technical Report: https://www.tractica.com/research/smart-augmented-reality-glasses/
79. Daqri official web page (currently only available through Internet Archive): https://bit.ly/35F5UQN
80. ODG official web page (currently only available through Internet Archive): https://bit.ly/2GZVUHt
81. ALVAR Augmented Reality / 3D Tracking official web page: http://virtual.vtt.fi/virtual/proj2/multimedia/alvar/index.html
82. Apple AR Kit Augmented Reality Apple Developer official web page: https://developer.apple.com/augmented-reality/arkit/
83. Google ARCore official web page: https://developers.google.com/ar/
84. List of ARCore supported devices: https://developers.google.com/ar/discover/supported-devices
85. AR-media official web page: http://www.armedia.it/index.php
86. ARToolkitX official web page: http://www.artoolkitx.org
87. ArUco official web page. A minimal library for Augmented Reality applications based on OpenCV: https://www.uco.es/investiga/grupos/ava/node/26
88. Augumenta Studio official web page: http://augumenta.com
89. Beyond Reality Face official web page: https://www.beyond-reality-face.com
90. Image Recognition, AR and AI Solutions. Catchoom official web page: https://catchoom.com
91. EasyAR official web page: https://www.easyar.com
92. HP Reveal official web page: https://www.hpreveal.com
93. Instant Reality official web page: http://www.instantreality.org
94. MAXST official web page: http://maxst.com
95. OpenSpace3D - Open Source Platform for 3D environments, OpenSpace3D official web page: http://www.openspace3d.com
96. Onirix official web page: https://www.onirix.com
97. Pikkart official web page: https://developer.pikkart.com/augmented-reality/sdk/
98. Reality Composer/Kit official web page: https://developer.apple.com/augmented-reality/reality-composer/
99. SSTT—Augmented Reality Tracking Library. Technotecture official web page: http://technotecture.com/projects/sstt
100. PTC. Vuforia official web page: https://www.vuforia.com
101. Wikitude Augmented Reality. Wikitude AR SDK official web page: https://www.wikitude.com
102. ZapWorks official web page: https://zap.works
103. Blender official web page: https://www.blender.org

Prof. Dr. Tiago M. Fernández-Caramés works as an Associate Professor at the University of A Coruña (Spain). His current research interests include augmented/mixed reality, IoT/IIoT systems, embedded systems, blockchain, and Industry 4.0 technologies. In such fields, he has contributed to more than 100 publications for conferences, journals, and books.

Dr. Paula Fraga-Lamas works as a Senior Researcher and Lecturer at the University of A Coruña (Spain). She has over 100 contributions to journals, conferences, and book chapters. She participated in over 30 research projects. Her current research interests include Industry 4.0/5.0, IoT, cyber-physical systems, augmented/mixed reality, fog and edge computing, blockchain and Distributed Ledger Technology (DLT), and cybersecurity.

Augmented Reality for Remote Assistance (ARRA)

27

Riccardo Palmarini, Iñigo Fernández del Amo, Dedy Ariansyah, John Ahmet Erkoyuncu, and Rajkumar Roy

Contents

27.1 Introduction 669

27.2 Background 670

27.3 ARRA 671
27.3.1 ARRA: A Practical Example 672
27.3.2 Technical Development 674

27.4 Test Design and Methodology 675
27.4.1 Validation Case Study 675
27.4.2 Quantitative Validation Test Methodology 676

27.5 Analysis and Results 679

27.6 Discussion 680

27.7 Conclusion and Future Work 683

References 683

Abstract

Augmented Reality (AR) reduces the technicians' cognitive effort mainly resulting in both time and error rate reductions. Still, its application in remote assistance has not been fully explored yet. This paper focuses on understanding the benefits of providing assistance to a remote technician through AR. Augmented Reality for Remote Assistance (ARRA) has been designed and developed for local novice maintainer to request assistance and communicate with a remote expert. The remote expert can manipulate virtual objects, which are then overlaid on the real environment of the novice maintainer. ARRA has been tested with the help of 60 participants. This involved performing an assembly/disassembly operation on a mock-up of a piping system. The participants were remotely assisted through ARRA or video-call. Quantitative spatial referencing error data has been collected. The results showed a 30% improvement in terms of spatial referencing when utilizing ARRA as remote assistance support as opposed to video-call. Future studies should investigate into quantifying the improvements due to other factors involved in remote assistance, especially language barriers and connectivity issues.

Keywords

Augmented reality · Digital engineering · Maintenance · Remote assistance · Spatial referencing

R. Palmarini · I. F. del Amo · D. Ariansyah · J. A. Erkoyuncu (✉)
School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK
e-mail: I.Fernandez@cranfield.ac.uk; Dedy.Ariansyah@cranfield.ac.uk; j.a.erkoyuncu@cranfield.ac.uk

R. Roy
School of Mathematics, Computer Science and Engineering, City, University of London, London, UK
e-mail: r.roy@city.ac.uk

27.1 Introduction

The increasing complexity of industrial machinery due to the constant push for improvements in productivity and reliability of industrial facilities has provided a flourishing ground for research and innovation [1]. Internet of Things, Digital Engineering, Smart Factory, Virtual Reality, Digital Twins, and Augmented Reality (AR) are only a few of the words utilized today for describing approaches and technologies which could enhance and support the fourth industrial revolution and take us to the nowadays well-acknowledged Industry 4.0 [2]. In this study, we explore the utilization of AR for Remote Assistance (RA) applications in maintenance. Several definitions of AR are provided in the academy. The first and most widely recognized one has been provided by Azuma in 1997 [3] and restated in 2001 [4]: "AR supplements the real world with virtual (computer-generated) objects that appear to coexist in the same space as the real world"; moreover an "AR system has the following properties: combines real and virtual objects in a real environment, runs interactively and in real-time, registers real and virtual objects with each other's."

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_27

Maintenance, Repair, and Overhaul (MRO) operations have a big impact on the lifecycle of industrial equipment [5] and strongly rely on the maintenance technician's expertise [6]. In this scenario, AR technology for remote assistance (RA) can potentially allow the "de-skilling" of the remote maintenance operations and, at the same time, improve flexibility and costs of maintenance [7]. The flexibility in the maintenance scenario is the capability of performing MRO operations without specific skill-requirements, location constraints, and effective with unexpected events [8]. The cost would be directly affected by avoiding the need for time consuming and expensive maintenance training as well as traveling [9]. It is not uncommon that machinery vendors are required to provide assistance in remote locations because their technicians are better trained to perform MRO on the vendor's product (industrial machinery, tooling, instruments). Similar maintenance dynamics may occur also within different departments of the same company. To provide such benefits, the ARRA tool should overcome three main limitations of current RA technologies based on voice and video call support as follows [10]:

1. Spatial referencing – identifying the correct location and orientations of the object in space.
2. Communication barriers – language describing actions can be vague and ambiguous.
3. Connectivity issues – relying on 4G or Wi-Fi internet connection can affect RA.

This paper focuses on improving spatial referencing through the utilization of AR for RA. The authors developed an AR approach that puts in communication two technicians situated in different locations: the expert and the novice. The novice here refers to the maintainer who does not sufficiently know how to perform the maintenance task and requires support from the expert (e.g., from the vendor). The AR approach has been called ARRA: Augmented Reality for Remote Assistance. It is based on the assumption that the AR system can recognize and track the objects in the Field of View (FOV) of the novice and that the CAD models of the objects for MRO are available. ARRA allows in execution-time order: (1) the novice to request assistance, (2) the expert to visualize virtually on his real environment, the objects to be MRO, (3) the expert to manipulate the virtual object to build a step-by-step MRO procedure, (4) the novice to visualize the step-by-step MRO procedure, and (5) the expert to monitor the progress of the MRO procedure. This paper is structured as follows. Section 27.2 provides the research background and motivation. Section 27.3 describes ARRA: how it works and its technical development. The detailed methodology for ARRA's validation is described in Sect. 27.4. It includes the description of the case study utilized (Sect. 27.4.1) and the quantitative test design (Sect. 27.4.2). Analysis and results are reported in Sect. 27.5. Finally, the discussion of the results and the conclusions and future works are proposed in Sects. 27.6 and 27.7, respectively.

27.2 Background

AR for MRO applications has been widely explored by academics, and the benefits that AR technology could bring to the industrial environment are mainly time reductions, error reductions, cognitive load reduction, training reduction, and cost reduction [11–13]. AR applications specific for RA in maintenance, on the other side, have been investigated and proposed only by the 8% of the academic studies of AR in maintenance [11]. It is worth to mention that some studies, rather than talking about "remote assistance," utilize the words "tele-presence," "tele-assistance," or "tele-maintenance" to indicate the capability of providing support to remote operators through the utilization of AR or other technologies (VR, the Cloud, Computer) [14–16]. Reference [17] in 2014 proposed a client-server AR system which allows the remote expert to overlay symbols and written instructions over the real internal combustion engine where a remote novice maintainer is carrying out the maintenance operation. This application has been designed for increasing customer satisfaction, cut costs, and allow rapid intervention always considering low connectivity. Reference [18] in 2015 attempted to utilize Mobile Internet Devices (MIDs) such as smartphones to remotely acquire data on a machine (equipped with its electronics and monitoring sensors) and apply corrective actions if required. The corrective actions can be suggested by the remote manufacturer or maintainer by means of AR annotations and/or directly modifying the machine parameters. This method requires a gateway architecture that is not always available and applicable only on heavily electronics equipped machinery. Reference [18] in 2015 developed and compared three remote support systems: Sketch3D, Point3D, and Demo3D. The utilization of Demo3D resulted in the shortest completion time of the assembly task selected in the study. The system enabled the remote expert to manipulate a virtual object through the utilization of a Head-Mounted Display (HMD) and a tracked mouse. The final configuration of the virtual replica was then overlaid on the real environment of the novice remote maintainer who could take advantage of the invariant spatial referencing of AR and verify the proper alignment of the real objects. Nevertheless, it did not consider this solution applicable to complex maneuvers. In 2017, another example of RA through AR that connects the cloud-based system to the assembly plant was demonstrated [19]. The MTBF of the machines to be maintained was calculated through an automated analysis of the maintenance logs and sensors data. If there is a requirement for preventive maintenance, the technician on the plant can then request assistance. The maintenance department is able to build a maintenance report which includes AR scenes, animations, and instructions. These are generated through a "smart dis/assembly algorithm." More specifically, the animations were built through the analysis of the physical constraint on CATIA. For instance, once the object to be maintained has been identified, the CAD model was automatically analyzed on CATIA, and the components' DOF were evaluated. If a component can move in at least one direction without colliding with other components, it can be disassembled. This solution overcomes communications barriers and provides an interesting attempt to automate, and hence, solve one of the main issues of the implementation of AR in maintenance: the AR contents creation [20–22]. Still, we believe it is slightly too simplistic. The solution was not only unable to provide different solutions for the same problem, but also it did not consider unpredicted events and did not take advantage of the human experience which is essential in maintenance [6].

The AR solutions for RA described testifying the effort in pushing forward the utilization of AR. Still, it is not clear how much benefit could we expect from its implementation. For this reason, in this paper, the authors will attempt to quantify the expected spatial awareness benefits resulting in the utilization of AR for RA.

27.3 ARRA

Augmented Reality for Remote Assistance (ARRA) is our proposed approach for overcoming spatial referencing issues, which affect current RA technologies: video/voice all support, VR, and AR. As anticipated, ARRA is based on two assumptions:

1. The system can recognize and track the object in the FOV of the remote novice maintainer.
2. The CAD models of the objects to be maintained are available.

The authors consider the assumptions plausible due to the recent advancement in image processing, depth sensors, and CAD modelling [11, 23]. Figure 27.1 reports a schematic concept of how ARRA works, what is the data flow, and what are the main processes involved. On the left, the remote novice maintainer carries out a maintenance operation without having the required knowledge. The initial current status (components positions and orientations) of the object to be maintained is sent to the remote maintainer. The remote maintainer can visualize the objects and virtually manipulate them. He performs the maintenance operation on the virtual objects. This is sent back to the novice maintainer that visualizes it overlaid on the real objects. The novice can then follow the steps of the procedure while the expert is monitoring the movements of the objects (3*) since the system is continuously sending the objects' current status. At any stage of the assistance, both the expert and the novice can request to restart from process 1*. This may occur in two main occasions:

1. The novice is not able to follow the overlaid procedure (2*).

27

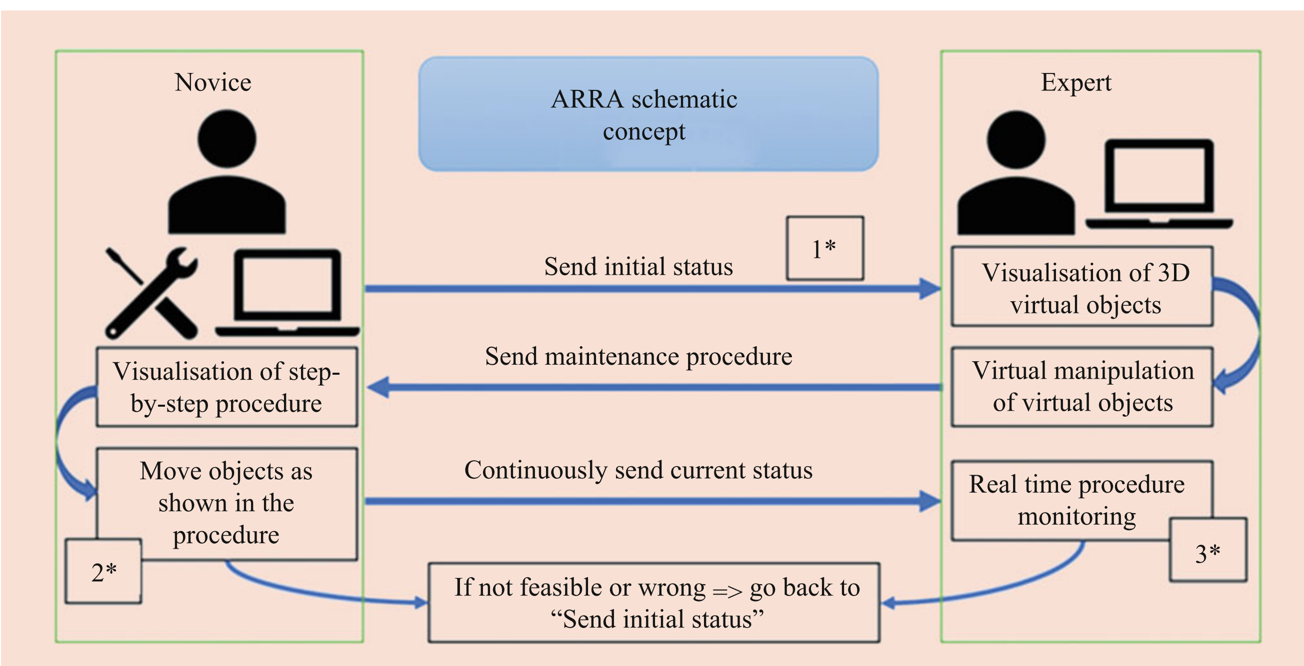


Fig. 27.1 ARRA schematic concept and functionalities

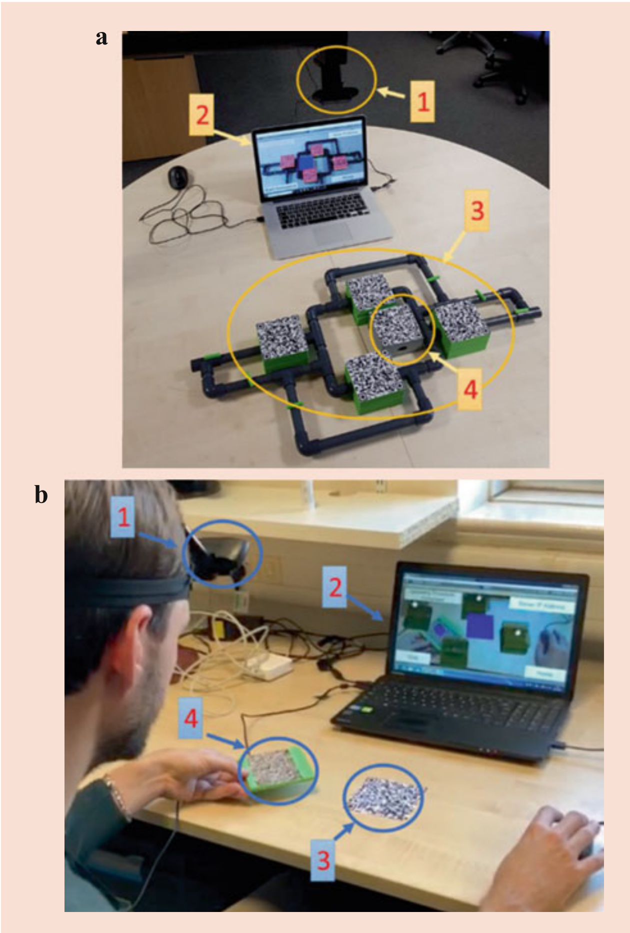


Fig. 27.2 Novice and expert environments when utilizing ARRA (**a**) Novice environment; (**b**) Expert environment

2. The expert has noticed something wrong in the movements of the object real time (3*).

The key to improving the spatial referencing with respect to video/voice supports lies in the AR technology and the utilization of the relative positions of the objects with respect to the anchor marker located in both the novice and the expert environments. In order to provide a better understanding of how ARRA works, the following sections will show a practical example in Sect. 27.3.1 and the technical development details in Sect. 27.3.2.

27.3.1 ARRA: A Practical Example

This section reports a practical example of how ARRA works. The pictures utilized for explaining ARRA have been taken during the validation tests and are utilized here to better explain how ARRA allows AR communication between the novice and the remote expert maintainer, what information is transferred, and how the novice maintainer becomes able to perform a maintenance operation through ARRA. The validation tests and the case study will be described in detail in Sect. 27.4.1. Two different environments are considered, for instance, the novice's shop-floor (Fig. 27.2a) and the expert's desk (Fig. 27.2b). The novice environment includes (1) an RGB camera facing the working area, (2) a laptop/display, (3) the object to be maintained, and (4) the anchor marker. The image is taken from the novice's point of view. Four markers have been placed on the object to be maintained for easing the four components recognition for testing purposes. The expert environment includes (1) an RGB camera facing the same direction as the expert, (2) a laptop, (3) the anchor marker, and (4) the virtual manipulator tool. The latter is a real object which, once recognized as the virtual manipulator through its marker, allows to move and rotate virtual objects. It is worth to mention that, in both environments, the RGB camera and laptop could potentially be substituted with an HMD. The description of the example will now progress following the actual operation time sequence.

Firstly, the novice approaches the object to be maintained and understood he is lacking the knowledge necessary to carry out the maintenance operation, and thus, requests assistance through the UI of the ARRA application on the display.

The remote expert accepts the request for assistance and visualizes the CAD models of the object to be maintained. More specifically, the four objects recognized by the novice's camera (Fig. 27.2a) and their position and orientation with respect to the novice's anchor marker are reproduced virtually on the experts' screen maintaining the same relative position with the maintainer's anchor marker (Fig. 27.3a). The expert understands what maintenance operation has to be carried out based on his/her expertise and places the virtual manipulator over the virtual component that has to be moved (Fig. 27.3b). Once he presses "Grab" (bottom left on Fig. 27.3b), the virtual component starts following the virtual manipulator movements. In Fig. 27.3c and Fig. 27.3d, respectively, it was shown that the expert rotates the virtual manipulator and the virtual component rotates as well. Also, on the top left of the expert's screen (Fig. 27.3d), it is possible to see the current action performed by ARRA: "Recording." Please note, ARRA is not recording the video information but only the object positions and orientations through time by storing them locally. Once the expert has moved the object as required by the maintenance operation, he can select "Release" (bottom right in Fig. 27.3d), and the information recorded was uploaded on a cloud server database. The remote expert can now keep monitoring the movements of the real novice's objects through the virtual components on his/her display.

On the top left of the novice's display, the statement "Playing Procedure x" is shown (Fig. 27.4b). All the objects that are positioned and orientated correctly will be overlaid with its own CAD model colored in green. The component

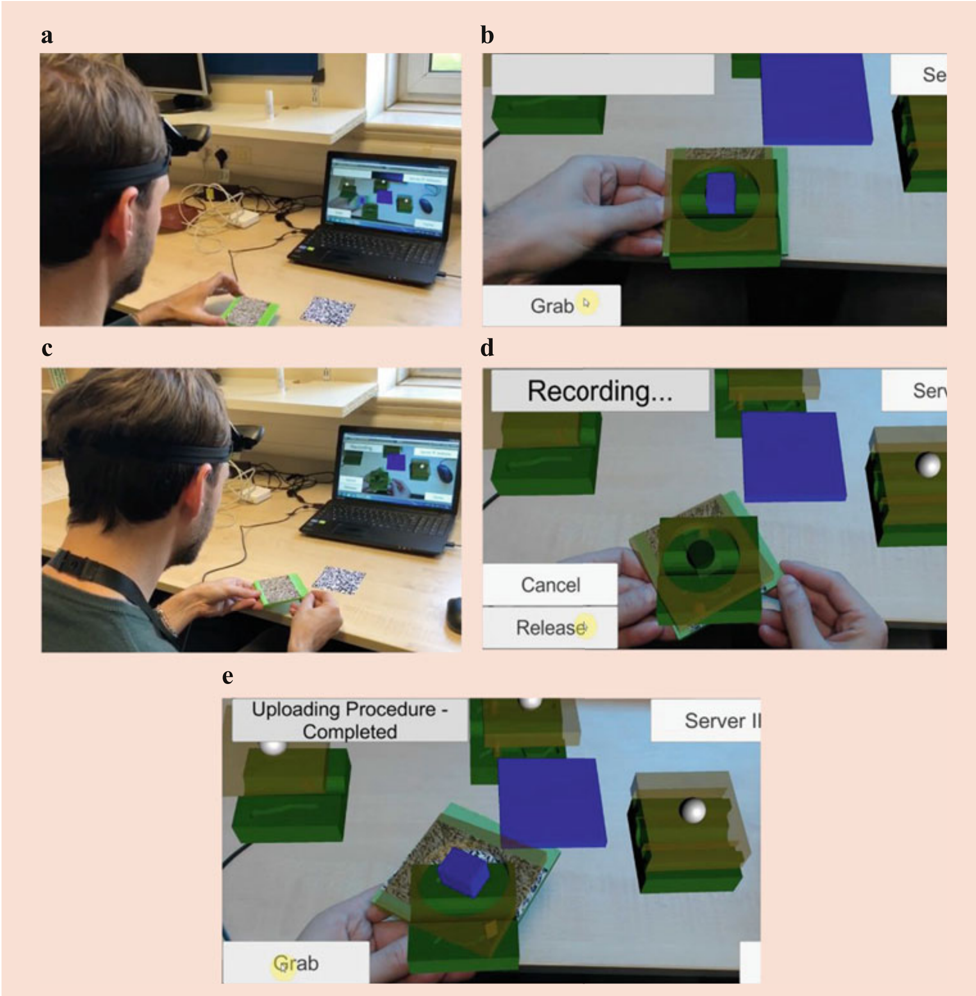


Fig. 27.3 Expert scenario since receiving the request for assistance (**a**)Virtual simulation, (**b**) Grab component, (**c**) Physical rotation, (**d**)Virtual rotation, (**e**) Procedure uploading

that has to be moved will be overlaid by its own CAD model in red. The latter is animated over the real one and moves as the expert has indicated previously (Fig. 27.3). The novice can now proceed and move the real object as indicated by the animation (Fig. 27.4c). In this specific case, the component has to rotate counterclockwise. Once the position and orientation indicated by the expert are reached by the real component, the overlaid CAD becomes green as shown in Fig. 27.4e. Both the novice and the expert can stop the procedure at any time through the specific UI button. For instance, the novice can stop it if, for any reason, he is not able to follow the animation; the expert should stop it if, while monitoring the movements of the virtual objects, he identifies an issue/mistake. It should be noted that independently from the orientation and position of the anchor marker with respect to the maintainer (novice and expert), the animations will always overlay on the correct object and move through the correct directions since these are "recorded" referencing to the anchor marker rather than the operator point of view. The novice should always address the correct component to be maintained and move it towards the correct direction, and therefore we expect the spatial referencing to improve with respect to voice/video support technologies.

Furthermore, the size of the workspace which can be covered in the FOV using this approach depends on the technologies utilized. For example, current implementation used one webcam on the novice side and one on the expert side; therefore the workspace is limited to the workbench where the camera is aiming at, which in this case the FOV is 130° wide. If using the head-mounted display, the size of workspace covered could be much bigger, but the FOV would be limited (e.g., FOV of Hololens is only 52°). Moreover, the increasing size and complexities of the virtual object can be handled with the integration of different technologies. For example, object recognition on novice side could be done through head-mounted display, therefore limiting the size and resolution to the novice FOV. On the other hand, object manipulation on the expert side could be done through haptic

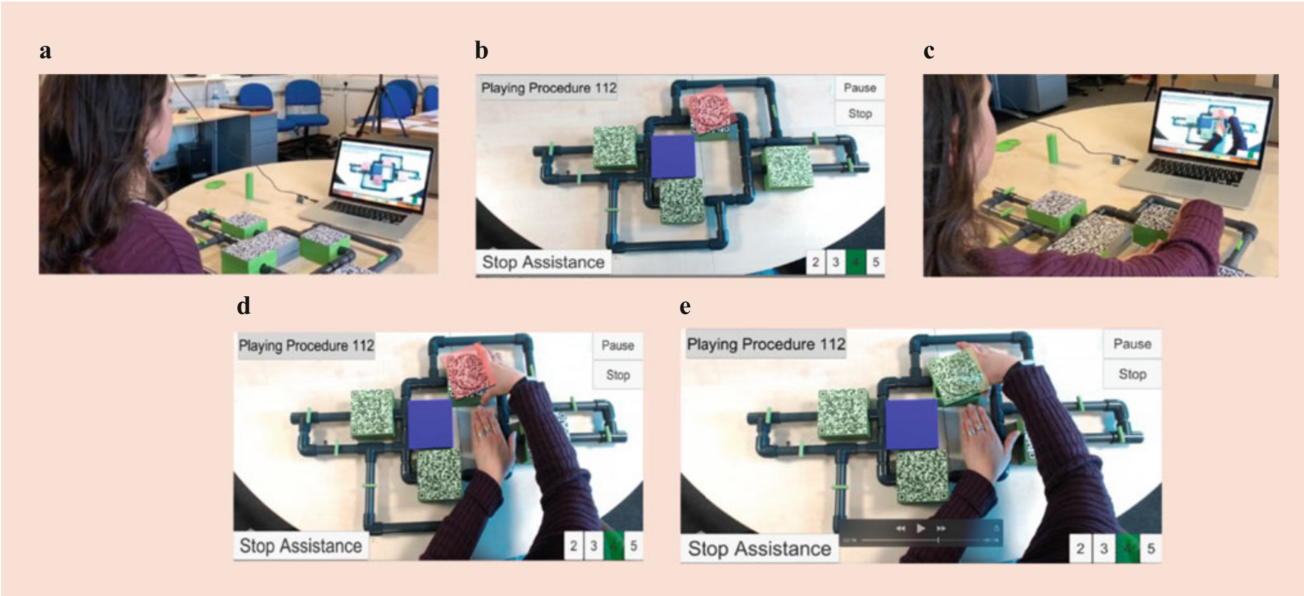


Fig. 27.4 Novice receives the remote support through ARRA (**a**) Novice environment, (**b**) Playing procedure, (**c**) Novice following procedure, (**d**) Novice moving component, (**e**) Component in final position

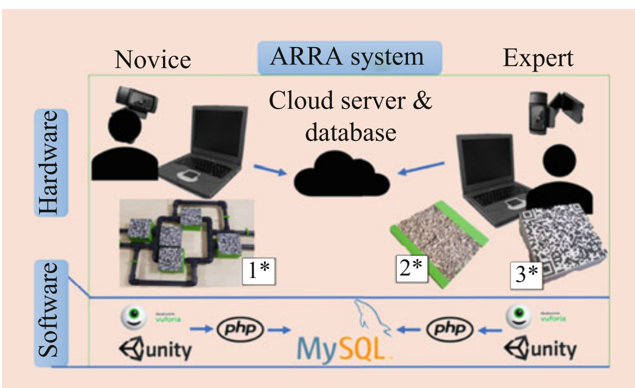


Fig. 27.5 ARRA system architecture

gloves in the future therefore enhancing the capability to manage complexity.

27.3.2 Technical Development

ARRA schematic concept (Fig. 27.1) and practical example have been described in the previous sections. ARRA approach formalizes in an AR system constituted by hardware and software. The hardware utilized is commercially available and can vary from one application to another as long as suitable for allowing the actions described in Fig. 27.1. The software has been developed specifically for this study and the hardware utilized in this study.

Figure 27.5 shows the system architecture utilized in this project. The novice maintainer is on the left, and the expert remote maintainer is on the right. Both of them were equipped with an RGB camera Logitech 1080, a laptop, and an anchor marker (3*, also described in Sect. 27.3.1). It is worth to mention that the RGB cameras' installation is different. The novice has it placed on his forehead through a strip while the remote expert has it installed facing the whole desk from a height of about 1.5 m. This is because the novice should intuitively face the object to be maintained (1*) while the expert does not know a priori where the object is located with respect to the anchor marker, and therefore, the whole desk needs to be in the FOV of the camera. Moreover, the remote maintainer also needs to have the virtual manipulator tool (2*): an object on which is placed the virtual manipulator marker. A different hardware solution utilizing HMD and hand gesture sensors could have been used to improve the AR experience and take rid of the virtual manipulator hardware since the manipulation of the virtual objects could have been done directly through the recognized hands. Unfortunately, the use of HMD would have obstructed the validation tests by not letting the tests observers understand what the tests participants were experiencing. The hand gesture recognition, on the other side, would have made easier the manipulation of the virtual object but required a more complex development without providing any advantage in quantifying the spatial awareness which is the scope of the study. The software for carrying on this study has been developed in Unity 3D and takes advantage of the Vuforia SDK for allowing the markers recognition. Rather than directly recognizing the objects, the authors decided to place markers (10×10 cm) on the objects for easing the validation test. The Unity application has been deployed for both Android and Windows. It has two user login kinds: (1) requesting assistance (novice maintainer) and (2) providing assistance (remote expert). It is worth to mention that the software does not allow video communication. The two maintainers communicate only using AR as described in the practical example in Sect. 27.3.1. The server has been firstly located on a local machine by utilizing XAMPP: open-source cross-platform web-server solution. Then it has been moved to a cloud server. The communication speed has not been affected due to the relatively small amount of information required to be exchanged to run ARRA. Only two tables of 8 columns are located on the server: (1) the Real Object DOF (RODOF) and (2) the Virtual Object DOF (VOROF). Quantitatively, the novice writes only about 70 Bytes per half-second, per object in RODOF. It corresponds to 6 numbers: 3 for the position and 3 for the orientation of the object with respect to the anchor marker. The expert reads these 70 Bytes and writes about the same amount of data on VOROF when manipulating the virtual object. In summary, considering 5 objects, the uploaded data goes from about 140 Bytes to 1.5 KB per second which is low compared to video-call support (300 KB/s for no HD). The architecture proposed in this section/project is the one utilized for carrying out the validation tests and therefore complies with the test observation requirements.

27.4 Test Design and Methodology

ARRA has been described in Sect. 27.3, both schematically and through a practical example. Among the expected benefit in the utilization of ARRA compared to video-call support for remote maintenance, the author intended to validate the improvement in terms of spatial referencing. To validate ARRA, the authors proceeded with the following three steps:

1. Quantification of the spatial referencing errors occurring when performing a maintenance operation supported through ARRA. The errors have been divided into three kinds:
   (a) Component identification
   (b) Component moving direction (for both translations and rotations)
   (c) Components coupling
2. Quantification of the spatial referencing errors occurring when performing the same maintenance operation as Step supported through "video-call support."
3. Comparison between 1 and 2.

The case study and therefore the maintenance operations utilized for testing purposes are reported in Sect. 27.4.1. The validation steps 1 and 2 have been calculated utilizing the test described in the following Sect. 27.4.2. The results have then been compared (step 3) and are shown in Sect. 27.5.

27.4.1 Validation Case Study

This section describes the case study utilized for validation purposes. The quantitative validation process is then described in detail in sec 27.4.2. The authors decided to utilize, as a case study, an operation that presents symmetries and difficulties in spatial referencing due to the resemblance of its component. Moreover, the case study had to comply with the following requirements:

1. Hard-copy manuals availability
2. Sufficient task complexity
3. Suitable dimensions for the available lab
4. Low occurrence maintenance hence suitable for the application of AR [11]
5. 3D printed simplified mock-up manufacturability

Therefore, it has been chosen to utilize complex hydraulic/pneumatic piping systems which are common in the oil and gas industry, pharmaceutical plants and energy factories.

For performing ARRA's validation test and quantify the improvements in terms of spatial referencing, the mock-up shown in Fig. 27.6 was 3D printed and assembled.

Starting from Fig. 27.6a, the mock-up consists of a piping system built utilizing ½″ PVC pipes connected through 90° elbows and tees. The piping path has been designed to have symmetries with respect to the two main piping directions. This has been done to add complications in terms of spatial referencing. Five "boxes" are visible in the figure. The four green ones will be called "locks" from now on. Each lock has a bottom component and a top component. These have been 3D printed and simulate any component which needs to be disassembled in order to be dismounted from its respective pipe. Each one of the four locks (A, B, C, and D) has a different locking system for coupling the top component with the bottom one. The gray box in the middle is the anchor marker support. In Fig. 27.6b, the markers for allowing object recognition has been applied. Moreover, lock C is opened (top and bottom component are separated) and it is possible to see its internal path. The latter is better shown in Fig. 27.6c. Locks C and A are opened and laid on the table (on the right). Locks B and D are closed and vertically shown on the left. Similarly, to the shaft-hole coupling, in this mock-up, the authors have designed the locks to have keys (indicated as Ck1 and Ck2 for lock C, as Ak1 and Ak2 for lock A) and paths/holes (indicated as Cp1 and Cp2 for lock C, as Ap1 and Ap2 for lock A). There is only one possible way to assemble the two components of each lock. For instance, Ck1 diameter can only get into Cp1. Finally, Fig. 27.6d showed an example of a defect that has to be fixed and lies under lock D. The locks have been designed in CATIA V5 and 3D printed in PLA utilizing the Ultimaker 2 printer. A material depositing head of 0.8 mm and layers of 0.6 mm is utilized.

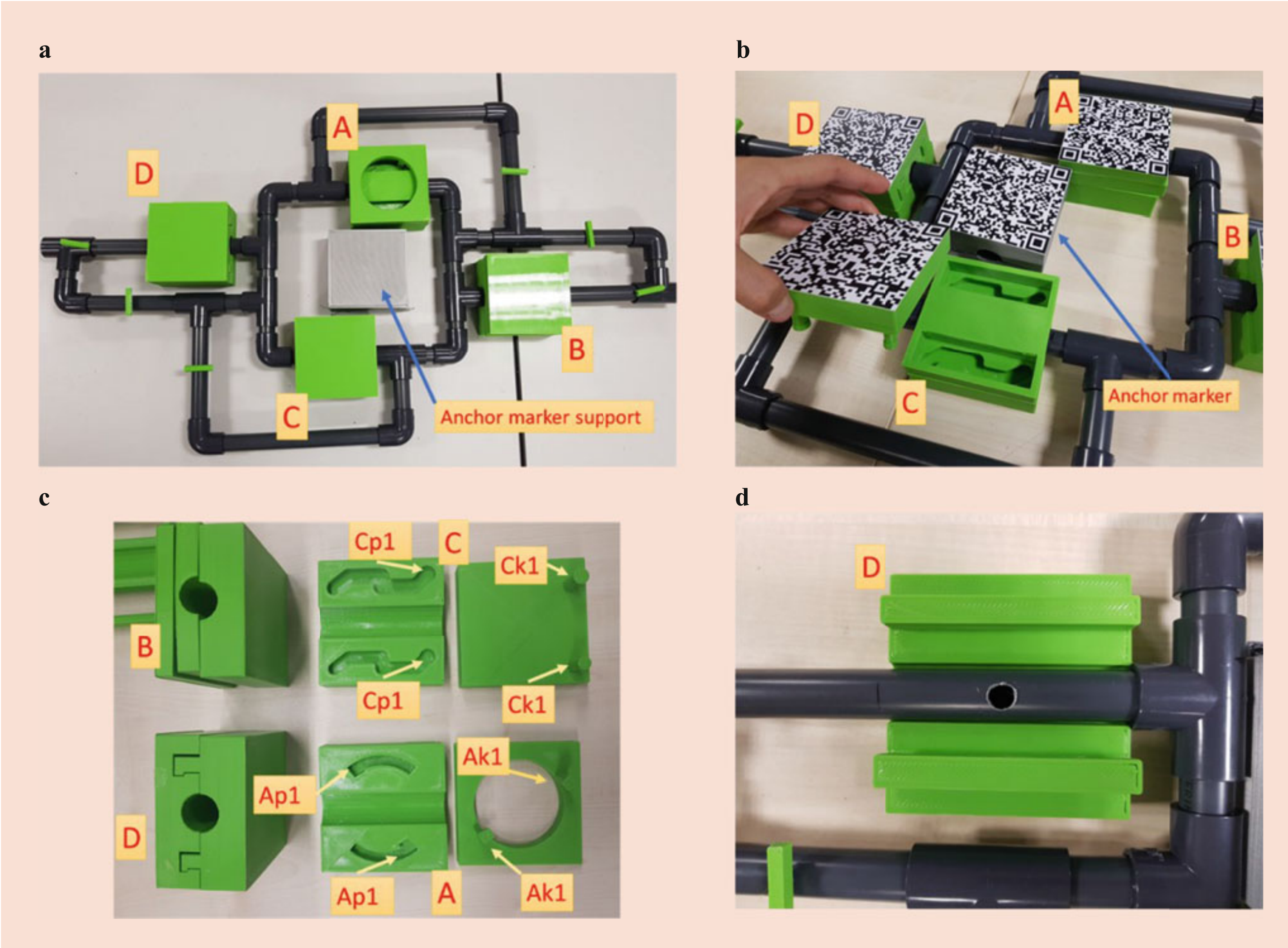


Fig. 27.6 3D printed mock-up of piping system for validation purposes (**a**) Full mock-up, (**b**) AR Markers, (**c**) Locks detail, (**d**) Piping defect example

27.4.2 Quantitative Validation Test Methodology

This section describes the method utilized for quantifying the spatial referencing error reduction due to the utilization of ARRA in comparison with video-call support.

Firstly, the quantification of the spatial referencing errors has been carried out separately for ARRA and video-call support utilizing respectively the method schematically described in Figs. 27.7 and 27.8. Then the results were statistically analyzed and compared for calculating the spatial referencing reduction. Following the timeline, on the top left, the participant is asked to read and sign the consent form as well as providing demographical data. The latter is used only for a qualitative analysis of the sample and does not affect the test results. The participant was then identified as a "novice" and was positioned in front of the assembly to be maintained (as in Fig. 27.4a) and introduced to ARRA by the observer. He could then request for assistance through ARRA, receive the procedure remotely built by an expert, and carry out the maintenance operation. The possible maintenance operations were eight and consisted of the assembly and disassembly of the four locks.

During maintenance operations, the observer will collect the spatial referencing information. The spatial referencing errors collected in this test can be of three kinds:

1. Wrong object identification
2. Wrong object direction
3. Wrong lock coupling (only applies assembly operations)

The first one occurred when the participant, after receiving the procedure, puts his hands on the wrong lock. The second one occurred when the component of the lock was moved towards an incorrect direction or rotated in the opposite sense. The last one consisted of associating the chosen top component of a lock with the bottom component of a different lock (only applies to assembly operations). The

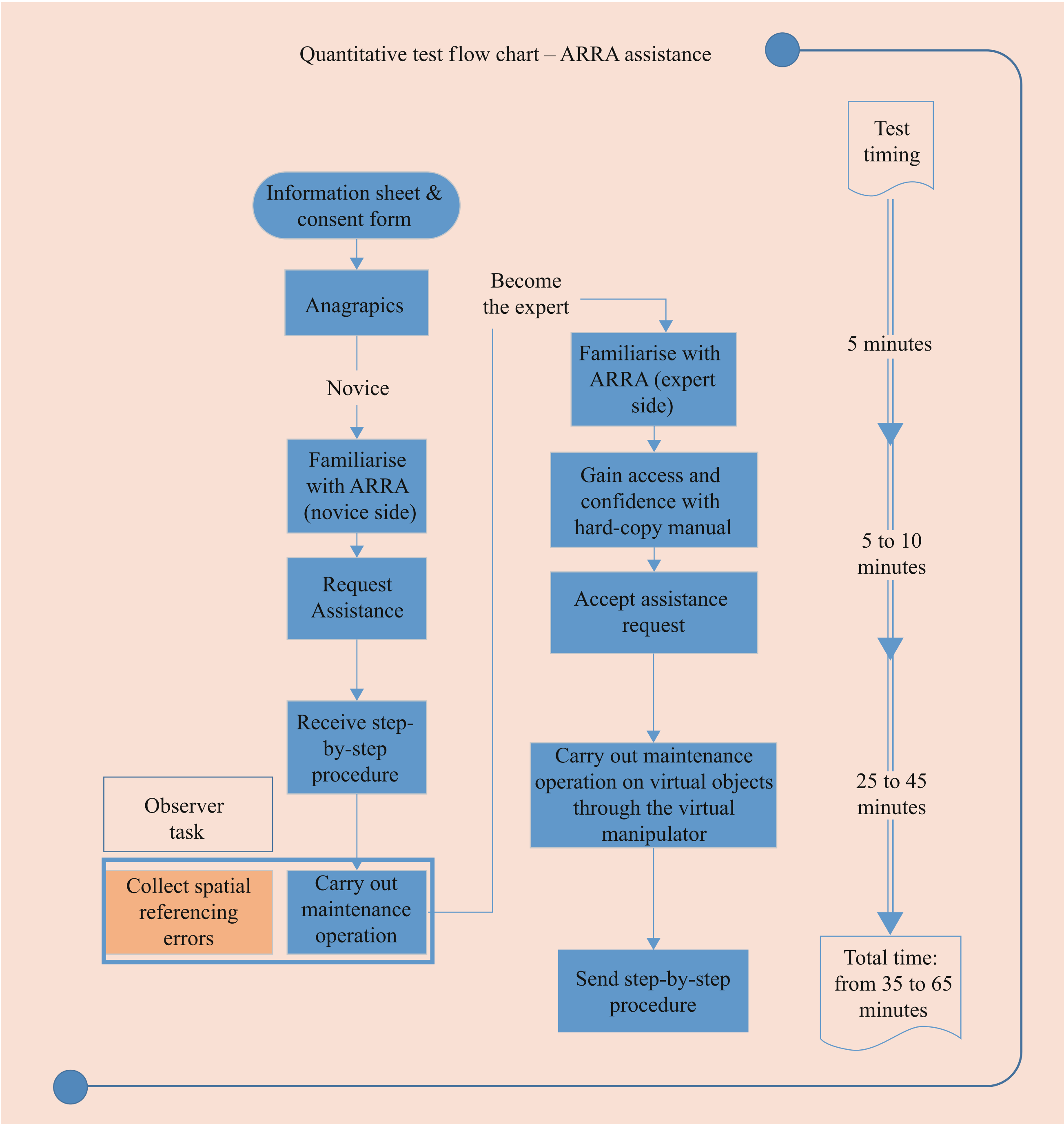


Fig. 27.7 Schematic representation of spatial referencing errors quantification test for ARRA

observer collected the data by filling a specifically designed form with a fixed multiple choice. For each of the spatial errors mentioned above, the observer can also choose among two descriptors: "opposite" and "other." The "opposite" was utilized when the participant:

1. Identifies the opposite component (with respect to the axis of symmetry)
2. Moves the object in the opposite direction
3. Couples the top component with the bottom component of the opposite lock

The "other" was utilized when the participant made a different kind of spatial referencing error. The "correct" was used when no spatial referencing error was made by the participant. The test was completed once the maintenance operation was carried out. The novice participant can now become a remote expert and provide assistance to the next

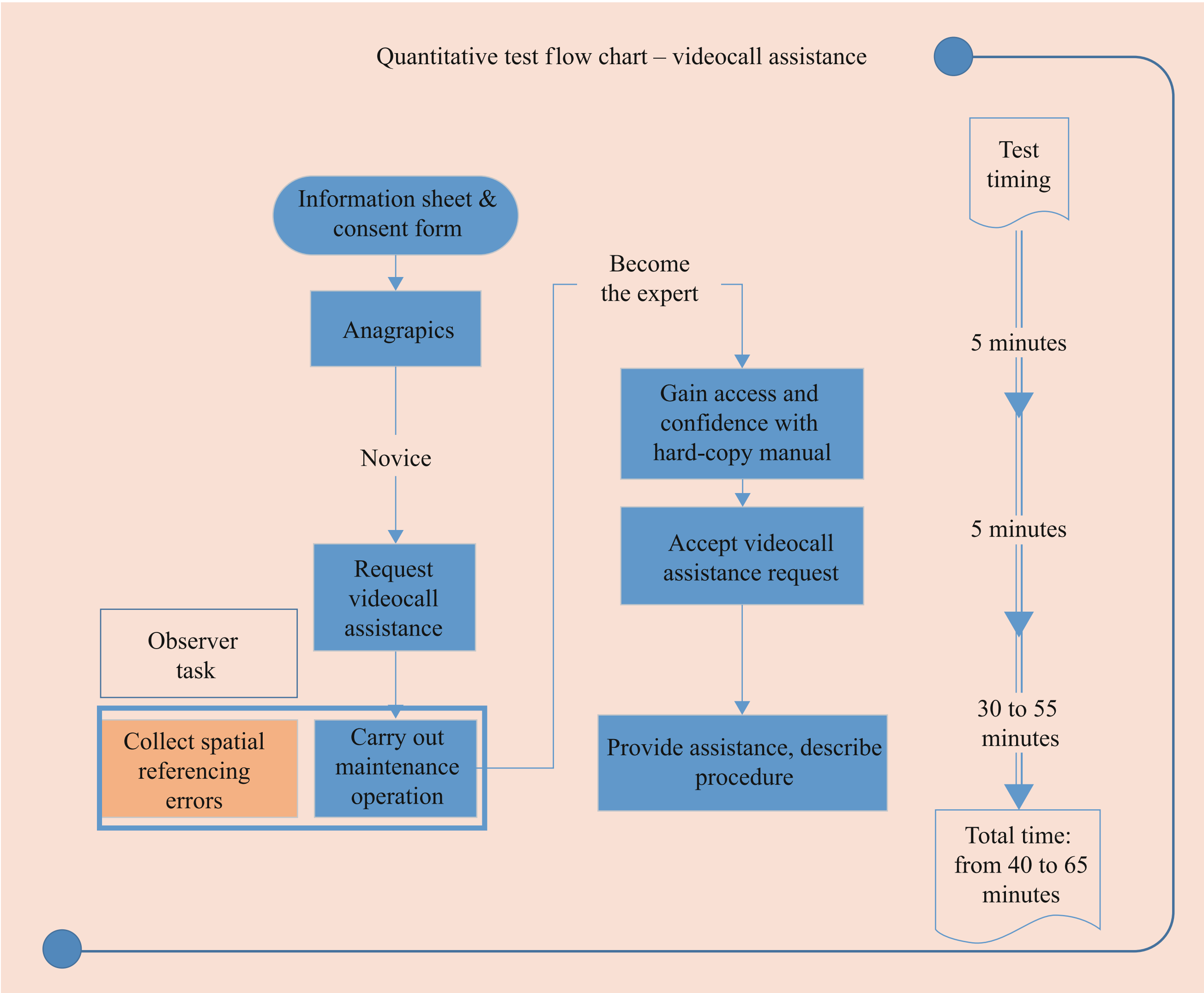


Fig. 27.8 Schematic representation of spatial referencing errors quantification test for video-call

novice participant. Taking advantage of the knowledge the first novice acquired during his test, providing him with more information about the assembly though a hardcopy manual, and showing him how the remote expert interface of ARRA works he is now able to virtually manipulate the locks as a remote expert. The spatial errors collected were compared with the one occurring when the same maintenance procedures were performed through video-call support. In this case, a new participant was placed in front of the assembly and was provided with an RGB camera for video-calling support. The orientation of the camera and the position of the participant with respect to the assembly were random. The randomness was eased by the utilization of a round table as a working area. On the other side, the expert was a participant that has already done the test as a novice who, moreover, was provided with the hardcopy manual. The observer collects the same data collected for quantifying the spatial errors considering ARRA support. The data collected in both scenarios were compared to calculate the final spatial errors reduction due to the utilization of ARRA versus video-call support. Table 27.1 is provided as an extract of the complete table of data collected during the tests.

In agreement with the methodology described in this section, Table 27.1 presents five columns. The first one lists the participant ID. The second column lists the method utilized for RA. The third column represents the operation carried out by the participant. These have been divided in 1–4 for disassembly and 5–8 for assembly of the four locks. The "spatial reference" and "error kind" columns report the data collected by the observer. For instance, in the first row, participant "1" correctly identified the object to be maintained in performing operation "4." The same participant has then wrongly moved the object in the opposite direction as reported in the second row. Participant "1" has been supported remotely through

Table 27.1 Extract of the complete dataset table utilized for further analysis

Participant ID	Remote assistance	Operation ID	Spatial reference	Error kind
1	ARRA	4	Correct	Identification
1	ARRA	4	Opposite	Direction
32	VIDEOCALL	3	Other	Coupling
32	VIDEOCALL	5	Correct	Coupling
34	VIDEOCALL	7	Other	Coupling
3	ARRA	1	Correct	Identification
4	ARRA	8	Correct	Direction
45	VIDEOCALL	1	Opposite	Identification
15	ARRA	6	Correct	Coupling

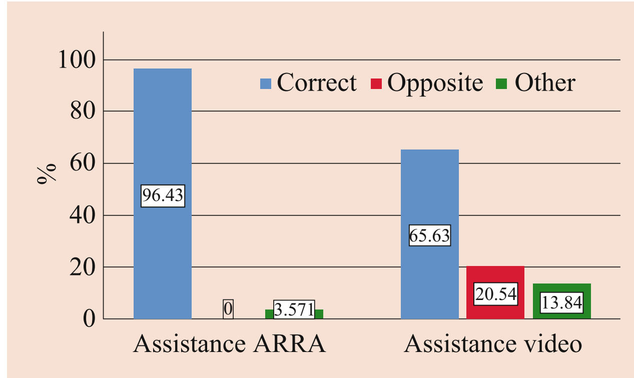


Fig. 27.9 Overall percentages of spatial referencing errors

ARRA. The analysis of the data and the results are reported in Sect. 27.5. The test aimed to quantify the improvement in terms of spatial referencing when performing a maintenance operation remotely supported by ARRA vs. video-call. A total of 60 participants (42 male/18 female) took part in this study. These included students and research staff from Cranfield University with higher education and/or engineering backgrounds as well as not academic people with no engineering background in a 50/50 ratio. Half of them performed the maintenance operation supported by ARRA; the other half were supported by video-call. The average age was 27.9 ($M = 21, 33$, $SD = 3.48$). Half of them performed the maintenance operation supported by ARRA; the other half were supported by video-call. On average, each participant carried out 3 of the 8 operations/tasks available. Each participant test took from 30 to 60 min for completion, and all the data collected has been stored in compliance with Cranfield University research ethics policy.

27.5 Analysis and Results

The data has been collected utilizing the methodology described in Sect. 27.4.2, and transcribed in a dataset shown in Table 27.1. The full table comprises of 450 rows. This number can be also calculated as reported in Eq. (27.1): where N is the number of rows, P is the number of participants, O is the operation performed by the participant, and E is the average number of error kinds.

$$N = P \times O \times E \tag{27.1}$$

From the equation above, the number of participants is 60, the operations performed by each participant for testing purposes were 3, and the average number of error kinds was 2.5. The latter was because, as already explained, for disassembly operations 1–4, the error kinds were 2: identification and direction. For assembly operations 5–8, the error kinds were 3: identification, direction, and coupling. Therefore, considering that each operation has been tested the same amount of time, the average is $(2 + 3)/2 = 5$.

To examine if a significant association exists between the RA methods utilized (ARRA vs. video-call) in terms of the amount of spatial referencing errors, it is required to perform a statistical test. Due to the nature of the sample, the authors decided to perform Pearson's chi-squared test. The sample, in fact, complies with the two test required assumptions:

1. The two variables should be measured at an ordinal or nominal level.
2. The two variables should consist of two or more categorical, independent groups.

The first assumption is verified since ARRA and video-call variables are measured at a nominal level through three categories that do not have an intrinsic order: correct, opposite, and other. The second assumption is verified since the two variables ARRA and video-call are two independent groups since the utilization of one excludes the utilization of the other. The result of Pearson's chi-square test is that there is a statistically significant association between ARRA and video-call, $\chi 2(2) = 72.68$, $p < 0.05$. The overall significant effect of the utilization of ARRA considering all the operations is shown in Fig. 27.9.

Figure 27.9 shows that, when utilizing ARRA, 96.43% of the tests resulted in "correct" spatial referencing. Only a small percentage of them resulted in other spatial referencing errors. On the other side, about 66% of the tests supported by video-call were performed correctly. About 20.5% of the tests resulted in presenting the spatial error defined as "opposite."

It occurred when the participant:

1. Identified the lock located in the opposite position with respect to the assembly symmetry
2. Moved the component in the opposite direction to the one he was expected to

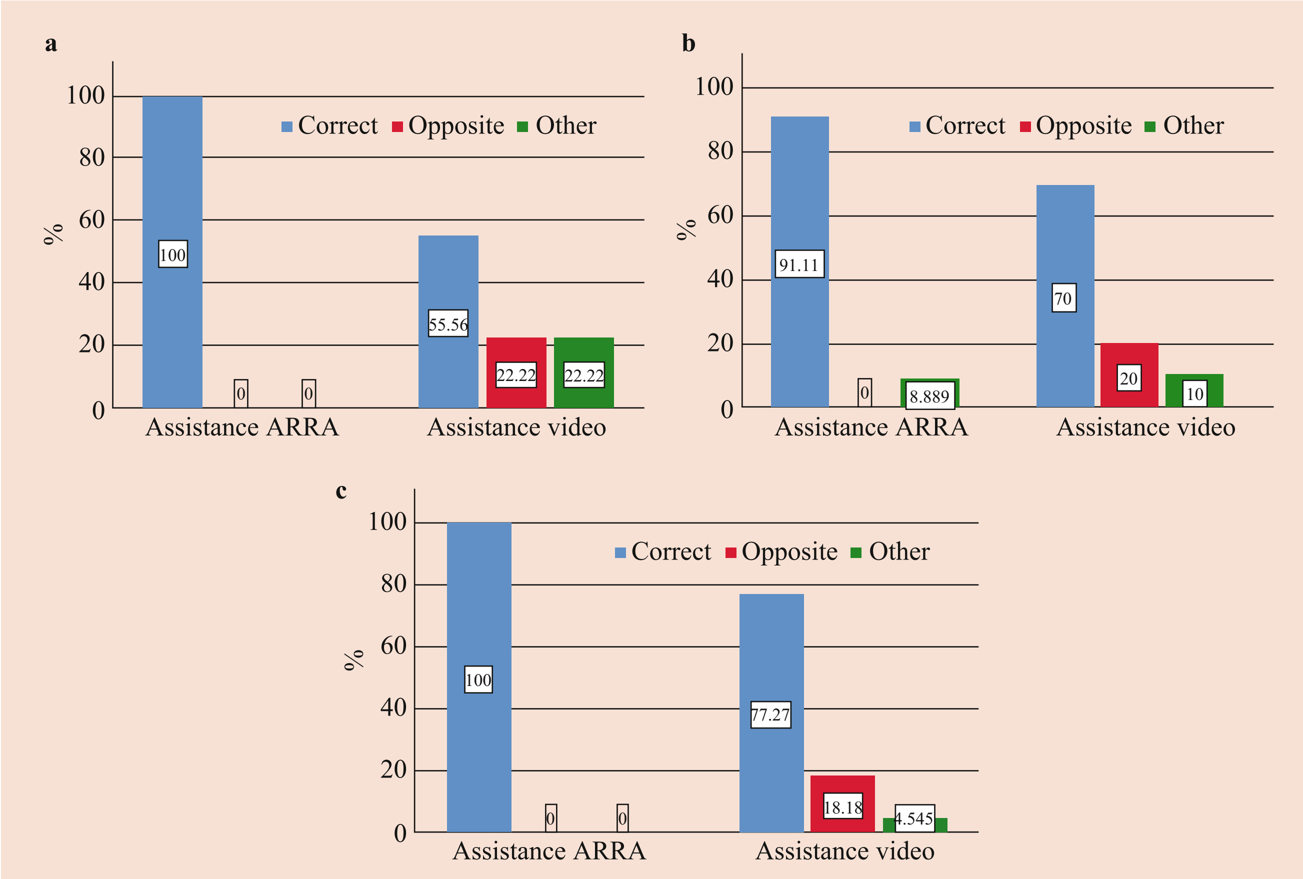


Fig. 27.10 Spatial referencing errors by kind: (**a**) identification, (**b**) directional, and (**c**) coupling

3. Intended to couple the top component of a lock with the opposite bottom of another lock

Moreover, about 14% of the tests resulted in other kinds of spatial referencing errors. Overall ARRA results in a 30% (correct-correct) improvement in terms of spatial referencing compared to video-call. For further understanding of the correlation between the errors and the operations, it has been found useful to plot the bar-chart of each "error kind" separately. These are shown in Fig. 27.10. It is worth to notice that ARRA performed perfectly (100% correct spatial referencing) for the identification of the objects (a) and the coupling (c) between the top and bottom components of the locks. About 9% of spatial errors were made in terms of moving directions (b). Furthermore, the authors investigated if the kind of operation (assembly or disassembly) affected the spatial referencing results (Fig. 27.11). Even though there is not a huge difference for ARRA in supporting an assembly or a disassembly operation, we can notice that video-call support results in slightly different outcomes. More specifically, for the disassembly operations, video-call support resulted in more "opposite" spatial errors than "other" (33% vs 8.7%). For assembly operations the percentages are inverted: 12% "opposite" vs 17% "other." Finally, each of the 8 operations has been plotted separately. Figure 27.12 reports the 4 disassembly operations.

In Fig. 27.12, the test, which utilized ARRA for disassembly operations (1–4) resulted in near-zero spatial errors. Only operation 2 (b) presents "other" spatial errors. Figure 27.13 reports the 4 assembly operations, each with the associated percentage of errors. As already shown also by Fig. 27.11, ARRA performed worst for assembly operations never reached the 100% correct spatial referencing.

27.6 Discussion

This section reports the discussion about the study methodology and results. The authors' intent in developing ARRA was to provide augmented reality support for RA. Moreover, the study focuses on quantifying the improvement in terms of spatial referencing due to the utilization of ARRA vs video-call support for maintenance. ARRA is based on two assumptions:

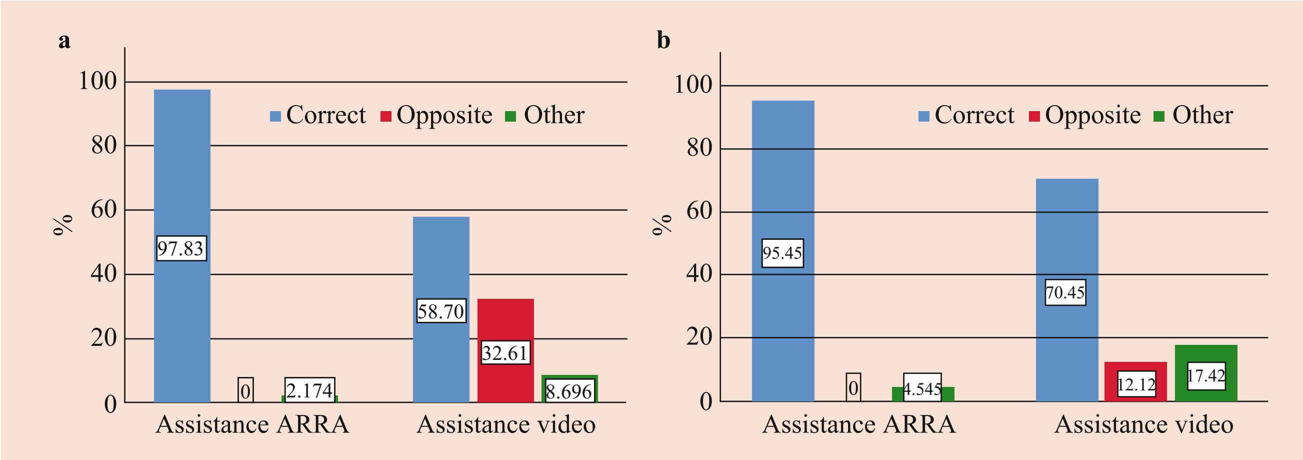


Fig. 27.11 Spatial referencing errors collected for disassembly and assembly operations (**a**) Disassembly, (**b**) Assembly

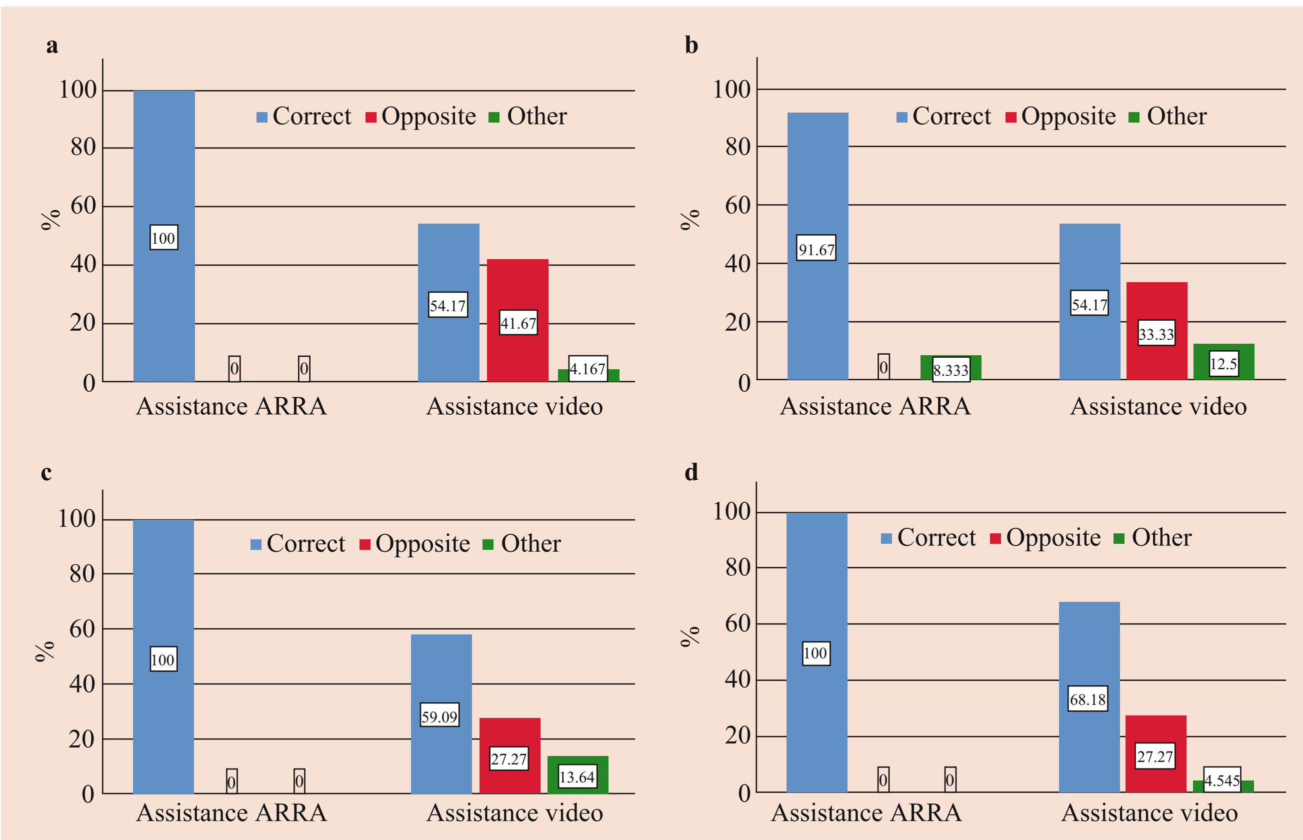


Fig. 27.12 Spatial referencing errors collected for each disassembly operation separately (**a**) Operation ID1, (**b**) Operation ID2, (**c**) Operation ID3, (**d**) Operation ID4

1. The system is able to recognize and track the object in the FOV of the remote novice maintainer.
2. The CAD models of the objects to be maintained are available.

The participants have been remotely assisted through ARRA or video-call support. Quantitative spatial referencing errors data has been collected. The results have shown a 30% of improvement in terms of spatial referencing when utilizing ARRA as remote assistance support vs video-call support. These improvements have been found to be due to an increase of spatial awareness. The AR system efficiency, in fact, is invariant with respect to the technician Point Of View (POV) since it relies only on the real environment configuration.

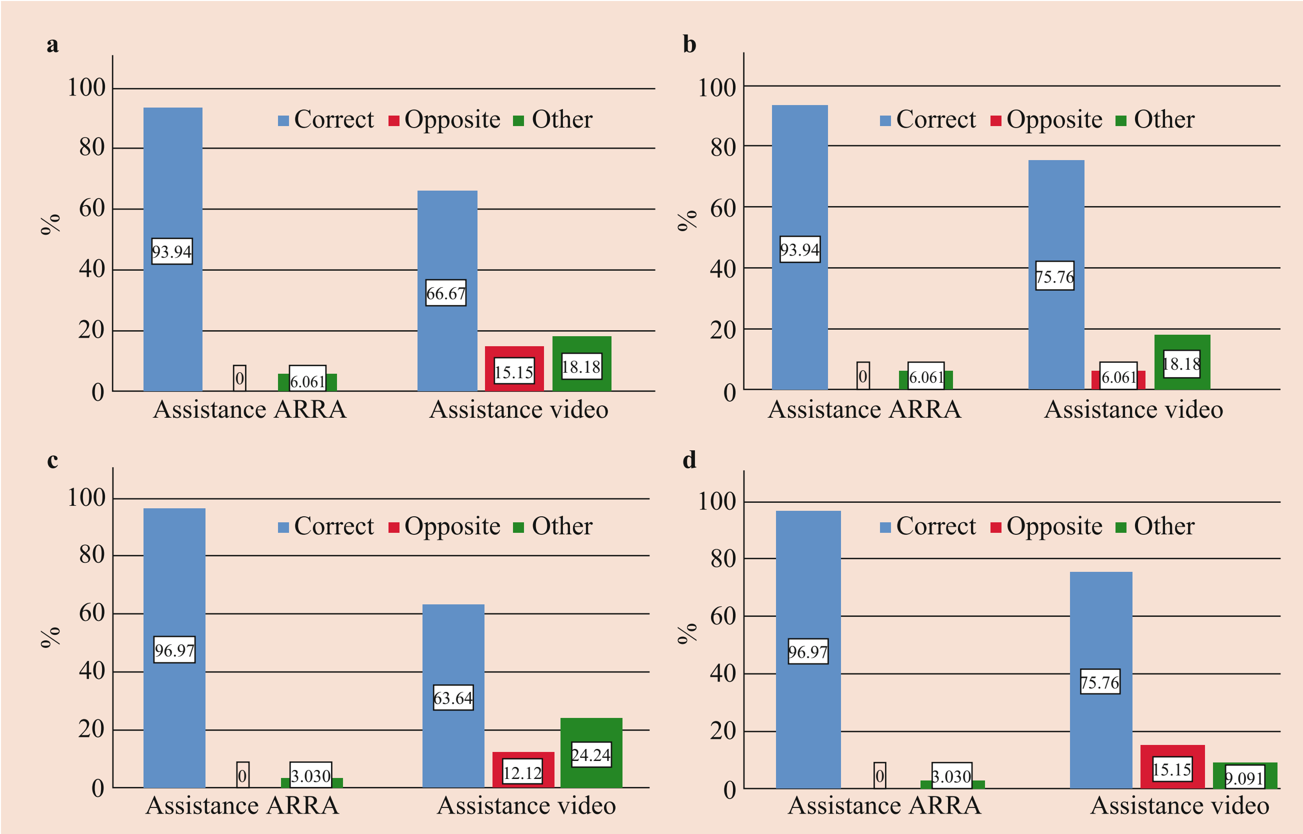


Fig. 27.13 Spatial referencing errors collected for each assembly operation separately (**a**) Operation ID5, (**b**) Operation ID6, (**c**) Operation ID37, (**d**) Operation ID8

The video-call support, on the other side, relies on the ability of the technicians to communicate and to understand each other's POV.

The authors consider these assumptions plausible due to the recent improvements in image processing, object tracking and recognition, and hardware (processors and sensors) [11, 23]. ARRA has been described in Sect. 27.3 through a practical example and technical development. Even though in this study the authors utilized some specific hardware and software solutions, ARRA can be developed and implemented differently. Considering the fast advancement of the technology related to AR, it could be useful to exploit the utilization of depth sensors for the recognition of the objects. Moreover, an HMD would be more suitable for industrial applications. It could not be utilized in this study only for validation reasons. The observer of the empirical tests that have been carried out needed to clearly understand the evolvement of each test for collecting the data required for assessing the spatial referencing improvements.

This study focuses on quantifying the spatial referencing errors occurring when utilizing ARRA vs. video-call support for RA. The methodology utilized for the empirical tests took inspiration from similar studies [26–28]. The case study utilized, even though apparently might not seem complex, hides several challenges. First of all, the full assembly presents symmetries and similitudes. All the components have the same external shape and color and, therefore, are difficult to be identified through voice indications or hard-copy manuals. Moreover, every one of the 4 locks has a different unlocking system. All together comprise x, y, and z translations and z rotation. The tests were planned carefully and the small space was given to subjectivity. The observer was provided with multiple-choice forms and a detailed schematic process for carrying on the tests. Regarding the results, ARRA performed always better than video-call support in terms of spatial referencing. This is because AR relies on the spatial references recognized by the software and is invariant with the orientation of the camera. Video-call support, on the other side, relies on the voice communication between the expert and the novice. The reference system, which is in the expert mind, might be different from the one of the novice. For instance, if the expert indicated to grab "the object on the right," the novice might have grabbed the object, which was at his right. Sometimes this resulted in grabbing the correct object, but sometimes not. This is the reason why, for all the operations (see Figs. 27.12 and 27.13), video-call support always presented an unneglectable percentage of spatial referencing errors of the kind "opposite." Furthermore, from the types of errors, identification, direction, and coupling (Fig. 27.10), we can see that ARRA only resulted in spatial errors within the direction category. It means that, when ARRA support indicated a direction of movement for any object in any operation, it resulted in a 10% error of the kind "other." In other words, the participant did not move the object towards the correct direction and not even the opposite direction. He moved the object towards a completely different direction. The authors found a plausible justification thanks to Fig. 27.12b. The latter shows that within the 4 disassembly operations, only operation "2" presented directional spatial errors when utilizing ARRA. Operation "2" consisted of the disassembly of the top component of lock B (see Fig. 27.6a). It was done by rotating the top component around the "z" axis and was also reported in the practical example in Fig. 27.4. Due to the inclination of the camera with respect to the assembly, the rotation was sometimes (10% of the time) confused with a pulling movement and therefore resulted in a spatial referencing error.

27.7 Conclusion and Future Work

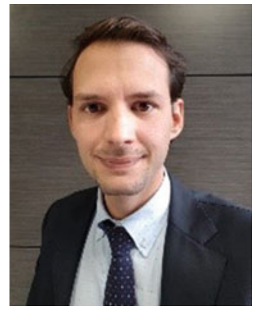
This study proposes Augmented Reality support for Remote Assistance: ARRA. ARRA allows a remote expert to visualize in real time the novices' maintenance problem and guide him through the solution. The remote expert can build step-by-step procedures through the virtual manipulation of the virtual objects and overlay the procedures into the real novice's working environment. Among the challenges in remote assistance, ARRA attempts to overcome the spatial referencing issues. These can be seen as the difficulties the remote expert has in explaining the novice what he has to do without knowing his spatial references and having full control of the maintenance environment. Therefore, ARRA has been tested and validated considering three spatial referencing errors: (1) the identification of the objects, (2) the movements of the objects, and (3) the coupling of two objects. The case study utilized was a mock-up of a piping system. The comparison of ARRA was made with remote assistance through video-call. The results indicated an overall improvement of 27% in terms of correct spatial referencing operation when utilizing ARRA in comparison with the video-call. Moreover, ARRA performed perfectly when considering identification and coupling errors. The tests regarding the direction of the objects, on the other side, showed an unneglectable percentage of errors of about 10%. Further research needs to investigate if the utilization of HMD and a more advanced UI in ARRA could overcome directional spatial referencing errors and close to 100% of correct operations for similar assemblies.

Acknowledgments This research is sponsored by HSSMI – High Speed Sustainable Manufacturing Institute available at https://www.hssmi.org.

References

1. Zhu, J., Ong, S.K., Nee, A.Y.C.: A context-aware augmented reality system to assist the maintenance operators. Int. J. Interact. Des. Manuf. **8**, 293–304 (2014). https://doi.org/10.1007/s12008-013-0199-7
2. Masoni, R., Ferrise, F., Bordegoni, M., Gattullo, M., Uva, A.E., Fiorentino, M., Carrabba, E., Di Donato, M.: Supporting remote maintenance in industry 4.0 through augmented reality. Procedia Manuf. **11**, 1296–1302 (2017). https://doi.org/10.1016/j.promfg.2017.07.257
3. Azuma, R.T.: A survey of augmented reality. Presence Teleoperators Virtual Environ. **6**, 355–385 (1997). https://doi.org/10.1162/pres.1997.6.4.355
4. Azuma, R., et al.: Recent advances in augmented reality. In: IEEE Comput. Graph. Appl, pp. 34–47 (2011)
5. Abramovici, M., Wolf, M., Adwernat, S., Neges, M.: Context-aware maintenance support for augmented reality assistance and synchronous multi-user collaboration. Procedia CIRP. **59**, 18–22 (2017). https://doi.org/10.1016/j.procir.2016.09.042
6. Neumann, U., Majoros, A.: Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In: Proc. – Virtual Real. Annu. Int. Symp, pp. 4–11 (1998). https://doi.org/10.1109/vrais.1998.658416
7. Haritos, T., Macchiarella, N.D.: A mobile application of augmented reality for aerospace maintenance training. AIAA/IEEE Digit. Avion. Syst. Conf. – Proc. **1**, 1–9 (2005). https://doi.org/10.1109/DASC.2005.1563376
8. Mura, M.D., Dini, G., Failli, F.: An integrated environment based on augmented reality and sensing device for manual assembly workstations. Procedia CIRP. **41**, 340–345 (2016). https://doi.org/10.1016/j.procir.2015.12.128
9. Funk, M., Kritzler, M., Michahelles, F.: HoloCollab: a shared virtual platform for physical assembly training using spatially-aware head-mounted displays. ACM Int. Conf. Proc. Ser. (2017). https://doi.org/10.1145/3131542.3131559
10. Oda, O., Elvezio, C., Sukan, M., Feiner, S., Tversky, B.: Virtual replicas for remote assistance in virtual and augmented reality. In: UIST 2015 – Proc. 28th Annu. ACM Symp. User Interface Softw. Technol, pp. 405–415 (2015). https://doi.org/10.1145/2807442.2807497
11. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. Robot. Comput. Integr. Manuf. **49**, 215–228 (2018). https://doi.org/10.1016/j.rcim.2017.06.002
12. Nee, A.Y.C., Ong, S.K., Chryssolouris, G., Mourtzis, D.: Augmented reality applications in design and manufacturing. CIRP Ann. – Manuf. Technol. **61**, 657–679 (2012). https://doi.org/10.1016/j.cirp.2012.05.010
13. Van Krevelen, D.W.F., Poelman, R.: A survey of Augmented Reality technologies, applications and limitations. Int. J. Virtual Real. **9**, 1–20 (2010). https://doi.org/10.20870/ijvr.2010.9.2.2767
14. Wang, X., Love, P.E.D., Kim, M.J., Wang, W.: Mutual awareness in collaborative design: an Augmented Reality integrated telepresence system. Comput. Ind. **65**, 314–324 (2014). https://doi.org/10.1016/j.compind.2013.11.012
15. Wang, M.J., Tseng, C.H., Shen, C.Y.: An easy to use augmented reality authoring tool for use in examination purpose. IFIP Adv. Inf. Commun. Technol. **332**, 285–288 (2010). https://doi.org/10.1007/978-3-642-15231-3_31

16. Bottecchia, S., Cieutat, J.M., Jessel, J.P.: T.A.C: augmented reality system for collaborative tele-assistance in the field of maintenance through internet. In: ACM Int. Conf. Proceeding Ser (2010). https://doi.org/10.1145/1785455.1785469
17. Bordegoni, M., Ferrise, F., Carrabba, E., Di Donato, M., Fiorentino, M., Uva, A.E.: An application based on Augmented Reality and mobile technology to support remote maintenance. Conf. Exhib. Eur. Assoc. Virtual Augment. Real. **1**, 131–135 (2014). https://doi.org/10.2312/eurovr.20141351
18. Cologni, A.L., Fasanotti, L., Dovere, E., Previdi, F., Bonfanti, S., Owen, F.C.: Smartphone based video-telemetry logger for remote maintenance services. IFAC-PapersOnLine. **28**, 822–827 (2015). https://doi.org/10.1016/j.ifacol.2015.06.185
19. Mourtzis, D., Vlachou, A., Zogopoulos, V.: Cloud-based augmented reality remote maintenance through shop-floor monitoring: a product-service system approach. J. Manuf. Sci. Eng. Trans. ASME. **139**, 1–11 (2017). https://doi.org/10.1115/1.4035721
20. Erkoyuncu, J.A., del Amo, I.F., Dalle Mura, M., Roy, R., Dini, G.: Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality. CIRP Ann. – Manuf. Technol. **66**, 465–468 (2017). https://doi.org/10.1016/j.cirp.2017.04.006
21. Csurka, G., Kraus, M., Laramee, R.S., Richard, P., Braz, J.: Computer Vision, Imaging and Computer Graphics – Theory and Applications. Springer Berlin Heidelberg (2012)
22. Knöpfle, C., Weidenhausen, J., Chauvigné, L., Stock, I.: Template based authoring for AR based service scenarios. In: Proc. – IEEE Virtual Real, pp. 237–240 (2005). https://doi.org/10.1109/vr.2005.1492779
23. Yu, L., Ong, S.K., Nee, A.Y.C.: A tracking solution for mobile augmented reality based on sensor-aided marker-less tracking and panoramic mapping. Multimed. Tools Appl. **75**, 3199–3220 (2016). https://doi.org/10.1007/s11042-014-2430-3
24. Wikikart99: Piping ystem on a chemical tanker, https://commons.wikimedia.org/wiki/File:A_UPW_Installation_using_PVDF_Piping.png. Last accessed 03 Sept 2020
25. Cozanet, H.: A UPW Installation using PVDF Piping, https://commons.wikimedia.org/wiki/File:Piping_system_on_a_chemical_tanker.jpg. Last accessed 03 Sept 2020
26. Tang, A., Owen, C., Biocca, F., Mou, W.: Experimental evaluation of augmented reality in object assembly task. In: Proc. – Int. Symp. Mix. Augment. Reality, ISMAR 2002, pp. 265–266 (2002). https://doi.org/10.1109/ISMAR.2002.1115105
27. Fiorentino, M., Uva, A.E., Gattullo, M., Debernardis, S., Monno, G.: Augmented reality on large screen for interactive maintenance instructions. Comput. Ind. **65**, 270–278 (2014). https://doi.org/10.1016/j.compind.2013.11.004
28. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE Trans. Vis. Comput. Graph. **17**, 1355–1368 (2011). https://doi.org/10.1109/TVCG.2010.245



Riccardo Palmarini completed his PhD from Cranfield University in 2019 and currently works as an Instrument and Automotion Engineer at TechnipFMC. Riccardo also completed an MSc at Cranfield University in Aircraft Vehicle Design. His PhD project focused on developing Augmented Reality (AR) based visualization system to support with the delivery of maintenance and\or manufacturing.

Iñigo Fernández del Amo completed his PhD at Cranfield University in 2020 in Augmented Reality for Maintenance. His primary interest is in how Augmented Reality systems can capture knowledge from maintainers to provide sustainable efficiency improvements. Iñigo graduated last year from a double-degree in Industrial Engineer and Manufacturing Consultancy between the Polytechnic University of Madrid (UPM) and Cranfield University. He has won the Cranfield Vice-Chancellor's Prize 2017 for the most outstanding MSc student.

Dedy Ariansyah completed his PhD from Politecnico di Milano in 2018 in methods and tools for product design and held a post-doctoral position in Augmented Reality (AR) and Virtual Reality (VR) development for manufacturing and maintenance until 2019. He currently holds a research fellow position at Cranfield University since 2019 in AR for Through-life Engineering. His current research is funded by EPSRC under Digital Toolkit for optimization of operators and technology in manufacturing partnerships (DigiTOP) project. His work is focused on the integration of Digital Twin and AR, including the Human Factors aspect.

John Ahmet Erkoyuncu completed his PhD from Cranfield University in 2011 in uncertainty modelling for maintenance and completed MSci degree in Applied Statistics from Imperial College London. Since 2020, John is a Professor of Digital Engineering in Digital Service Engineering at Cranfield University. John is active with Innovate UK and EPSRC funded projects in the UK around research topics: digital twins, augmented reality, digitization (of degradation assessment), and simulation of complex manufacturing and maintenance procedures. Prof. Erkoyuncu has published over 100 conference and journal papers in internationally leading venues. John is a Chartered Engineer and a Member of IET.

Rajkumar Roy is the Dean of School of Mathematics, Computer Science and Engineering at City, University of London. He joined City, University London from Cranfield University, where he was Director of Manufacturing. Professor Roy holds a PhD in Computing from the University of Plymouth (UK) and BEng and MEng degrees in Production Engineering from Jadavpur University in India. He started his career as an engineer at Tata Motors; pioneered research in Through-life Engineering Services (TES) with Rolls-Royce, BAE Systems, Bombardier Transportation, the Ministry of Defence and Babcock International; and established an internationally known TES Centre. Professor Roy's cost engineering and obsolescence research has transformed contemporary understanding of the engineering effort required to design, make, and support high-value products, resulting in tools used by BAE Systems, Airbus, the Ministry of Defence, Rolls-Royce, and Ford Motor Company. Professor Roy is a Founding Editor-in-Chief of the Elsevier Applied Soft Computing journal and a Fellow of the CIRP (International Academy for Production Engineers), the Institution of Engineering and Technology (IET), the Institute of Engineering Designers (IED), and the Higher Education Academy (HEA).

Part VI

Applications in Health Science

Augmented Reality for Computer-Guided Interventions

28

Stephane Cotin and Nazim Haouchine

Contents

28.1 From Medical Image Computing to Computer-Assisted Interventions 689

28.2 Imaging and Anatomical Modeling 693

28.3 Surgical Vision 694

28.4 Registration and Pose Estimation 696

28.5 Physics-Based Modeling 697
28.5.1 Soft Tissue Biomechanics 697
28.5.2 Strategies for Real-Time Computation 699
28.5.3 Beyond Surgery 699

28.6 Visualization and Perception 700

28.7 Related Topics 700
28.7.1 Training and Knowledge Transfer 701
28.7.2 Surgical Planning 702

28.8 The Future of AR in Medicine 703

References 703

Abstract

With the development of augmented reality devices, new imaging hardware, and better computer simulations, the field of computer-assisted surgery is rapidly evolving. From an augmented visualization of medical images during an intervention to a fully automatic fusion of the patient's virtual anatomy with the surgical view, many ways of augmenting a surgery can be envisioned. This chapter provides a description of the main challenges, state of the art, and results in this revolutionary area by looking at the contributions of augmented reality in several medical application fields with an emphasis on deformable techniques. We will address several key questions, such as the creation of patient-specific models of the anatomy, the registration of the virtual model onto the actual view, or the real-time computation of biophysical phenomena on this anatomy.

Keywords

Computer-assisted intervention · Augmented and virtual reality · Patient-specific biomechanical modeling · Medical image registration · Pose estimation · Surgical vision · Visualization and perception · Knowledge transfer · Surgical planning and training

28.1 From Medical Image Computing to Computer-Assisted Interventions

Since the first use of X-rays in 1895, medical imaging has improved and diversified and progressively transformed medicine. This major evolution was, however, only feasible thanks to the joint development of computer sciences and image processing methods, which made it possible to interpret and process increasingly larger and more complex images (such as computed tomography or magnetic resonance images) with greater precision and efficiency (see Fig. 28.1). Many interventions now involve the acquisition of medical images of the patient prior to the actual operation, even in emergency procedures. Associated with efficient image processing, this makes it possible to improve the diagnostic and choose the most appropriate therapy.

Today, it is quite frequent to also rely on medical imaging (such as ultrasound or X-ray) to guide the intervention, as these images can be acquired in real time at the point of care. This field of medicine, where images are used to visualize the internal anatomy of the patient during the procedure, is rapidly expanding to cover many types of interventions, performed under various types of images. Originally named *interventional radiology*, this modern approach of medicine

S. Cotin (✉)
MIMESIS Team, Inria, Strasbourg, France
e-mail: stephane.cotin@inria.fr

N. Haouchine
Surgical Planning Laboratory, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_28

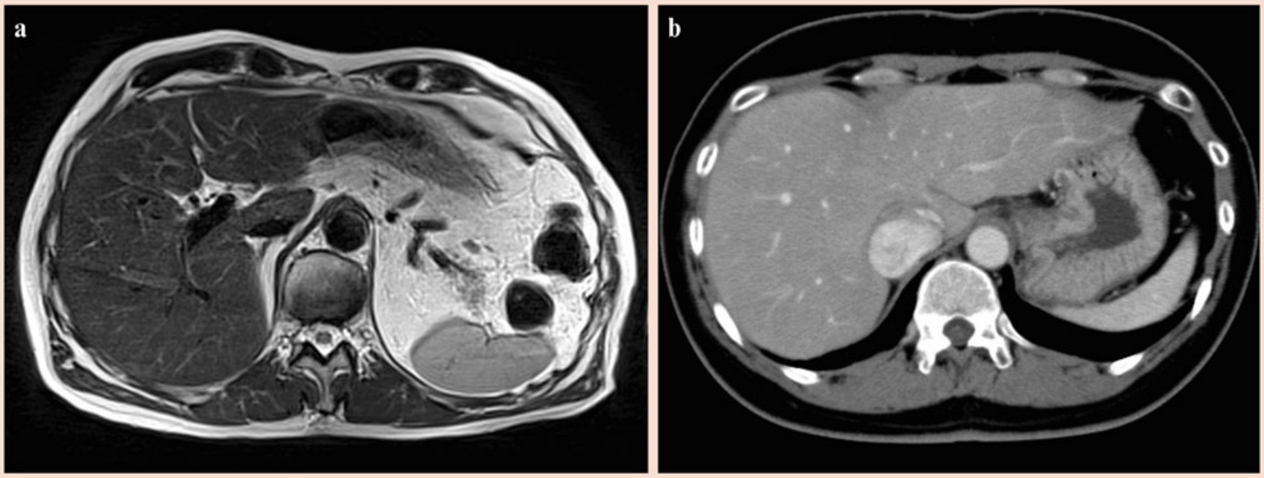


Fig. 28.1 Examples of medical images. (**a**) a slice of an image of the abdomen obtained from a magnetic resonance scanner; (**b**) computed tomography image of a similar area

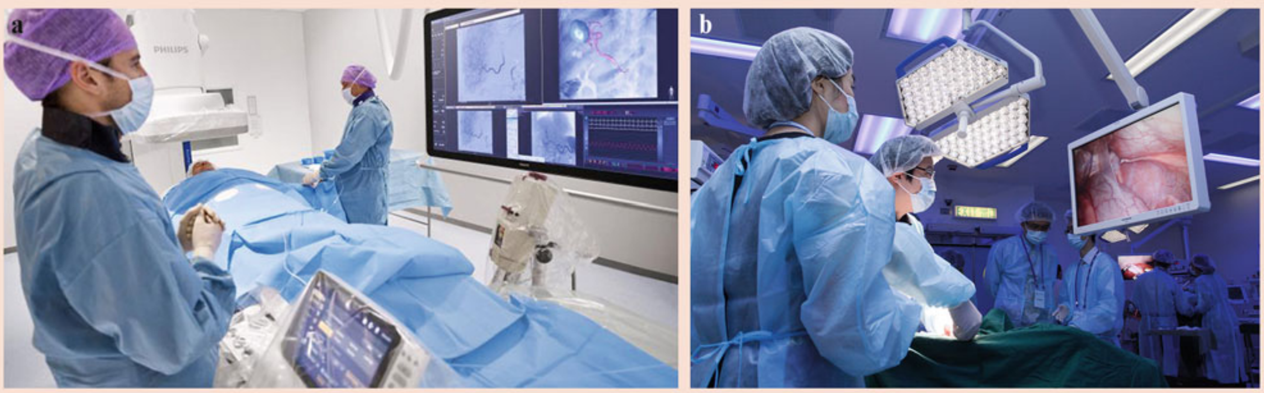


Fig. 28.2 (**a**) Vascular surgery is based on the use of flexible micro-instruments (catheters and guidewires) navigating to the pathology via the vascular network. The intervention is visualized using a real-time X-ray imaging system called fluoroscopy. (**b**) General principle of laparoscopic surgery – miniaturized instruments, as well as a camera, are introduced into the abdomen via small incisions. The surgeon then operates by looking at a monitor on which the view from the camera is displayed

is now often called *image-guided therapy*. It aims at being less invasive and more precise, therefore leading to shorter hospital stays. Laparoscopic [1] (see Fig. 28.2b) and vascular [2] surgery (see Fig. 28.2a) are two examples of such image-guided therapies.

Yet, while imaging has radically evolved, how images are displayed and accessed is basically the same as it was 50 years ago. Visual data are shown on 2D monitors, except in some rare cases, such as in (robotic) surgery, where the camera inserted in the patient's body is sometimes equipped with dual optics allowing for stereoscopic visualization of a 3D monitor. However, the display setup generally forces healthcare providers to look away from the patient, and even away from their own hands while operating [3] (see Fig. 28.2).

The images are also not displayed from the perspective of the viewer but rather from that of the imaging device [4]. As a result, doctors have to rely on their experience and anatomical knowledge to analyze and mentally project the images into the patient while they are performing the surgery. To make things worse, it is frequent that several images and data are needed to assess the situation in the operating room for the clinician to decide on the next best step. When relying on multiple images and monitors, it is easy to miss vital cues regarding the patient's status. With this information being displayed on different monitors, doctors have to mentally fuse multiple image types, such as ultrasound and computed tomography, into a coherent representation of the patient. Acquiring this skill takes years of training.

Augmented reality (AR) has the potential to change all of this [3]. Using dedicated hardware to track the user and the equipment, new display technologies to visualize data, and three-dimensional models of the anatomy and algorithms to compute their change of state over time, AR can superimpose digital information on the physical world, making it easier for the clinician to analyze. If the commercial applications of augmented reality in medicine remain limited, research in this area is quite significant (see, e.g., [5–7]). An example of AR already being tested in the operating room consists of a display that integrates multiple medical images and patient data [8]. It allows clinicians to keep their eyes on the patient, therefore reducing procedure-related complications. For instance, a surgeon could see preoperative MR images and the electrocardiogram of the patient directly in her/his field of view during a cardiac intervention. A similar example is depicted in Fig. 28.3. Augmented reality can also make it possible to visualize anatomical structures that would be otherwise invisible by merging a 3D model of a patient, created before the intervention, with images obtained at the time of the operation [9]. This is probably the application area for augmented reality with the highest potential.

Visually augmenting a surgery is a way of providing *surgical navigation*. Surgical navigation is a term that describes a clinical workflow where patient preoperative images, instrument tracking, and sometimes numerical simulation are combined into real-time spatial information that provides (visual) guidance to reach the target location during an intervention (see Fig. 28.4 for an example of surgical guidance during liver surgery). Compared to other applications of AR in the operating room, such as the ones described above, the main benefit of surgical navigation is the possibility to precisely indicate where structures of interest are located even when they cannot be directly seen during surgery.

One of the first studies on augmented reality for surgery was proposed by Fuchs et al. [3]. This work focused on the extraction of depth information in laparoscopic images in order to improve visualization in augmented reality during surgery. Still in the context of visualization, Suthau et al. [8] described the general principles that still prevail for augmented surgery applications. In 2004, in the journal *Medical Augmented Reality for Patient*, Wesarg et al. [10] describe a system of augmented reality for minimally invasive interventions, in which only rigid transformations between the pre- and intraoperative images are considered. The same year, Marescaux et al. [11] reported the first laparoscopic adrenalectomy assisted by augmented reality, based on manual alignment between the virtual model and the operating image. Similar results, in other clinical fields such as vascular surgery, have been obtained [12]. In most cases, the deformation of the anatomy is ignored or supposed to be negligible. The first approaches to augmented reality on deformable organs were made using markers or navigation systems placed near the operating field [13]. These methods have shown the feasibility of automatic augmented reality systems in surgery but generally impose constraints on the equipment of the operating room or require manual interactions (cf. Fig. 28.4).

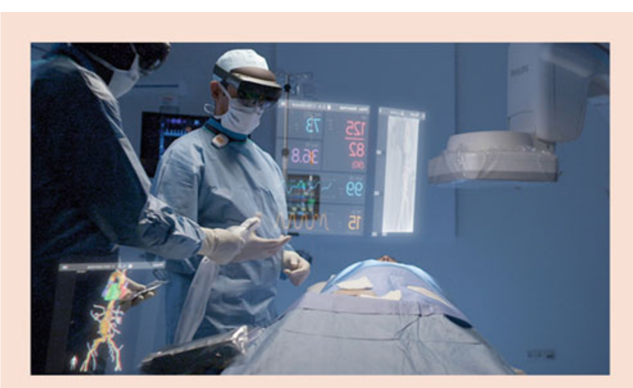


Fig. 28.3 An example of augmented reality used to display patient information on a virtual monitor aligned with the surgeon's line of sight (prototype developed by Philips in collaboration with Microsoft). Photo credit: Philips

To achieve the best possible results when augmenting interventional images, several challenges need to be addressed. A problem shared by many clinical applications is the alignment of the preoperative images (or 3D models reconstructed from these images) with the patient. This can be achieved relatively easily in the case of rigid structures (skull, spine, and other bones) by placing markers at the same position on the patient during preoperative imaging and during the intervention [14, 15]. Theses markers are, by design, both visible to the tracking system and effortlessly segmented from the preoperative scan. Since bone structures are rigid, they can easily be imaged with computed tomography (CT) imaging and intraoperative X-rays. Once segmented, the position of the tracker relative to the patient can be calculated, leading to a registration between the coordinates of the patient and his/her scan. It can then be used as a 3D map that will guide the surgeon. Figure 28.5 illustrates a proof of concept of augmented spine surgery where the instrument is tracked, and the 3D model of the anatomy (reconstructed from preoperative CT data) is overlaid onto the patient. Of course, accuracy in executing interventions such as in spine surgery is essential [16]. These surgeries often require screws and other hardware to be placed into vertebrae without damage to the spinal cord, nerves, and nearby blood vessels. Achieving a precise registration is therefore essential to the accuracy of the guidance system. This can be partially addressed through the development of new tracking systems; however, other solutions need to be devised to compensate for positioning errors due to organ motion and deformation.

An intermediate scenario (in terms of complexity) that fits naturally within the paradigm of surgical guidance is

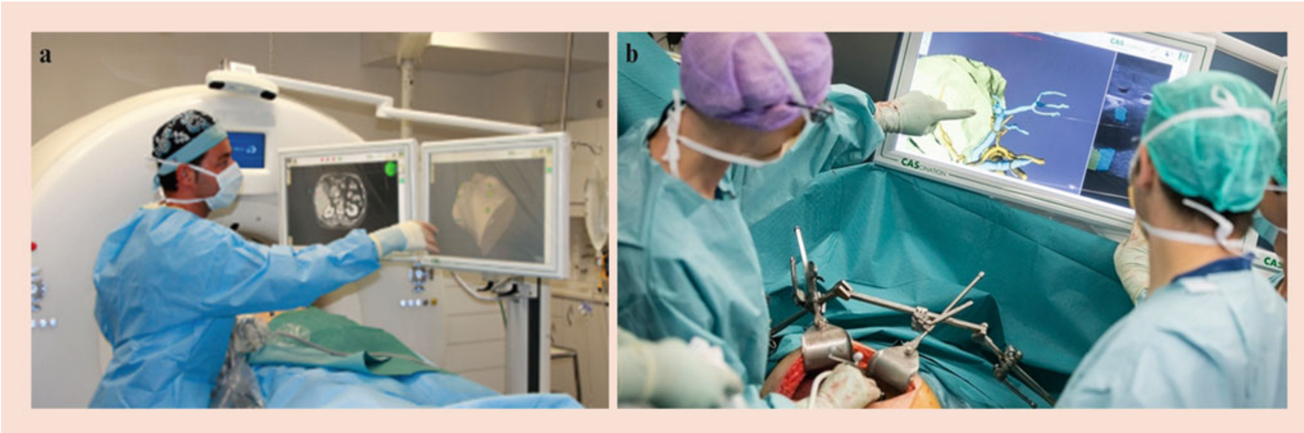


Fig. 28.4 Example of the use of a navigation system in surgery. (**a**) we can see the camera system used to track instruments (top center) in order to facilitate the repositioning of the virtual view according to the surgical view (**b**). This approach does not handle organ deformation nor the visual overlay of the virtual model onto the surgical image. Photo credit: CAScination

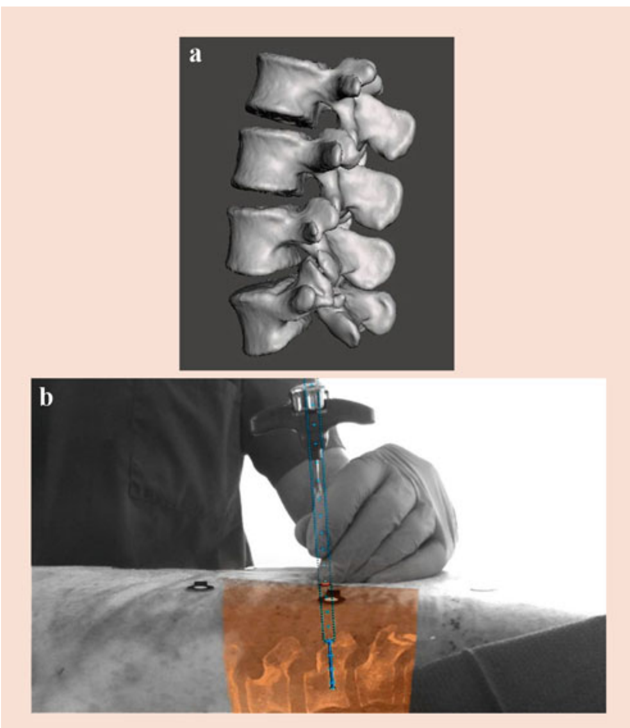


Fig. 28.5 Augmented reality in the operation theater. (**a**) 3D virtual model of the patient's vertebra reconstructed from CT data. (**b**) augmented view helping the surgeon position a surgical screw in a vertebra. Photo credit Philips

percutaneous needle placement [17]. It is one of the least invasive techniques practiced today, as the access is simply a puncture through the skin. The basic process involves the use of patient images to identify targets within the patient and planning needle trajectories, inserting the needles and verifying their placement, performing some action such as an injection or taking a biopsy sample, and assessing the results. In most needle placements, an accuracy of only a few millimeters is sought, which is not easy to achieve without guidance since the target is not directly visible. Today, this guidance is provided by visual feedback through intraoperative imaging (generally ultrasound or X-ray). This is however not always sufficient to efficiently reach the target, mainly due to poor image quality or lack of a third dimension in the visual feedback. This makes it an area where AR can be very useful.

Organ motion and deformation naturally take place under physiological conditions such as cardiac or respiratory motion, or during the medical intervention itself [18–20]. Respiratory motion can, for instance, lead to large displacement and deformation of the liver (several centimeters) as this organ sits below the diaphragm. Providing augmented reality guidance in this case requires the development of more complex solutions, which will be discussed in Sect. 28.4.

To improve the quality of guidance provided in these applications and reach a broader range of clinical targets, several research directions need to be further investigated. There is a robust and diverse research community addressing a broad range of research topics, including the creation of patient-specific models from medical images, techniques for updating these models based upon real-time image and other sensor data, and the use of these models for planning and monitoring of surgical procedures. Sensor data, images, and geometries are then linked through equations describing the physiological or biomechanical behavior of the organ. The following sections address these technical challenges, in the context of computer-assisted surgery, which is at the crossroads of these different research areas, with only one objective: to make future surgical interventions safer and more efficient.

While the literature on augmented reality for medical application is vast [21–24], in this chapter, we will describe

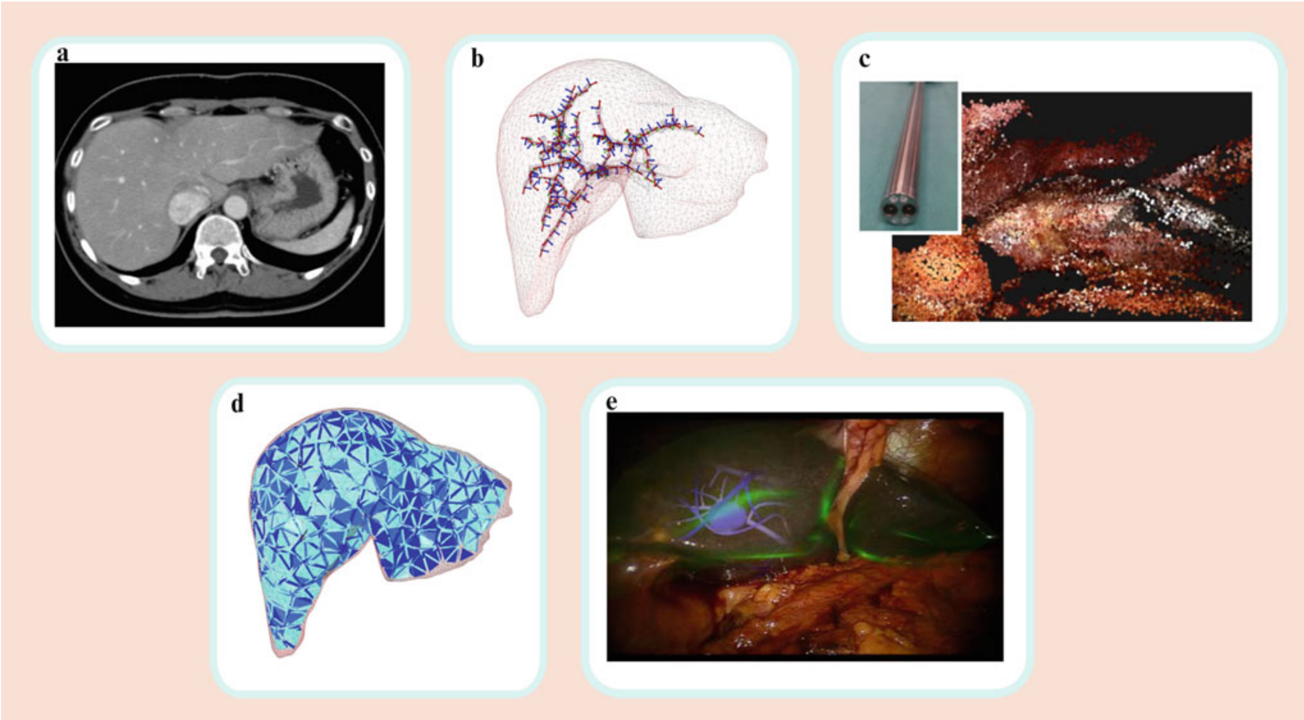


Fig. 28.6 Concepts of augmented reality for surgery: (**a**) medical imaging, (**b**) anatomical modeling, (**c**) surgical vision, (**d**) physics-based simulation, (**e**) visualization and perception

the technical concepts that are required to bring *deformable augmented reality* in the operating room. Indeed, we will focus on methods that deal with soft tissue organs, which introduces new challenges in establishing an augmented reality pipeline, from deriving models using medical image processing to performing non-rigid registration using surgical vision. The sections are as follows:

– Section 28.2 describes the imaging techniques used during surgical navigation and the procedure that transforms the resulting images to more advanced models
– Section 28.3 explains the different computer vision methods that can be used to process visual information during surgery
– Section 28.4 presents the concepts and challenges of the registration process, a core technique of many augmented reality methods
– Section 28.5 traces the different strategies to build deformable models used to represent and predict the motion and deformation of organs and soft tissues
– Section 28.6 presents a brief overview of the visualization methods used to provide a comprehensive rendering of the combined virtual and real images during augmented reality

We will present some examples of AR techniques that can assist clinicians through the fusion of intraoperative images and preoperative data (mainly images and virtual models of the anatomy). We will introduce the different concepts brought into play to achieve this objective and discuss related application domains such as the use of real-time computer simulation to train health practitioners (Sect. 28.7.1) and surgical planning as a step toward computer guidance in the operating room (Sect. 28.7.2).

28.2 Imaging and Anatomical Modeling

Surgical navigation can be partially addressed by using intraoperative imaging to monitor the operative field and visualize real-time changes to the displayed 3D model. It is possible to use intraoperative ultrasound [25, 26], CT, or MRI to update the knowledge on the anatomical shape or its relationship to a medical device. However, the amount of time required to capture and process medical images remains significant when considering volume acquisitions. Using a CT scanner [27] or its operating room version, the cone beam CT [28], exposes the patient to X-ray radiations. To limit risks for the patient, such images can, therefore, only be acquired a limited number of times. An open MRI is a feasible option for intraoperative imaging but requires longer acquisition times and the use of specialized MRI-compatible instruments and a dedicated operating theater [29]. Although imaging devices regularly improve by being faster, by being less harmful to the patient, and by providing better-quality images, their presence in the operating room remains scarce (due to their cost and large footprint in already crowded operating rooms). The exception remains for ultrasound and X-ray imaging, which stay the modalities of choice. Yet they provide only 2D images or small 3D volumes, thus limiting their role in surgical guidance.

An alternative to the options listed above is to overlay a virtual model of the anatomy, reconstructed from the preoperative image onto the patient. The quality and automation of 3D reconstructions from medical images have significantly improved in the past few years, in particular thanks to the development of deep learning [30]. It makes the generation of 3D models easier and allows for more detailed 3D virtual models of the anatomy (see Fig 28.7). Today, it is widespread that a 3D model of the patient's anatomy is computed and employed to facilitate the diagnostic and help plan the intervention [31].

If accurate 3D models of the preoperative anatomy can be obtained, intraoperative imaging can therefore be of lower resolution or dimension, or even replaced by depth sensing cameras or video cameras. However, since these images may not properly show the structures of interest, the preoperative data needs to be fused with this intraoperative image in order to augment it. This has been addressed in many application contexts. However, all currently proposed solutions [32, 33] [34] need further improvements [35].

The main challenges and therefore research directions from this point on take different names: image fusion, registration, or augmented reality, but they all correspond to very similar ideas and objectives. Image fusion corresponds to the idea of merging information from different image modalities. In our context, they consist of preoperative and intraoperative images. It is then assumed that the fused image either is directly visualized by the clinician or will be segmented and a 3D surface mesh will be created from the segmentation. Registration is the process of computing a common reference frame or a non-rigid transformation between images or an image and a 3D model. This is a step actually needed before performing any image fusion. Augmented reality can be seen as just the visualization step of the previous methods, where the registered 3D model (constructed from the preoperative image) is overlaid onto the live image of the patient in the operating room. In the context of this chapter, we will consider AR as the complete process, from image acquisition, processing of the image, registration, and visualization. This obviously calls for very efficient algorithms for each of these steps in order to achieve a smooth visualization at about 20Hz.

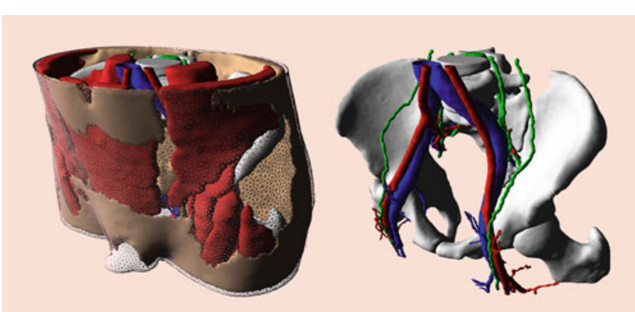


Fig. 28.7 The modeling of anatomy, as well as the creation of geometric models adapted to different computational objectives, is a key first step in numerical simulation for planning or peroperative guidance. It can include multiple anatomical structures and cover different levels of detail

28.3 Surgical Vision

In the last decade, several research groups have investigated new approaches to process and understand surgical scenes. With the introduction of optical sensors in operating rooms like cameras or laser range sensors, computer vision techniques naturally appear as an appropriate support for such a task. One of the main research topics is the estimation of organ motion, the 3D reconstruction of the organ, and its temporal visual tracking. When dealing with laparoscopic surgery, many methods have attempted to recover 3D organ shape from intraabdominal images [36], using both active and passive techniques.

With active techniques, images are enriched to facilitate their processing using projecting patterns or lights, generally through hardware modifications on optical devices or additional sensors. For instance, the *time-of-flight* technique has been tested for endoscope optics [37, 38] and measures the time that the signal takes between the organ and the camera. One of its advantages is that only a small depth range is required to build a depth map, which can be suitable knowing the restricted working space of laparoscopic surgery. Another active method used in laparoscopic surgery is *structured light* [39,40]. With this technique, a predefined pattern is projected onto the organ using a calibrated projector; the projected image is acquired with a monocular camera and processed following a decode-triangulate scheme, thus leading to the 3D reconstruction of the organ shape. This method can also be used during open surgery [41, 42], where camera placement is almost unrestricted, leading to more robust results. As an alternative to strong hardware tuning, *photometric stereo* has been [43, 44] to surgery. Based on a monocular laparoscope modified with adequate color filters, this method can reconstruct surface normals of an organ with images taken from the same viewpoint under several illumination conditions.

On the other hand, passive techniques are based on processing endoscopic images solely. Most of the proposed methods for 3D reconstruction of surgical data are inspired

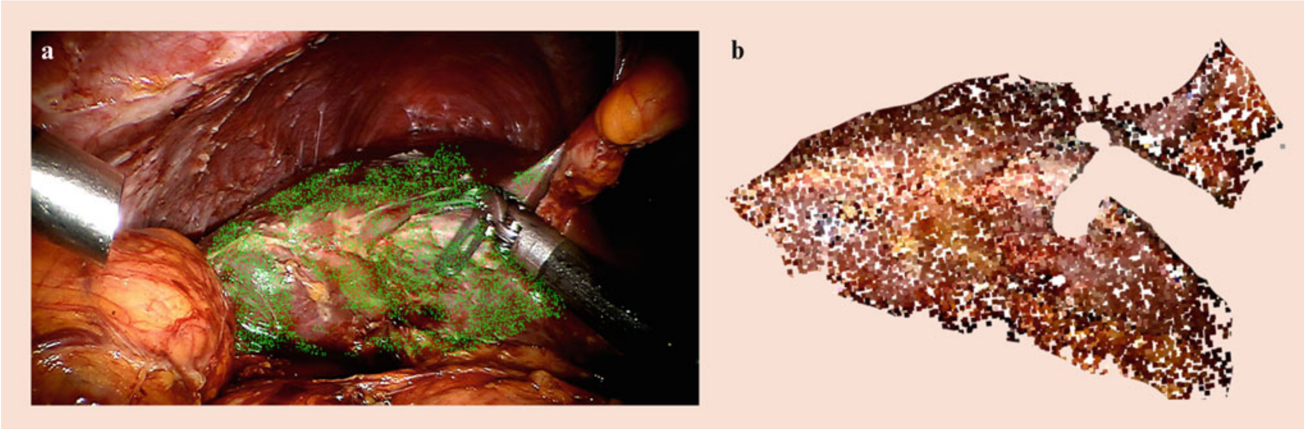


Fig. 28.8 3D reconstruction of the liver surface from a stereo-endoscopic image. (**a**) left image with extraction of points of interest (in green). (**b**) 3D partial reconstruction of the liver surface from these points of interest

by classic computer vision algorithms. However, translating these algorithms to surgical scenes is a difficult task without a specific tuning and additional visual cues.

Surgical stereoscopy has been widely studied in the last decades [9, 45–51] with very effective methods and significant results. Many of the techniques can achieve real-time performance, thanks to GPU implementations, and have been tested on *ex vivo* and *in vivo* data, making stereoscopy a promising technique for 3D reconstruction of laparoscopic scenes (see Fig. 28.8). Nevertheless, robustness regarding specular highlights, texture-less organ tissues, and occlusions remains its major limitation. Moreover, despite the fact that stereo-endoscopic devices are currently used in clinical routine, they are not as often used when compared to classical monocular laparoscopic cameras.

For this reason, several monocular-based approaches have been proposed to overcome the need for a stereo-endoscope. These methods take advantage of camera motion to simulate multiple camera or exploit shading properties of the organ like *structure from motion* [52] or *simultaneous localization and mapping* [53]. Both techniques assume a rigid environment, which is not the case when dealing with soft organs, where deformations due to respiratory or cardiac motion occur. For that purpose, they were extended to handle non-rigid behavior and predict organ deformations [54, 55].

Once the 3D shape of the organ has been recovered, most of the time, tracking this organ over time is necessary to maintain a coherent augmented reality system. Visual tracking aims to temporally locate the position of the targeted organ and to provide the evolution of tissue behavior over time (Fig. 28.9). Depending of the objective, these positions can be obtained directly from the image (and therefore are in 2D) or can be computed in 3D from the reconstruction stage.

The main contributions in this area rely on the detection of salient landmarks (or features) on the surface of the organ [57–59]. Once the landmarks are selected, an algorithm for

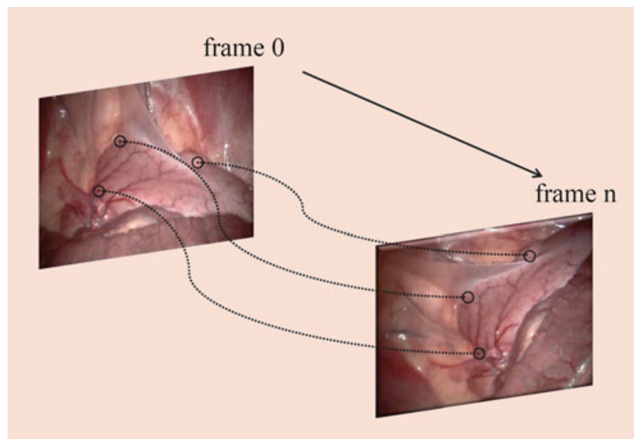


Fig. 28.9 Tracking organ's soft tissue locations in endoscopy. The aim is to be able to locate robust image features in endoscopic images through time and update the augmented reality view. Image taken from [56]

searching correspondences over frames is usually applied. These correspondences can be found using an assumption on the organ motion in two consecutive frames [46], or by searching for image similarities between the frames [60], or using optical flow constraints [47, 56, 61]. More advanced methods can rely on the use of temporal or physical models [50, 62] to predict organ's motion efficiently.

Another group of methods consist in using optical trackers and more advanced acquisition sensors (such as infrared or depth cameras). These methods are widely used in neurosurgery [21], orthopedics [23], otolaryngology [22], and spine surgery [24] where skin fiducials can be considered.

More particularly, in neurosurgery, well-established commercial systems exist to help register preoperative images to intraoperative patient coordinates so surgeons can view the locations of their surgical instruments relative to their surgical plan [63, 64]. These systems can be extended to provide an augmented surgical view using the shape of the patient's skull or brain surface to estimate the registration [65]. Marreiros et al. [66] used three near-infrared cameras to capture brain surface displacement to register the cortical surface to MRI scans using the coherent point drift method by considering vessel centerlines as strong matching features. A new deformed MRI volume is generated using a thin-plate spline model. The approach presented by Luo et al. [67] uses an optically tracked stylus to identify cortical surface vessel features; a model-based workflow is then used to estimate brain deformation correction from these features after a dural opening. Jiang et al. [68] use phase-shifted 3D measurement to capture 3D brain surface deformations. This method highlights the importance of using cortical vessels and sulci to obtain robust results. The presence of a microscope in the operating room makes it very convenient to deploy such methods clinically. Sun et al. [69] proposed to precompute brain deformations to build an atlas from a sparse set of image points extracted from the cortical brain surface. These image points are extracted using an optically tracked portable laser range scanner. Other methods used hardware-free approaches relying solely on the stereo-microscope [70, 71] to first recover brain's surface 3D shape and potentially register the preoperative models using a 3D-3D registration.

With the recent advances in deep learning, many researchers proposed to use neural networks to tackle previous surgical vision tasks [72, 73] in order to bring more robust method and more accurate results. Indeed, neural networks have led to significant improvements in classical computer vision tasks, such as scene segmentation, motion tracking, and even depth estimation from a single image. Extending these results to surgical vision is a natural continuation. However these methods are data greedy, and currently very few surgical datasets are publicly available, making this objective still difficult to achieve.

28.4 Registration and Pose Estimation

Registration is a fundamental task in surgical guidance and medical image processing. The problem of co-registering coordinate systems associated with instruments, images, and the patient, in a setup where the transformation is (or can be assumed to be) rigid, has been extensively addressed in the literature (see, e.g., [74] or [75]). Although the problem seems relatively simple, determining accurately the rigid transformation, in a clinical setup where noise, occlusions, and other phenomena can lead to uncertain and noisy data, is not trivial [76]. Yet, the most general and complex problem remains the one of estimating a non-rigid registration. Among its most important applications of deformable registration, we find multimodality fusion, where information acquired by different imaging devices is fused to facilitate diagnosis, treatment planning, and intervention. It can also be used to study anatomical changes over time and other things which are beyond the scope of this chapter.

Following [77], we briefly summarize the main concepts here. Although, in general, registration can be performed on any number of images, in our context, we often consider only two: one acquired preoperatively and one acquired during the intervention. One is usually referred to as the source (S) or moving image, while the other is referred to as the target (T) or fixed image. The two images are related by a transformation W which is generally unknown. The goal of registration is to estimate the optimal transformation W that aligns the two images according to two criteria: a measure of alignment between T and S and a regularization term which introduces some smoothness in the transformation. The measure of alignment between T and S can be computed in many ways, using various similarity metrics and distance measures. One a metric has been chosen, the images get aligned under the influence of a transformation W, that maps locations of T to corresponding locations of S. Typical features can include artificial fiducials (pins, implanted spheres, etc.) or anatomical features such as vessel branching points, ridge curves, or local surface patches. The transformation W then needs to be determined at every position in the image where no matching information is known. To this end, it generally includes a regularization term which "enforces" specific properties in the solution, such as smoothness of the displacement field. W can also directly integrate the regularization term and an *a priori* knowledge about the deformation, such as the elastic properties of the organ being registered.

When the objective is to augment the intraoperative image, two main scenarios are possible: (1) the preoperative data S is an actual image, or (2) it is a mesh generated from the preoperative image. In the first case, once the non-rigid registration is performed, the augmentation can be provided by volume rendering of the preoperative image [78] (see Fig. 28.10). If the registration process involves a mesh and an image, the general principles described above still hold. However, an appropriate set of matching features need to be found, such that they are both present in the geometrical model and intraoperative image. This is usually more challenging than when T and S are images, in particular when avoiding the use of artificial fiducials. Possible solutions to address this are presented in the following sections. Then, irregardless of the nature of T and S, two additional challenges remain. First, finding an optimal transformation W when the intraoperative data is, for instance, an image from a laparoscopic camera, an X-ray image, or an ultrasound image is an ill-posed problem for which finding W is difficult. However, considering that, in the context of AR, the registration is computed on data from the same patient, it is possible to introduce an a priori knowledge about the characteristics of the deformation which

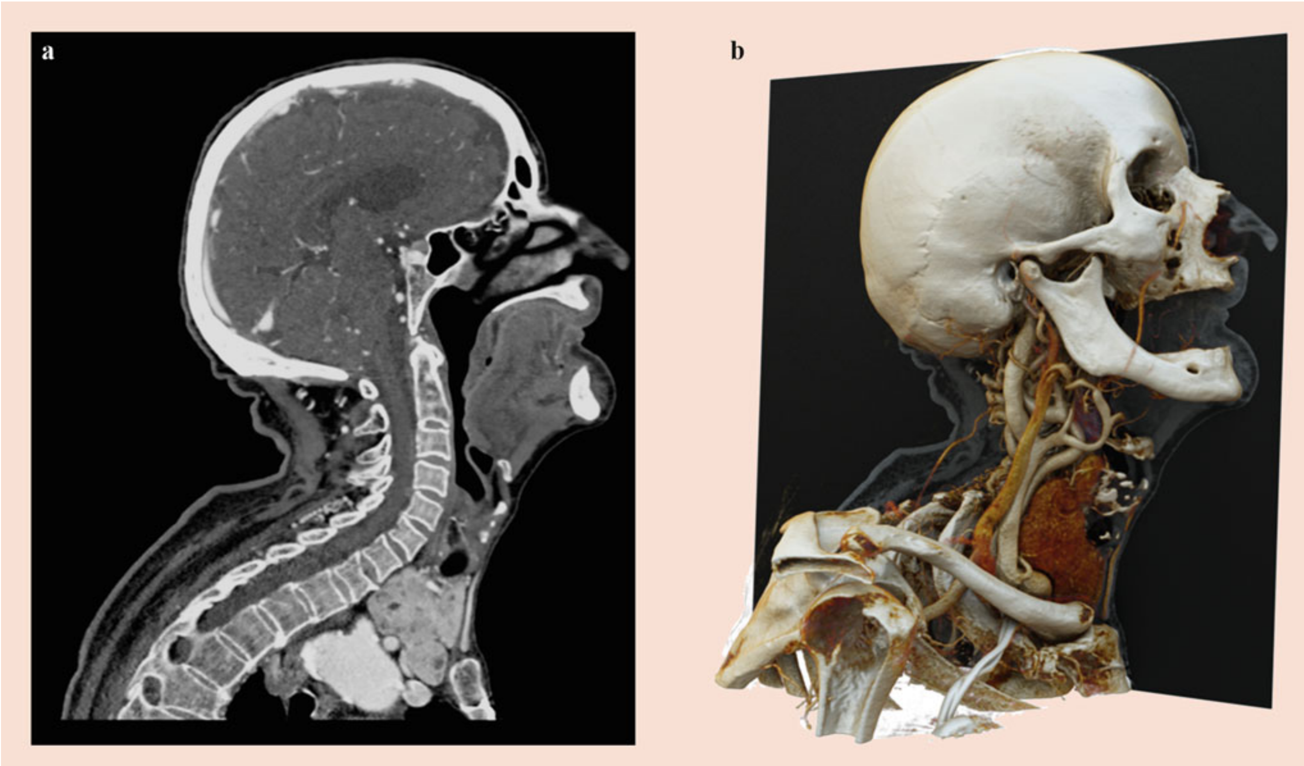


Fig. 28.10 Example of visualization based on volume rendering. (**a**) Original computed tomography (CT) data. (**b**) Cinematic rendering of the same dataset. Data courtesy of Radiologie im Israelitischen Krankenhaus, Hamburg, Germany

can limit the number of possible solutions. This is often done by using a biomechanical model of the organ.

28.5 Physics-Based Modeling

Registration algorithms based on biomechanics principles have become popular in recent years [6, 20, 79–82]. They allow for explainable and accurate computation of the deformation between two images or an image and a 3D shape. However, due to the complexity of the governing equations of the underlying model, these methods are generally slow to compute. While this is acceptable for image registration, this is not sufficient for augmented reality applications, which require real-time updates of the 3D model according to the intraoperative image. Computing soft tissue deformation in real time, with an appropriate level of accuracy, is a challenging task, which is discussed in this section.

28.5.1 Soft Tissue Biomechanics

Soft anatomical structures such as the liver, heart, brain, and blood vessels are involved in many medical interventions. Developing augmented reality solutions for these soft organs is very challenging: tracking their deformation and motion cannot be solved with tracking devices and methods used for rigid bodies. It requires to model these structures, not only from an anatomical point of view but also from a biomechanical point of view. This step is very important to provide patient-specific simulations and aim for a personalized medicine. In addition to simulating the complex deformation of these structures, it is necessary to significantly decrease computation times to provide a smooth, interactive, augmented view. It is generally considered that the computation must take less than 50ms to ensure at least a 20Hz visual update. This is an important challenge given the complexity of the biomechanical models that need to be developed to ensure the highest possible accuracy in predicting the updated shape.

A preliminary step generally consists in generating a three-dimensional model of the anatomy. Since most interventions are specific to a single organ, this modeling is limited to a few anatomical structures. If we consider the liver, for instance, this involves the parenchyma, the veins and arteries inside the organ, and some potential tumors. Figures 28.7 and 28.18 illustrate such representations of a patient's anatomy. These models are generally built from a medical image of the patient used initially for diagnostic. These geometric models will then serve to augment the surgical view but are also the starting point for creating the biophysical model of the organ. For this reason, the

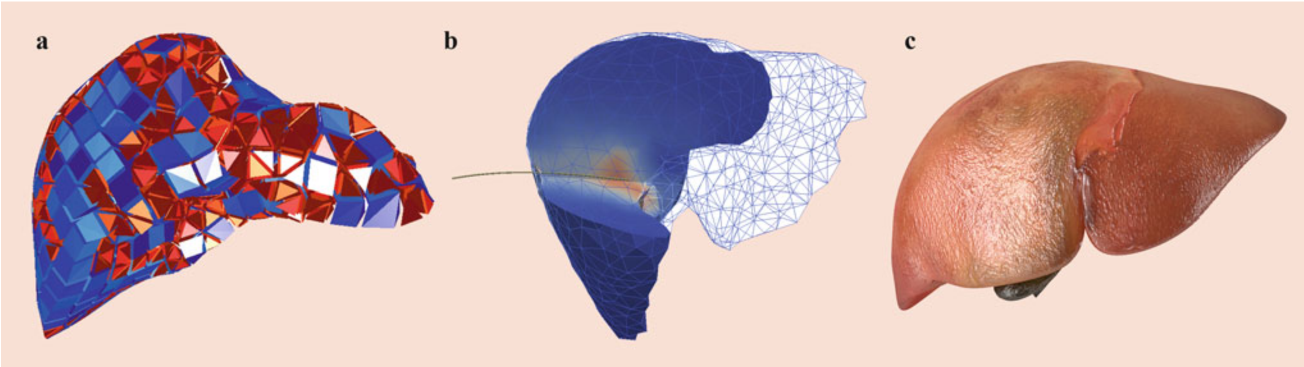


Fig. 28.11 (**a**) finite element mesh of the liver composed of tetrahedra and hexahedra. (**b**) simulation of the interaction between a flexible needle and the liver, requiring the computation of both the deformation of the organ and its interactions with the instrument. (**c**) visual model of the liver, with a realistic rendering using textures and different shaders

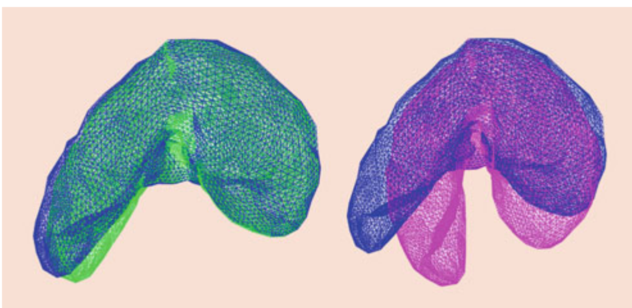


Fig. 28.12 Comparison of three hyperelastic models of deformation: (Blue) hyperelastic model, (Pink) visco-hyperelastic model, (Green) poro-hyperelastic model. All models are subject to the same external force. We can see how strongly the model choice influences the prediction. Image taken from [83]

characteristics which one seeks to obtain in these meshes are not only related to visualization. Geometric and topological properties are also important. For instance, "smoothness" of the surface is usually sought since most anatomical structures have this property. It is also important to ensure that meshes are closed when defining volumes. This will make it possible to manage on the one hand the contacts between virtual objects (if there is a hole in the mesh, one can pass through the object without detecting collision) and on the other hand to create volume meshes, which will be used as support for the computation of the biophysics simulation (see Fig. 28.11 for an illustration of these different characteristics).

Once the geometry has been defined, the next step consists in modeling the deformation of the organ as it may occur during surgery. This is most often based on an approach inspired by the laws of physics but with varying degrees of approximation compared to the real behavior. In the case of soft tissue, it is accepted that the general behavior is *hyperelastic*. A hyperelastic material is a material which tolerates large reversible elastic deformations and for which the force-deformation relationship is nonlinear. Depending on the case, different approximations can be made to simplify the computation. When the deformations remain small, as can be the case in neurosurgery, the deformation model can be simplified by a linear elastic law, much faster to compute. There exist several constitutive models for soft tissues, representing the actual behavior in more or less complex and accurate ways. Figure 28.12 shows how the choice of the model can influence the end result. This choice also depends on the application context. The brain, for instance, remains a hyperelastic "material," and the assumption of linearity is valid only for certain interventions where the applied forces are small and therefore compatible with the linearization of the constitutive law.

Once the constitutive law that is most suited to the problem has been defined, it remains to compute the motion and deformation of the set of virtual models involved in the augmented surgery. The elasticity equations (linear or not) involve partial derivatives and are too complex to have an analytical solution except for very simple geometries. For shapes as complex as an organ, it is necessary to use a numerical approximation. For this, the finite element method [84] is favored by many researchers due to its well-established numerical properties, even if alternative solutions exist. Indeed the finite element method guarantees a convergence of the approximate solution toward the exact solution if the mesh of the domain is sufficiently fine. This mesh, in our case, is generally a volume mesh (cf. Fig. 28.11), composed of simple geometrical elements such as tetrahedra or hexahedra. Creating such a mesh requires specialized software but also a good understanding of the influence of the mesh on the result. The type and number of these elements influences both the precision and the speed of the calculations, as illustrated in Fig. 28.13. Twenty years ago, performing real-time simulations with a finite element method was not possible due to the heaviness of the computation. With the development of new numerical approaches, this has become possible in certain cases, in particular if it is acceptable that the computation

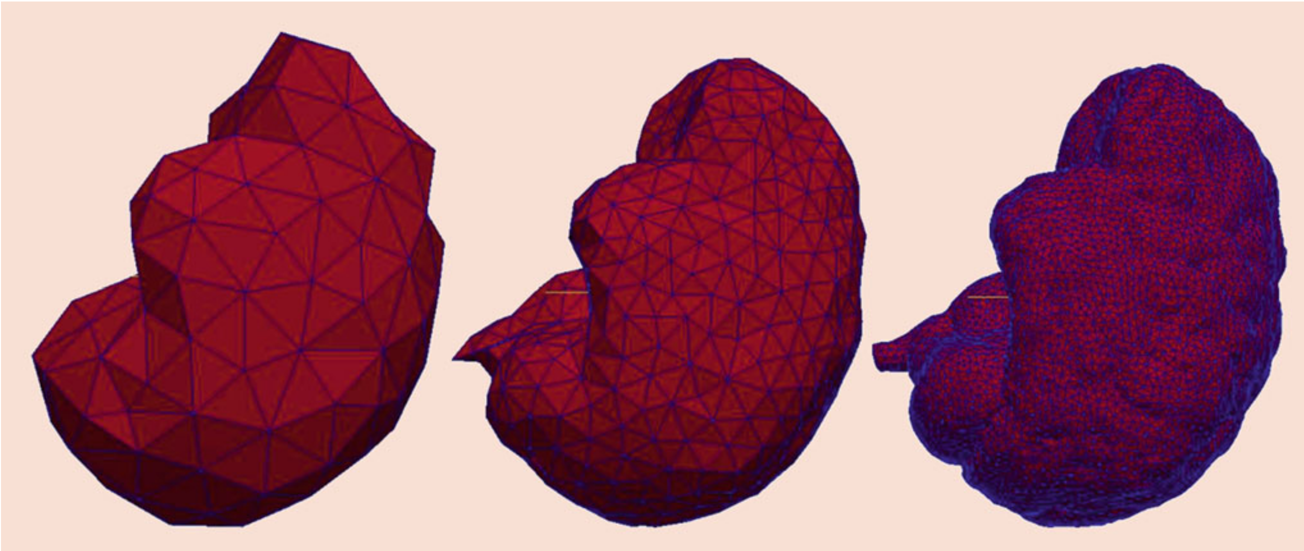


Fig. 28.13 Meshes of the brain with different resolutions

of the deformation remains an approximation of the exact solution [85, 86]. To achieve visual update rates adapted to AR while achieving sufficient accuracy (as required for surgical guidance), different strategies can be developed. We review the main ones in the following section.

28.5.2 Strategies for Real-Time Computation

Among the numerous publications in the field of biomechanics, real-time deformable models, and haptics, many methods have been proposed to address the requirements listed above. We can cite, for instance, methods based on spring-mass networks [87, 88] or their derivatives [89], based on linear elasticity [85], using explicit time integration schemes for nonlinear materials [86], and several others. Here we only describe strategies related to the finite element method.

Generally speaking, the finite element method, for an elastic body, leads to solving (in the static case) the following system of equations: $K(u)u = f$, where $K(u)$ is the stiffness matrix of size $3N \times 3N$ with N the number of nodes of the mesh (in 3D). $K(u)$ depends on the current displacement $u = x - x_0$ computed between the rest shape x_0 and the current shape x. A first mean of reducing computation time is to simplify the constitutive law. If linearity is assumed (i.e., small strain and small displacement), then the stiffness matrix of the system can be efficiently solved [85, 90], since K in this case is considered constant. This allows to precompute responses and obtain a significant speedup. Since the small strain hypothesis leads to incorrect results when rotations or large deformation are involved, Felippa [91] introduced the co-rotational method which accounts for large displacements (but small strains). This allows for more realistic soft tissue behavior [92] and fast simulations for reasonable mesh complexity (see Fig. 28.13).

If higher frame rates are needed, or more complex deformation models, it is also possible to obtain a significant gain by parallelizing the computation, on the CPU or GPU. This can lead to speedups up to 50× [86, 93]. In [94], the authors use an explicit time integration method, which is well-suited for a GPU implementation. Such an approach has been demonstrated for the non-rigid registration of a 3D brain model during neurosurgery.

More recently, the use of deep learning techniques has also proven to be a very efficient alternative [95]. This data-driven method based on a U-Net architecture approximates the nonlinear relation between external forces and the displacement field. Thanks to its ability to encode highly nonlinear models in a compact form, this neural network can perform very fast simulations on various geometries and resolutions with small errors when compared to a full FEM computation. The use of machine learning methods also opens new possibilities for creating patient-specific simulations, where not only the shape of the organ but also material properties can be taken into account.

28.5.3 Beyond Surgery

Surgery, in the classical sense, relies on mechanical interactions to treat patients. Soft tissues are manipulated and dissected to remove the tumor, leading to deformations of the organ that the methods described above will attempt to model and simulate. However, there exist several new, minimally invasive therapies that aim at treating patients using non-mechanical approaches. Augmented reality can also

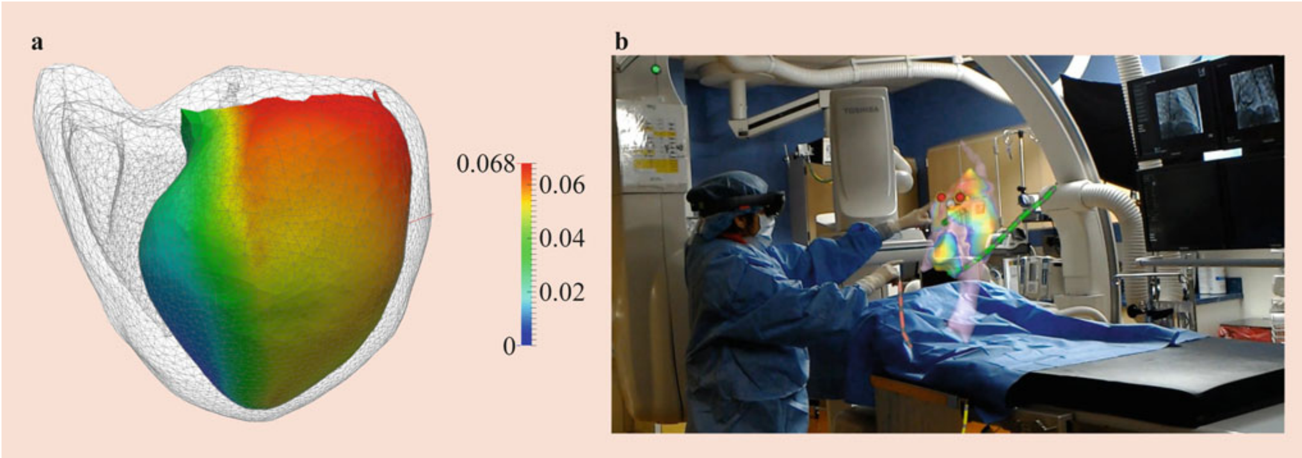


Fig. 28.14 (**a**) simulation of the electrophysiological activity of the heart. (**b**) visualization of the cardiac information in the operating room. Photo credit: SentiAR Inc.

provide significant assistance in these new interventions by displaying information that cannot be visualized directly on the anatomy. For instance, many of these minimally invasive therapies rely on the use of intense heat or cold to destroy cancerous cells (e.g., radiofrequency ablation or cryotherapy). However, the effect of the medical device cannot be visualized during the intervention, only the final outcome, thus making the monitoring and control of these procedures very complex. By visualizing computer-generated heat maps or physiological processes (see Fig. 28.14) overlaid onto the actual organ, it becomes easier for the clinician to control the therapy [96].

28.6 Visualization and Perception

Registering a 3D model onto an image in a robust and accurate manner will produce a new image, where important areas are enhanced such as risk map, or hidden parts are made visible such as internal tumors or vessels. Visualizing those augmented parts in a way that surgeons and clinicians can understand them easily is an important part of the complete AR pipeline [97, 98]. One of the most challenging process in AR visualization in surgery is the improvement depth perception when overlaying a surgical scene with complex organ or anatomy models [99]. Most of the time, these models appear to be "floating" above the actual organ or anatomy. To solve this issue, many strategies have been proposed and evaluated by research groups. Stereopsis has been used extensively to improve depth perception. In this method, 3D glasses are used to replicated the two-eye effect [100, 101]. Other methods rely on ChromaDepth to update the rendering so it fits the light spectrum [102, 103]. Using a color-coded scheme, close objects are represented in red, whereas far objects are represented in blue. Edges are considered strong features to enhance AR rendering [104, 105]. Several methods are based on curating edges to provide a rendering that can outline the region of interest while blending it with its surrounding tissue or anatomy. Motion has also been used as a depth cue, since moving around an object gives strong structural information to surgeons [106]. This technique is often merged with previous rendering approaches in order to give users the freedom of "exploring" the scene and improve their perception of depth (Fig. 28.15).

Combining rendering techniques is common when dealing with complex surgical scenes. In [105], a combination of adaptive alpha blending and contour rendering was used to effectively perceive the vessels and tumors' depth during hepatic surgery. In their method, Lerotic et al. [107] used a pq-space-based non-photorealistic rendering. This method rendered the surface of the anatomy in a translucent manner while enhancing navigation depth cues. This former work was successfully used in AR systems for lobectomy, heart disease [108], and nephrectomy [109].

Providing surgeons with a comprehensive augmented image and an understandable visualization is an important aspect of any AR system in surgery. Many existing methods in computer graphics and human-machine interaction have been used and evaluated for surgical purposes [110]. Despite that, it is still a little studied problem where no real ground truth about how surgeons perceive and understand these rendering exists, where the heterogeneity of users and the diversity of applications makes it difficult to find the best way to evaluate each approach in a quantitative, objective way.

28.7 Related Topics

Learning and training is an ongoing process that does not end with medical school; it can take several years for a medical

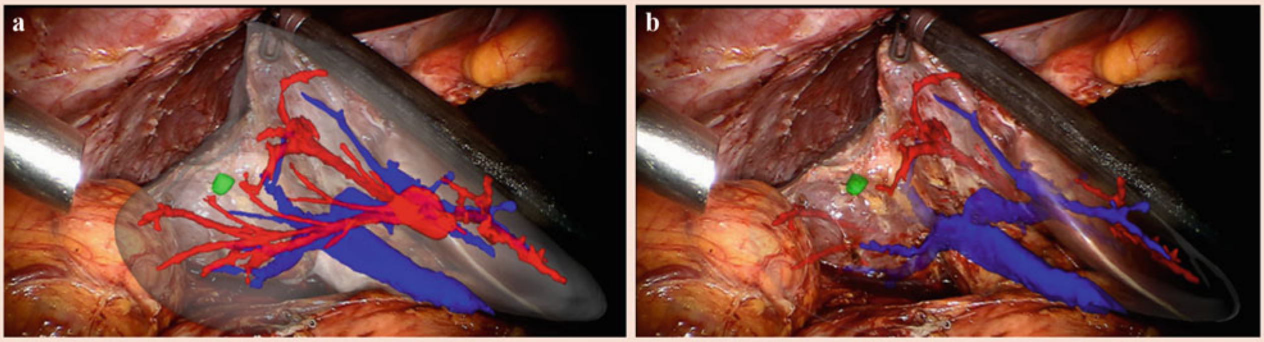


Fig. 28.15 Visualization approach using contour rendering and adaptive alpha blending. (**a**) 3D model overlaid on a endoscopic image without improved depth perception. (**b**) The vessels are rendered w.r.t the depth of the organ to better perceive the vascular network depth. Image taken from [105]

practitioner to become a domain expert. In addition to being a surgical assisting tool, augmented reality can be seen as knowledge transfer technology that can help scale expertise to less-experienced personnel: for example, to capture MRI procedural information from experienced technicians and scale it to junior-level personnel. A senior-level physician can also interact with other medical personnel that are geographically dispersed. Using real-time annotation techniques and surgical planning can also help less-experienced doctors identify patterns and build a diagnosis. We will describe in this section how AR-related technology allows for surgical planning, knowledge transfer, and training.

28.7.1 Training and Knowledge Transfer

In medical education, the traditional paradigm for surgical technical training is often summarized as "see one, do one, teach one." This way of training can be effective in open surgery since the surgical gestures can be observed directly. The trainee sees the surgeon's hands and instruments' motion and follows the organ manipulation. Following this paradigm in endoscopic surgery is made more difficult since surgical tools and motions are inside the body and only visible through a monitor. Moreover, endoscopic surgery requires different skills than open surgery, mainly spatial orientation and navigation. These skills are hard to transfer and need extensive training. Surgical simulation systems have been introduced as a direct response to these considerations. They are useful to mimic different kinds of procedures with varying degrees of complexity and realism for endoscopic and other minimally invasive procedures. They have nowadays achieved widespread acceptance in the field of flexible endoscopy, anesthesia, interventional radiology, intensive care, surgery, and other fields. Many teaching hospitals have extensive simulation training centers, and the use of simulators is so common that working groups have been set up in order to evaluate these training systems based on shared guidelines (Fig. 28.16).

In what follows, we briefly present some examples of the use of computer-based simulation in the context of learning in surgery. The objective is not to cover extensively the body of work and products existing in this field but rather to introduce some general concepts and their close relationship with augmented reality. In general, computer-based simulations are intended for minimally invasive surgeries such as laparoscopy. These new approaches offer many benefits for patients, such as reduced risk of infection and hemorrhage and reduced hospital and recovery times. However, the reduced operating field, since viewed via an endoscopic camera, and the absence of tactile perception during such interventions require dedicated training. Fortunately, this surgical technique offers features that have facilitated the development of virtual reality and simulation tools. The fact of not directly manipulating the organs, nor of directly visualizing the operating field (cf. Fig. 28.2), makes it possible to develop devices which today faithfully reproduce what the surgeon perceives.

These same concepts extend to the field of microsurgery or even vascular surgery. In the first case, the operating field is most often viewed via a stereoscopic microscope, and the instruments are sometimes similar to those used in laparoscopic surgery. With regard to vascular surgery, also called interventional radiology, the anatomy is visualized through an X-ray imaging system, and the intervention is performed using flexible instruments navigating to the area of interest via the arterial or venous system (see Fig. 28.17). Using this technique, vascular surgeons perform interventions on the arteries (e.g., stenosis or aneurysm) or even treat pathologies directly accessible via the vascular network (e.g., heart valves).

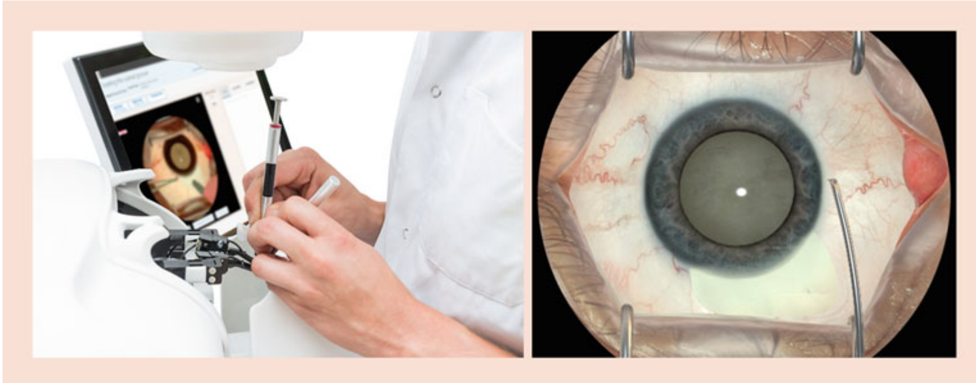


Fig. 28.16 Microsurgery is also an area of application suitable for the development of learning simulations. Here a simulation of cataract surgery and its force feedback system. Photo credit: HelpMeSee

Fig. 28.17 Examples of interactions between virtual models of organs and instruments. (**a**) Simulation of navigation of a catheter in vascular surgery (Photo credit: Mentice). (**b**) Simulation of an incision in laparoscopic surgery. (**c**) Simulation of suturing in laparoscopic surgery. Photo credit: 3D systems (LAP Mentor). In these three cases, the interactions are complex and, for the first and the last examples, involve other deformable structures outside the organ itself

28.7.2 Surgical Planning

In some cases, planning, an operation is an essential step to the success of surgery. In the case of hepatectomy (the surgical procedure which allows the removal of a part of the liver for the treatment of hepatic tumors), for instance, this planning will make it possible to maximize the volume of liver remaining after surgery to increase the patient's chances of survival. In other cases, this planning will lead to a reduction in the duration of the intervention, also reducing hospitalization time.

Today, medical images remain the main basis for intervention planning. They are studied first by radiologists in order to establish the diagnosis and then by surgeons. However, in some cases, it is difficult to define the best strategy on the basis of these images alone, or at least on the basis of their native visualization (see Fig. 28.1). This is why these images are most often processed by different software, allowing them to be better viewed and manipulated in 3D. The most common way to view these medical images in 3D is to use volume rendering techniques. This approach is widely available on the workstations of radiology departments and may be sufficient to provide a good 3D visualization of anatomical and pathological structures. However, some more advanced calculations and manipulations are not possible using this technique. In many cases, it is necessary to calculate the volume of tumors, and in order to plan a hepatectomy, it is important to determine the volume of the liver after the operation, as this is a key criterion for the success of an intervention. For this, each anatomical and pathological structure of the medical image must be segmented and reconstructed. The resulting 3D models (arteries, veins, nerves, tumors, etc.) can then be viewed and manipulated individually, offering a solution more suitable for use in surgical planning (see Fig. 28.18). There is a large number of software available today to perform these manipulations such as Myrian (Intrasense, Montpellier France), MeVisLab (MeVis Medical Solution, Germany), or Scout Liver (Pathfinder Therapeutics, USA).

By offering surgeons the ability to manipulate this digital clone of the patient, they can plan the intervention and even rehearse precisely the gestures to be performed during the operation. When realistic simulation environments are available, clinicians can also acquire and maintain technical and cognitive skills. These skills can then be transferred to the operating room, where they can contribute to safer

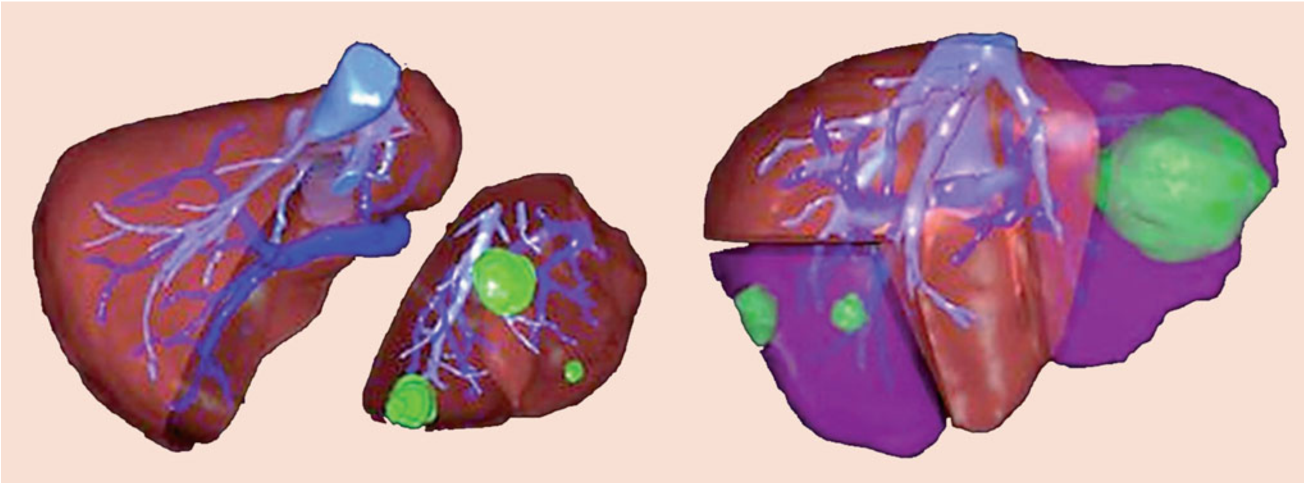


Fig. 28.18 3D patient-specific reconstructions of the liver's anatomy. The model includes the parenchyma (in red), the veins (in blue), and tumors (in green). The regions of the liver comprising the tumors are delimited in order to estimate the volume of the remaining liver, an essential criterion for postoperative survival. This is used for planning the intervention. Photo credit: IRCAD and Visible Patient

interventions. Similarly, the computational models involved for skills training can also be applied to augmented reality in surgery. Altogether, these technologies have the potential to reduce errors in medicine.

28.8 The Future of AR in Medicine

Thanks to the technical advances in numerical simulation, computational biomechanics, imaging, computer vision, and augmented reality displays, we see today a new revolution taking place in the medical field. These advances expand the use of medical imaging from a diagnostic tool only to a way assisting surgeons during complex interventions. Augmented reality is gradually moving away from research labs and entering the operating rooms. With a reduction in surgical risk and shorter hospital stays, these new surgical techniques will likely form the basis of future surgical techniques. However, this remains an emerging technology, requiring a great deal of validation and experimentation. Research in this field is also far from being over, particularly in terms of robustness of the algorithms and their accuracy but also regarding interaction techniques that could facilitate the use of these new tools in the operating theater.

This evolution of interventional medicine echoes what we observed 20 years ago with the arrival of computer science dedicated to medical image processing. Through information processing, modeling, visualization, and easier manipulation of complex concepts, augmented reality widens the field of possibilities in surgical guidance. This also facilitates the development of new technologies and their use by clinicians. Among the different directions these developments are taking place, the field of medical robotics is certainly a direct match. Augmented reality can be naturally combined with robotics to facilitate the control of master-slave systems, in particular when the robot has to target structures inside deformable organs, such as inserting a needle into a tumor.

In all cases, computer-assisted surgery, as its name suggests, aims to help surgeons in their actions and their decisions but at no time to replace them. They remain the ones who make the final decisions and operate on the patient.

Acknowledgments We would like to thanks Phillips, Siemens, CAScination, HelpMeSee, IRCAD, Visible Patient, SentiAR, and InSimo for making their illustrations available. The rendering in Fig. 28.10 is produced with Siemens Healthineers research tools that are not commercially available. Future availability cannot be guaranteed.

References

1. Stephen, W., Eubanks, M.: Mastery of Endoscopic and Laparoscopic Surgery, 2nd edn. Lippincott Williams and Wilkins, Philadelphia (2004)
2. Lemmon, G.W.: Vascular surgery: principles and practice. J. Vasc. Surg. **67**(3), 988 (2018). https://doi.org/10.1016/j.jvs.2017.10.008
3. Fuchs, H., Livingston, M., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J., Rademacher, P., Drake, S., Meyer, A.: Augmented reality visualization for laparoscopic surgery. In: Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 934 – 943. Springer, Berlin
4. Marvik, R., Lango, T., Tangen, G., Lindseth, F., Yavuz, Y., Hernes, T.N.: Image-guided laparoscopic surgery, review and current status. Minerva Chirurgica **60**, 305–325 (2005)
5. Fischer, J., Eichlera, M., Bartza, D., Straßer W.: A hybrid tracking method for surgical augmented reality. Comput. Graph. **31**(1), 39–52 (2007)
6. Haouchine, N., Dequidt, J., Peterlik, I., Kerrien, E., Berger, M.-O., Cotin, S.: Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. In: International Symposium on Mixed and Augmented Reality (ISMAR), pp. 199–208. IEEE, Piscataway (2013)

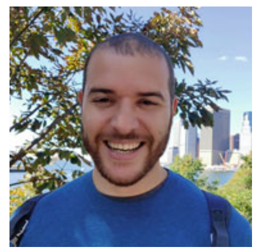
7. Leizea, I., Alvarez, H., Aguinaga, I., Borro, D.: Real-time deformation, registration and tracking of solids based on physical simulation. In: International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, Piscataway (2014)
8. Suthau, T., Vetter, M., Hassenpflug, P., Meinzer, H.-P., Hellwich, O.: A concept work for augmented reality visualisation based on a medical application in liver surgery. Int. Arch. Photogramm. Remote Sensing Spatial Inform. Sci. **34**(5), 274–280 (2002)
9. Su, L.-M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D.: Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-CT to stereoscopic video registration. Urology **73**(4), 896–900 (2009)
10. Wesarg, S., Schwald, B., Seibert, H., Zogal, P., Schnaider, M., Sakas, G.: An augmented reality system for supporting minimally invasive interventions. In: Workshop Augmented Environments for Medical Imaging, pp. 41–48. Springer, Berlin (2004)
11. Marescaux, J.: Augmented reality assisted laparoscopic adrenalectomy. J. Am. Med. Assoc. **292**(18), 2214–2215 (2004)
12. Anxionnat, R., Berger, M.-O., Kerrien, E.: Time to go augmented in vascular interventional neuroradiology? In: Linte, C., Chen, E., Berger, M.-O., Moore, J., Holmes, D. (eds.) Augmented Environments for Computer-Assisted Interventions, vol. 7815 of Lecture Notes in Computer Sciences, pp. 3–8. Springer, Berlin (2013)
13. Teber, D., Guven, S., Simpfendorfer, T., Baumhauer, M., Guven, E.O., Yencilek, F., Gozen, A.S., Rassweiler, J.: Augmented reality: a new tool to improve surgical accuracy during laparoscopic partial nephrectomy? Preliminary in vitro and in vivo results. Eur. Urol. **56**(2), 332–338 (2009)
14. Teber, D., Guven, S., Simpfendorfer, T., Baumhauer, M., Guven, E.O., Yencilek, F., Gozen, A.S., Rassweiler, J.: Augmented reality: a new tool to improve surgical accuracy during laparoscopic partial nephrectomy? preliminary in vitro and in vivo results. Eur. Urol. **56**(2), 332–338 (2009)
15. Kong, S.-H., Haouchine, N., Soares, R., Klymchenko, A., Andreiuk, B., Marques, B., Shabat, G., Piechaud, T., Diana, M., Cotin, S., Marescaux, J.: Robust augmented reality registration method for localization of solid organs' tumors using CT-derived virtual biomechanical model and fluorescent fiducials. Surg. Endosc. **31**, 2863–2871 (2017). https://doi.org/10.1007/s00464-016-5297-8
16. Wang, M.-L., Wu, J.-R.: Real-time advanced spinal surgery via visible patient model and augmented reality system. Comput. Methods Programs Biomed. **113**(3), 869–881 (2014)
17. Grasso, R., Faiella, E., Luppi, G., Schena, E., Giurazza, F., Vescovo, R., D'Agostino, F., Cazzato, R., Beomonte Zobel, B.: Percutaneous lung biopsy: Comparison between an augmented reality CT navigation system and standard CT-guided technique. Int. J. Comput. Assisted Radiol. Surg. **8**(5), 837–848 (2013). https://doi.org/10.1007/s11548-013-0816-8
18. Haouchine, N., Dequidt, J., Marie Odile, B., Cotin, S.: Deformation-based augmented reality for hepatic surgery. Stud. Health Technol. Inform. **184**, 182–188 (2013). https://doi.org/10.3233/978-1-61499-209-7-182
19. Stoyanov, D.: Stereoscopic Scene Flow for Robotic Assisted Minimally Invasive Surgery, pp. 479–486. Springer, Berlin (2012)
20. Suwelack, S., Talbot, H., Röhl, S., Dillmann, R., Speidel, S.: A biomechanical liver model for intraoperative soft tissue registration. In: Proceedings of the SPIE International Conference, vol. 7964, pp. 79642I–79642I–6. SPIE (2011)
21. Meola, A., Cutolo, F., Carbone, M., Cagnazzo, F., Ferrari, M., Ferrari, V.: Augmented reality in neurosurgery: a systematic review. Neurosurg. Rev. **40**(4), 537–548 (2017). https://doi.org/10.1007/s10143-016-0732-9
22. Wong, K., Yee, H.M., Xavier, B.A., Grillone, G.A. (2018). Applications of augmented reality in otolaryngology: a systematic review. Otolaryngol. Head Neck Surg. **159**(6), 956–967. https://doi.org/10.1177/0194599818796476
23. Jud, L., Fotouhi, J., Andronic, O., Aichmair, A., Osgood, G., Navab, N., Farshad, M.: Applicability of augmented reality in orthopedic surgery—a systematic review. BMC Musculoskelet Disord. **21**, 103 (2020). https://doi.org/10.1186/s12891-020-3110-2
24. Burström, G., Persson, O., Edström, E., Elmi-Terander, A.: Augmented reality navigation in spine surgery: a systematic review. Acta Neurochir. **163**, 843–852 (2021)
25. Kang, X., Azizian, M., Wilson, E., Wu, K., Martin, A.D., Kane, T.D., Peters, C.A., Cleary, K., Shekhar, R.: Stereoscopic augmented reality for laparoscopic surgery. Surg. Endosc. **28**(7), 2227–2235 (2014). https://doi.org/10.1007/s00464-014-3433-x
26. Gonzalez, S., Guo, Y., Lee, M.: Feasibility of augmented reality glasses for real-time, 3-dimensional (3d) intraoperative guidance. J. Am. Coll. Surgeons **219**(3), 64 (2014)
27. Shekhar, R., Dandekar, O., Bhat, V., Philip, M., Lei, P., Godinez, C., Sutton, E., George, I., Kavic, S., Mezrich, R., Park, A.: Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography. Surg. Endosc. **24**(8), 1976–1985 (2010). https://doi.org/10.1007/s00464-010-0890-8
28. Kenngott, H.G., Wagner, M., Gondan, M., Nickel, F., Nolden, M., Fetzer, A., Weitz, J., Fischer, L., Speidel, S., Meinzer, H.-P., Böckler, D., Büchler, M. W., Müller-Stich, B.P.: Real-time image guidance in laparoscopic liver surgery: first clinical experience with a guidance system based on intraoperative CT imaging. Surg. Endosc. **28**(3), 933–940 (2014). https://doi.org/10.1007/s00464-013-3249-0
29. Tsutsumi, N., Tomikawa, M., Uemura, M., Akahoshi, T., Nagao, Y., Konishi, K., Ieiri, S., Hong, J., Maehara, Y., Hashizume, M.: Image-guided laparoscopic surgery in an open mri operating theater. Surg. Endosc. **27**(6), 2178–2184 (2013). https://doi.org/10.1007/s00464-012-2737-y
30. Zhou, T., Ruan, S., Canu, S.: A review: deep learning for medical image segmentation using multi-modality fusion. Array **3–4**, 100004 (2019)
31. Knoops, P., Papaioannou, A., Borghi, A., Breakey, R., Wilson, A., Jeelani, O., Zafeiriou, S., Steinbacher, D., Padwa, D., Dunaway, D., Schievano, S.: A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. Nat. Sci. Rep. **9**, 13597 (2019)
32. Pessaux, P., Diana, M., Soler, L., Piardi, T., Mutter, D., Marescaux, J.: Robotic duodenopancreatectomy assisted with augmented reality and real-time fluorescence guidance. Surg. Endosc. **28**, 2493–2498 (2014)
33. Kersten-Oertel, M., Gerard, I., Drouin, S., Mok, K., Sirhan, D., Sinclair, D.S., Collins, D.L.: Augmented reality in neurovascular surgery: feasibility and first uses in the operating room. Int. J. Comput. Assisted Radiol. Surg. **10**(11), 1823–1836 (2015)
34. Onda, S., Okamoto, T., Kanehira, M., Suzuki, F., Ito, R., Fujioka, S., Suzuki, N., Hattori, A., Yanaga, K.: Identification of inferior pancreaticoduodenal artery during pancreaticoduodenectomy using augmented reality-based navigation system. J. Hepato-Biliary-Pancreatic Sci. **21**(4), 281–287 (2014)
35. Vavra, P., Roman, J., Zonca, P., Ihnat, P., Nemec, M., Jayant, K., Habib, N., El-Gendi, A.: Recent development of augmented reality in surgery: a review. J. Healthcare Eng. 2017:4574172 (2017)
36. Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., Stoyanov, D.: Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. Med. Image Anal. **17**(8), 974–996 (2013)
37. Penne, J., Holler, K., Sturmer, M., Schrauder, T., Schneider, A., Engelbrecht, R., Feussner, H., Schmauss, B., Hornegger, J.:

Time-of-flight 3-d endoscopy. In: Medical Image Computing and Computer-Assisted Intervention, vol. 5761 of Lecture Notes in Computer Science, pp. 467–474 (2009)
38. Groch, A., Haase, S., Wagner, M., Kilgus, T., Kenngott, H., Schlemmer, H., Hornegger, J., Meinzer, H., Maier-Hein, L.: A probabilistic approach to fusion of time-of-flight and multiple view based 3d surface reconstruction for laparoscopic interventions. In: International Conference on Computer Assisted Radiology and Surgery, pp. 397–398. Springer, Berlin (2012)
39. Maurice, X., Graebling, P., Doignon, C.: Real-time structured light coding for adaptive patterns. J. Real-Time Image Process. **8**, 169–178 (2011). https://doi.org/10.1007/s11554-011-0200-2
40. Albitar, C., Graebling, P., Doignon, C.: Robust structured light coding for 3d reconstruction. In: IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, pp. 1–6. Springer, Berlin (2007)
41. Golse, N., Petit, A., Lewin, M., Vibert, E., Cotin, S.: Augmented reality during open liver surgery using a markerless non-rigid registration system. J. Gastrointest. Surg. **25**(3), 662–671 (2020). https://doi.org/10.1007/s11605-020-04519-4
42. Adagolodjo, Y., Golse, N., Eric, V., De Mathelin, M., Cotin, S., Courtecuisse, H.: Marker-based registration for large deformations—application to open liver surgery (2018). https://doi.org/10.1109/ICRA.2018.8462909
43. Collins, T., Bartoli, A.: 3d reconstruction in laparoscopy with close-range photometric stereo. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2012, vol. 7511 of Lecture Notes in Computer Science, pp. 634–642. Springer, Berlin (2012)
44. Herrera, S., Malti, A., Morel, O., Bartoli, A.: Shape-from-polarization in laparoscopy. In: 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI), pp. 1412–1415 (2013). https://doi.org/10.1109/ISBI.2013.6556798
45. Stoyanov, D., Mylonas, G.P., Deligianni, F., Darzi, A., Yang, G.Z.: Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In: Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 139–146. Springer, Berlin (2005)
46. Stoyanov, D.: Stereoscopic scene flow for robotic assisted minimally invasive surgery. In: Proceedings of the 15th international conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI2012, pp. 479–486. Springer, Berlin (2012)
47. Richa, R., Bó, A. P. L., Poignet, P.: Robust 3d visual tracking for robotic-assisted cardiac interventions. In: Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention, MICCAI2010, pp. 267–274. Springer, Berlin (2010)
48. Chang, P.-L., Stoyanov, D., Davison, A., Edwards, P.E.: Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery. In: Medical Image Computing and Computer-Assisted Intervention, vol. 8149, pp. 42–49. Springer, Berlin (2013)
49. Rohl, S., Bodenstedt, S., Suwelack, S., Kenngott, H., Muller-Stich, B.P., Dillmann, R., Speidel, S.: Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. Med. Phys. **39**(3), 1632–1645 (2012)
50. Pratt, P., Stoyanov, D., Visentini-Scarzanella, M., Yang, G.-Z.: Dynamic guidance for robotic surgery using image-constrained biomechanical models. In: Medical Image Computing and Computer-Assisted Intervention, vol. 6361, pp. 77–85 (2010)
51. Haouchine, N., Cotin, S., Peterlik, I., Dequidt, J., Sanz-Lopez, M., Kerrien, E., Berger, M.-O.: Impact of soft tissue heterogeneity on augmented reality for liver surgery. IEEE Trans. Visualization Comput. Graph. **21**(5), 584–597 (2015)
52. Hu, M., Penney, G., Edwards, P., Figl, M., Hawkes, D.: 3d reconstruction of internal organ surfaces for minimal invasive surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007, vol. 4791 of Lecture Notes in Computer Science, pp. 68–77 (2007)
53. Mountney, P., Stoyanov, D., Davison, A., Yang, G.-Z.: Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006, vol. 4190 of Lecture Notes in Computer Science, pp. 347–354. Springer, Berlin (2006)
54. Collins, T., Compte, B., Bartoli, A.: Deformable shape-from-motion in laparoscopy using a rigid sliding window. In: Medical Image Understanding and Analysis Conference, pp. 173–178. Springer, Berlin (2011)
55. Mountney, P., Yang, G.-Z.: Motion compensated slam for image guided surgery. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010, vol. 6362 of Lecture Notes in Computer Science, pp. 496–504. Springer, Berlin (2010)
56. Yip, M.C., Lowe, D.G., Salcudean, S.E., Rohling, R.N., Nguan, C.Y.: Tissue tracking and registration for image-guided surgery. IEEE Trans. Med. Imaging **31**(11), 2169–2182 (2012)
57. Mountney, P., Stoyanov, D., Yang, G.-Z.: Three-dimensional tissue deformation recovery and tracking. IEEE Signal Process. Mag. **27**(4), 14–24 (2010). https://doi.org/10.1109/MSP.2010.936728
58. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. Int. J. Comput. Vis. **65**(1–2), 43–72 (2005)
59. Giannarou, S., Visentini-Scarzanella, M., Yang, G.-Z.: Probabilistic tracking of affine-invariant anisotropic regions. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 130–143 (2013). https://doi.org/10.1109/TPAMI.2012.81
60. Puerto, G., Mariottini, G.-L.: A comparative study of correspondence-search algorithms in mis images. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2012, vol. 7511 of Lecture Notes in Computer Science, pp. 625–633. Springer, Berlin (2012)
61. Elhawary, H., Popovic, A.: Robust feature tracking on the beating heart for a robotic-guided endoscope. Int. J. Med. Robot. **7**(4), 459–468 (2011)
62. Richa, R., Poignet, P., Liu, C.: Three-dimensional motion tracking for beating heart surgery using a thin-plate spline deformable model. Int. J. Rob. Res. **29**(2–3), 218–230 (2010)
63. Fraser, J.F., Schwartz, T.H., Kaplitt, M.G.: BrainLab Image Guided System, pp. 567–581. Springer, Berlin (2009)
64. Bucholz, R., McDurmont, L.: The History, Current Status, and Future of the StealthStation Treatment Guidance System, pp. 543–565. Springer, Berlin (2009)
65. Kersten-Oertel, M., Gerard, I., Drouin, S., Mok, K., Sirhan, D., Sinclair, D., Collins, D.: Augmented reality in neurovascular surgery: feasibility and first uses in the operating room. Int. J. Comput. Assist Radiol. Surg. **10**(11), 1823–1836 (2015). https://doi.org/10.1007/s11548-015-1163-8
66. Marreiros, F.M.M., Rossitti, S., Wang, C., Smedby, Ö.: Non-rigid deformation pipeline for compensation of superficial brain shift. In: MICCAI 2013, pp. 141–148. Springer, Berlin (2013)
67. Luo, M., Larson, P.S., Martin, A.J., Konrad, P.E., Miga, M.I.: An integrated multi-physics finite element modeling framework for deep brain stimulation: Preliminary study on impact of brain shift on neuronal pathways. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention—MICCAI 2019, pp. 682–690. Springer, Cham (2019)
68. Jiang, J., Nakajima, Y., Sohma, Y., Saito, T., Kin, T., Oyama, H., Saito, N.: Marker-less tracking of brain surface deformations by non-rigid registration integrating surface and vessel/sulci features. Int. J. Comput. Assisted Radiol. Surg. **11**(9), 1687–701 (2016). https://doi.org/10.1007/s11548-016-1358-7

69. Sun, K., Pheiffer, T., Simpson, A., Weis, J., Thompson, R., Miga, M.: Near real-time computer assisted surgery for brain shift correction using biomechanical models. IEEE J. Transl. Eng. Health Med. **2**, 1–13 (2014). https://doi.org/10.1109/JTEHM.2014.2327628
70. Haouchine, N., Juvekar, P., Wells III, W.M., Cotin, S., Golby, A., Frisken, S.: Deformation aware augmented reality for craniotomy using 3d/2d non-rigid registration of cortical vessels. In: Medical Image Computing and Computer Assisted Intervention—MICCAI 2020, pp. 735–744. Springer, Cham (2020)
71. Ji, S., Fan, X., Roberts, D.W., Hartov, A., Paulsen, K.D.: Cortical surface shift estimation using stereovision and optical flow motion tracking via projection image registration. Med. Image Anal. **18**(7), 1169–1183 (2014). https://doi.org/10.1016/j.media.2014.07.001
72. Mahmood, F., Durr, N.J.: Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy. Med. Image Anal. **48**, 230–243 (2018)
73. Lin, J., Clancy, N.T., Qi, J., Hu, Y., Tatla, T., Stoyanov, D., Maier-Hein, L., Elson, D.S.: Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks. Med. Image Anal. **48**:162–176 (2018)
74. Maintz, J.A., Viergever, M.A.: A survey of medical image registration. Med. Image Anal. **2**(1), 1–36 (1998)
75. Mani, V., Rivazhagan, S.: Survey of medical image registration. J. Biomed. Eng. Technol. **1**(2), 8–25 (2013)
76. Altantsetseg, E., Khorloo, O., Konno, K.: Rigid registration of noisy point clouds based on higher-dimensional error metrics. Visual Comput. **34**(6), 1021–1030 (2018)
77. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. IEEE Trans. Med. Imaging **32**(7), 1153–1190 (2013). https://doi.org/10.1109/TMI.2013.2265603
78. Comaniciu, D., Engel, K., Georgescu, B., Mansi, T.: Shaping the future through innovations: from medical imaging to precision medicine. Med. Image Anal. **33**, 19–26 (2016)
79. Plantefève, R., Peterlik, I., Haouchine, N., Cotin, S.: Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery. Ann. Biomed. Eng. **44**(1), 139–153 (2016)
80. Clements, L., Cash, D., Chapman, W., Galloway, R.J., Miga, M.: Robust surface registration using salient anatomical features in image-guided liver surgery. In: Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display, vol. 6141, pp. 105–115. International Society for Optics and Photonics, SPIE (2006). https://doi.org/10.1117/12.654209
81. Sun, K., Pheiffer, T., Simpson, A., Weis, J., Thompson, R., Miga, M.: Near real-time computer assisted surgery for brain shift correction using biomechanical models. IEEE J. Transl. Eng. Health Med. **2**, 1–13 (2014). https://doi.org/10.1109/JTEHM.2014.2327628
82. Haouchine, N., Juvekar, P., Wells III, W.M., Cotin, S., Golby, A., Frisken, S.: Deformation aware augmented reality for craniotomy using 3d/2d non-rigid registration of cortical vessels. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (eds.) Medical Image Computing and Computer Assisted Intervention—MICCAI 2020, pp. 735–744. Springer, Cham (2020)
83. Marchesseau, S., Chatelin, S., Delingette, H.: Non linear Biomechanical Model of the Liver: Hyperelastic Constitutive Laws for Finite Element Modeling, chapter 10, vol. 1, pp. 1–21. Elsevier, Amsterdam (2017)
84. Zienkiewicz, O., Taylor, R.: The Finite Element Method for Solid and Structural Mechanics. Elsevier, Amsterdam (2013)
85. Cotin, S., Delingette, H., Ayache, N.: Real-time elastic deformations of soft tissues for surgery simulation. IEEE Trans. Visualization Comput. Graph. **5**(1), 62–73 (1999)
86. Comas, O., Taylor, Z. A., Allard, J., Ourselin, S., Cotin, S., Passenger, J.: Efficient nonlinear fem for soft tissue modelling and its GPU implementation within the open source framework sofa. In: Proc. of the International Symposium on Biomedical Simulation, pp. 28–39. Springer, Berlin (2008)
87. Sorensen, T., Mosegaard, J.: An introduction to gpu accelerated surgical simulation. In: Proceedings of the Third international conference on Biomedical Simulation, vol. 4072 of Lecture Notes in Computer Science, pp. 122–129. Springer, Berlin (2006)
88. Delingette, H.: Triangular springs for modeling nonlinear membranes. IEEE Trans. Visualization Comput. Graph. **14**(2), 329–341 (2008)
89. Faure, X., Zara, F., Jaillet, F., Moreau, J.: An implicit tensor-mass solver on the GPU for soft bodies simulation. In: Proc. 9th Workshop on Virtual Reality Interactions and Physical Simulations, pp. 1–10. Eurographics Association (2012)
90. James, D.L., Pai, D.K.: Artdefo: accurate real time deformable objects. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, pp. 65–72. ACM, Berlin (1999)
91. Felippa, C., Haugen, B.: A unified formulation of small-strain corotational finite elements: I. theory. Comput. Methods Appl. Mech. Eng. **194**(21), 2285–2335 (2005)
92. Peterlik, I., Duriez, C., Cotin, S.: Modeling and real-time simulation of a vascularized liver tissue. In: Proc. Medical Image Computing and Computer-Assisted Intervention, pp. 50–57. Springer, Berlin (2012)
93. Courtecuisse, H., Jung, H., Allard, J., Duriez, C., Lee, D.Y., Cotin, S.: GPU-based real-time soft tissue deformation with cutting and haptic feedback. Progr. Biophys. Mol. Biol. **103**(2), 159–168 (2010)
94. Miller, K., Joldes, G., Lance, D., Wittek, A.: Total lagrangian explicit dynamics finite element algorithm for computing soft tissue deformation. Commun. Numer. Methods Eng. **23**(2), 121–134 (2007)
95. Mendizabal, A., Marquez-Neila, P., Cotin, S.: Simulation of hyperelastic materials in real-time using deep learning. Med. Image Anal. **59**, 1–11 (2020)
96. Silva, J.N., Southworth, M., Raptis, C., Silva, J.: Emerging applications of virtual reality in cardiovascular medicine. JACC Basic Transl. Sci. **3**(3), 420–430 (2018)
97. Jannin, P., Kersten-Oertel, M., Collins, D.: DVV: a taxonomy for mixed reality visualization in image guided surgery. IEEE Trans. Visualization Comput. Graph. **18**(2), 332–352 (2012)
98. Heinrich, F., Bornemann, K., Lawonn, K., Hansen, C.: Depth perception in projective augmented reality: an evaluation of advanced visualization techniques. In: 25th ACM Symposium on Virtual Reality Software and Technology, VRST'19. Association for Computing Machinery, New York (2019)
99. Kruijff, E., Swan, J.E., Feiner, S.: Perceptual issues in augmented reality revisited. In: 2010 IEEE International Symposium on Mixed and Augmented Reality, pp. 3–12 (2010)
100. Kersten, M., Stewart, J., Troje, N., Ellis, R.: Enhancing depth perception in translucent volumes. IEEE Trans. Visualization Comput. Graph. **12**(5), 1117–1124 (2006)
101. Tendick, F., Bhoyrul, S., Way, L.: Comparison of laparoscopic imaging systems and conditions using a knot-tying task. Comput. Aided Surg. **2**, 24–33 (1997)

102. Bailey, M.J., Clark, D.: Using chromadepth to obtain inexpensive single-image stereovision for scientific visualization. J. Graph. GPU Game Tools **3**, 1–9 (1998)
103. Joshi, A., Qian, X., Dione, D., Bulsara, K., Breuer, C., Sinusas, A., Papademetris, X.: Effective visualization of complex vascular structures using a non-parametric vessel detection method. IEEE Trans. Visualization Comput. Graph. **14**(6), 1603–1610 (2008)
104. Interrante, V., Fuchs, H., Pizer, S.: Enhancing transparent skin surfaces with ridge and valley lines. vol. 1995, pp. 52–59 (1995)
105. Marques, B., Haouchine, N., Plantefeve, R., Cotin, S.: Improving depth perception during surgical augmented reality. In: ACM SIGGRAPH 2015 Posters, SIGGRAPH'15. Association for Computing Machinery, New York (2015)
106. Ware, C., Mitchell, P.: Reevaluating stereo and motion cues for visualizing graphs in three dimensions. In: Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization, pp. 51–58. Association for Computing Machinery, New York (2005)
107. Lerotic, M., Chung, A., Mylonas, G., Yang, G.-Z.: PQ-space based non-photorealistic rendering for augmented reality. In: Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 4792 of Lecture Notes in Computer Science, pp. 102–109. Springer, Berlin (2007)
108. Stoyanov, D., Mylonas, G., Lerotic, M., Chung, A., Yang, G.-Z.: Intra-operative visualizations: perceptual fidelity and human factors. J. Display Technol. **4**, 491–501 (2009). https://doi.org/10.1109/JDT.2008.926497
109. Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., Yang, G.Z.: An effective visualisation and registration system for image-guided robotic partial nephrectomy. J. Robot. Surg. **6**, 23–31 (2012). https://doi.org/10.1007/s11701-011-0334-z
110. Kersten-Oertel, M., Chen, S., Collins, L.: An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. IEEE Trans. Visualization Comput. Graph. **20**, 391–403 (2014). https://doi.org/10.1109/TVCG.2013.240

Stephane Cotin is a Research Director at Inria and Leader of the Mimesis team. His main research interests are in physics-based simulation, real-time simulation of soft tissue deformations, finite element modeling, augmented reality, and medical applications of this research. Before joining Inria, he was Research Lead for the Medical Simulation Group at CIMIT and an Instructor at Harvard Medical School.



Nazim Haouchine received his PhD from the University of Lille (France) in 2015 and was a Postdoctoral Researcher at the Institute of Image-Guided Surgery of Strasbourg (France). He is now a Researcher at the Surgical Planning Lab (Harvard Medical School/BWH). His research focuses on image-guided simulation, augmented reality, and 3D shape recovery of deformable objects with an application to surgery.

28

Youth and Augmented Reality

29

Elana R. Kriegel, Bojan Lazarevic, Deborah S. Feifer, Christian E. Athanasian, Nelson Chow, Jamie P. Sklar, Yaa O. Asante, Clara S. Goldman, and Ruth L. Milanaik

Contents

29.1 Introduction ... 710

29.2 Augmented Reality in Video Games ... 710
29.2.1 Brief History of Augmented Reality Video Games (ARGs) ... 711
29.2.2 Modern ARGs Utilize Smartphone Technology ... 711
29.2.3 Benefits of Augmented Reality Video Games ... 712
29.2.4 Downsides to Augmented Reality Games' Growing Popularity ... 712
29.2.5 Safety Precautions When Using Augmented Reality Games ... 713
29.2.6 The Future of Augmented Reality Games ... 714

29.3 Augmented Reality and Social Media ... 714
29.3.1 Snapchat ... 714
29.3.2 Facebook, Messenger, and Instagram ... 715
29.3.3 TikTok ... 715
29.3.4 Up and Coming AR: Pinterest ... 716
29.3.5 Precautions and Risks of AR Use on Social Media Among Children and Adolescents ... 716
29.3.6 The Future of AR on Social Media ... 716

29.4 Augmented Reality in Sports ... 717
29.4.1 Enhancing Athletic Training ... 717
29.4.2 Spectator Interactions ... 718
29.4.3 Limitations of Augmented Reality in Sports ... 718
29.4.4 The Future of Augmented Reality in Sports ... 718

29.5 Augmented Reality in Art ... 718
29.5.1 Art Appreciation and Experience ... 719
29.5.2 Augmented Reality, Art, and Young Children ... 719
29.5.3 Limitations of Art and AR ... 719
29.5.4 Future of Art and AR ... 720

29.6 Augmented Reality in Education ... 720
29.6.1 Preschool ... 720
29.6.2 Elementary School ... 720
29.6.3 Middle School ... 721
29.6.4 High School ... 721
29.6.5 College ... 722
29.6.6 Graduate Education and Vocational Training ... 722
29.6.7 Augmented Reality and the Disabled ... 724
29.6.8 Cautions ... 724

29.7 Augmented Reality in Medical Practice ... 725
29.7.1 Educational Opportunities ... 725
29.7.2 Therapeutic Uses ... 725

29.8 Augmented Reality Safety Precautions ... 727
29.8.1 Safety Prior to AR Use ... 727
29.8.2 Safety Measures During AR Use ... 728
29.8.3 Mental Health ... 729

29.9 Considerations and Causes for Concern ... 731
29.9.1 Age Considerations for Augmented Reality ... 731
29.9.2 Gender Considerations ... 732
29.9.3 Pandemic Considerations ... 733
29.9.4 Accessibility for All ... 733

29.10 Conclusion ... 733

References ... 734

E. R. Kriegel · B. Lazarevic · D. S. Feifer · C. E. Athanasian · N. Chow · J. P. Sklar · Y. O. Asante · C. S. Goldman · R. L. Milanaik (✉)
Division of Developmental and Behavioral Pediatrics, Steven and Alexandra Cohen Children's Medical Center of New York, Lake Success, NY, USA
e-mail: ekriegel@northwell.edu; Ek634@cornell.edu; blazarevic@northwell.edu; bojanL@alumni.princeton.edu; deborah.s.feifer.21@dartmouth.edu; nelsonchow@princeton.edu; jsklar1@northwell.edu; yasante@northwell.edu; rmilanai@northwell.edu

Abstract

Augmented reality (AR) is being implemented in the lives of children and young adults and is becoming increasingly relevant in various fields. AR initially influenced popular culture through gaming by encouraging children to go outside in order to play. It then evolved for use in other fields, such as social media. AR is commonly used in social media by youth when they use lenses and filters to augment their pictures. AR is even used in arts and sports to improve children's understandings and abilities. Additionally, in education, AR can assist teachers by providing engaging opportunities for learning, such as 3D models or field trips without leaving the classroom. Finally, an important application of AR use is for children in medical settings to improve patient experiences and outcomes. The fields of gaming, social media, sports, arts, education, and medicine can increasingly incorporate AR technology as it continues to advance and improve. This chapter will discuss the current and future applications of AR in these

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_29

fields as it relates to children and young adults, as well as important safety and accessibility considerations.

Keywords

Augmented reality · Children · Adolescents · Video games · Education · Social media · Arts · Sports · Safety · Medicine

29.1 Introduction

Augmented reality (AR) utilizes technological applications to create a "mixed reality" of the real environment and the virtual world. AR users' visual perception of the real environment is augmented through the use of devices that superimpose interactive 2D or 3D computer-generated objects over images of the real world [1, 2]. Numerous aspects of youth and young adult's lives are beginning to incorporate more AR technology in everyday use, especially in the fields of gaming, social media, sports, arts, education, and medicine. However, it is important to note that research on young individuals and various applications of AR is limited. Therefore, in addition to research on AR with children and adolescent, this chapter will explore and discuss research in AR that was conducted among adults that could be relevant for younger users. Conclusions and insights presented may be based on works with adults in addition to studies conducted with children and young adults. Further research is necessary to fully understand augmented reality usage and applications among younger individuals.

The concept of AR was originally developed in the 1950s by Morton Heilig, a cinematographer, who built an AR prototype because he believed AR could enhance the experience of moviegoers. In 1966, Ivan Sutherland continued to explore the development of AR technology when he engineered the first AR head-mounted display. Following these initial innovations, the term "Augmented Reality" was coined by the aircraft company Boeing when they were using the technology in order to help workers assemble wires and cables used in aircrafts. Around this time, discussions began about the benefits of AR over VR, noting that AR requires fewer pixels and potentially less equipment [2]. AR technology has continued to evolve and improve over the decades. Now, camera systems in mobile personal devices and mobile applications (apps) are able to integrate virtual objects with the real world, making this technology more accessible and usable [2]. Because of these continued advancements, the global market for AR is expected to drastically increase in the near future. The AR market's value is estimated to increase from 3.5 billion USD in 2017 to over 198 billion USD in 2025 [3].

AR applications are beginning to influence many areas of life for children and young adults, making everyday experiences enhanced with virtual features appearing everywhere from the classroom and one's social media to doctor's offices and the tennis court. AR made its debut to popular culture in the summer of 2016 when Niantic® released the mobile augmented reality game (ARG) Pokémon GO® for iOS® and Android® products. This game encourages children to go outside, get exercise, and explore, while collecting Pokémon™ characters through the use of AR [4, 5]. The influence of AR technology has since transcended the world of entertainment and gaming, as it begins to influence daily interactions in multiple fields. On social media platforms such as Snapchat®, Instagram®, Facebook®, and Pinterest®, there are features where users can use AR filters and lenses to take pictures or play AR selfie games, and more [6]. AR can also be used in sports to help athletes better prepare and visualize game possibilities as they train or help spectators engage with events [7]. In art museums users can use AR guide systems to enhance their learning, engagement, and appreciation for art [8], while students can bring their artwork to life using AR applications [9]. In education, especially in the wake of the SARS-CoV-2, 2019 coronavirus disease pandemic (COVID-19) where most education has been transitioned online, AR has the potential to strongly influence learning. AR can be used to motivate students to explore material, help teach subjects through first-hand experiences, improve collaboration between students and teachers, foster creativity among students, and help students create their own learning pace [10]. AR is also very useful in the realm of medical education, training, and practice. AR technology can be used to help teach medical students' anatomy, understand patient interactions from different perspectives, and observe novel procedures. It can also be used in practice for laparoscopic procedures, neurosurgery, and cardiology. It is used to guide patients through procedures and to help educate them about their medical conditions [11]. Specifically for pediatric patients, AR can be used to make children feel more comfortable during office medical visits [12], in addition to its uses in the hospital setting.

The fields of entertainment, gaming, social media, sports, arts, education, and medical practice can increasingly incorporate AR technology as it continues to advance and improve. This chapter will discuss the current and future applications of AR in these fields as it relates to youth and young adults, as well as important safety and accessibility considerations.

29.2 Augmented Reality in Video Games

AR technology has many potential applications, one exciting dynamic application is augmented reality games (ARGs). ARGs are extremely interactive because they require players to make decisions in real time that impact the outcome of their game. However, because the concept of ARGs is quite new, much remains unknown about their impact on users. This is an especially important point of discussion because ARGs, such as Pokémon GO, are very popular in younger age groups [13].

29.2.1 Brief History of Augmented Reality Video Games (ARGs)

AR technology began to infiltrate the video game industry in 2000 shortly after American game developer id Software released its source code for Quake®, a first person-perspective shooter game where players explore a virtual world where they can shoot monsters, collect various items, and complete missions [14]. Researchers then modified the game's original source code to incorporate AR when Bruce Thomas of the Wearable Computer Lab took the opportunity to create ARQuake. In ARQuake, game-generated images were superimposed over the player's view of the physical world [14]. This was done by situating a computer system on a rigid backpack worn by the player, with a transparent head-mounted display placed over the player's head. Inside of the head-mounted display, a one-way mirror superimposed computer-generated images onto players' views of the real world [14]. The original Quake game was played with a keyboard and mouse, but in ARQuake, players control their movement by actually moving around in their physical space and by using a two-button input device [15]. Thomas had essentially developed the world's first ARG, and for the first time in history, gamers saw their video games meet the real world.

A major limitation with ARQuake was that the user was required to wear a backpack with computers, gyroscopes, and other equipment [15]. This issue was resolved in 2006 when AR Tennis was developed for Nokia devices which are hand-held mobile telephones that could operate on digital networks internationally [16]. AR Tennis was the first face-to-face collaborative AR game developed for mobile phones. Because mobile phones have a digital display, powerful processing power, and a camera, they are efficient ARG platforms. Further, all the technology required for ARGs are within the capabilities of a mobile phone. In AR tennis, the player's Nokia® overlays a virtual ball and net over the field of view of the camera. Players serve by hitting the "2" key on the Nokia's keypad. Then, players simply rally with one another by placing their phone to be in front of the ball. Bluetooth wireless networking was used to synchronize the ball's position [17]. These initial innovations set the stage for more complex ARGs to move into the public's view as they continued to develop on more modern platforms and became more accessible.

29.2.2 Modern ARGs Utilize Smartphone Technology

The emergence of smartphones in the 2000s allowed for a new and accessible platform on which ARGs could be played [18]. The growing popularity of smartphones such as iPhones and Android devices, among other personal mobile devices, put AR machines in a large portion of the population's pockets [19]. App stores like the Apple Store® and Google Play Store® allow people to easily download apps from a common marketplace [20, 21]. It takes less than a minute to search for, download, and play a free ARG from any app store. Many mainstream companies quickly jumped on the idea of producing ARGs for their most popular products; one such example is Microsoft's Minecraft Earth® [22]. When players first download and open Minecraft Earth, they are greeted by a map view, where their character stands on a map closely resembling their physical location. This is because Minecraft Earth renders the map according to the player's real-time location using their smartphone's GPS technology. Players can go on adventures, where game-rendered images are overlaid onto the camera's view so that players can collect resources and fight enemies in their homes, outside, or virtually anywhere else. Finally, players can go to Buildplates where they can use their resources to build virtual structures that are overlaid onto the camera's view [22].

Another AR-based app that has been successful is Jurassic World Alive, which brought the beloved reptiles of the Jurassic Park® franchise into the AR scene. In Jurassic World Alive, players are once again loaded onto a map based on their physical location constructed using GPS technology [23]. The creatures are overlaid onto the map, and players can choose to engage with dinosaurs of their choosing. The player views the dinosaur from above and controls a virtual drone that collects their DNA which they can then use to create hybrid dinosaurs in the lab. Players are also able to view dinosaurs on their team in AR mode. This overlays a computer-generated dinosaur onto the camera's view [23].

Since Thomas' innovative approach to Quake, AR has made tremendous strides in accessibility, reliability, and popularity. Perhaps the most influential ARG to date is Pokémon GO. In 2016, Pokémon GO brought AR into the mainstream media's spotlight as people all over the world went out to experience the phenomenon themselves. No company was more successful in producing an ARG than Nintendo, who brought their beloved pocket monsters, Pokémon, into reality, or more specifically, an augmented reality. Pokémon GO quickly became the most popular US mobile game ever measured by the peak number of daily users. A week after its release, Pokémon GO had around 21 million daily active users in the United States, which is approximately 6% of the United States' population [24]. Pokémon GO is especially popular among children and adolescents. In mid-2017, about

32% of users were under the age of 18 [25]. In fact, users were spending more time playing Pokémon GO than engaging with applications of social media giants like Facebook, Snapchat, Twitter®, and Instagram [26]. In Pokémon GO, players are once again loaded into a map generated based on the player's physical location. Players need to physically move around in the real world in order to encounter Pokémon which appear as animated characters. Different Pokémon can be found on the map around the player. The player's main goal is to capture different Pokémon. This can be done by clicking on Pokémon on the map. The app then generates the Pokémon and overlays it onto the phone's camera. Players can then throw Pokéballs at the Pokémon by swiping on the screen. Pokéballs are essentially spherical devices that are used to catch and contain Pokémon. Pokéballs, along with other useful in-game items, can be obtained at Pokéstops, real-world locations marked on the map that players can physically go to [27]. Between discovering Pokémon and finding Pokéstops, there is a strong motivation for players to physically explore new locations in their surrounding neighborhood and communities in order to capture as many different Pokémon as possible.

29.2.3 Benefits of Augmented Reality Video Games

The sudden success of Pokémon GO can be attributed to several factors. Firstly, Pokémon debuted in 1996 meaning that, to date, there are nearly 25 years' worth of fans who grew up with the franchise [28]. Further, Nintendo® is not only wildly popular in the United States but also on a global scale. The cute and unique Pokémon are attractive to fans of all ages and can particularly invoke feelings of nostalgia for many. There are also immediate feelings of adrenaline and thrill as players try to catch new Pokémon. A psychologist explains that the gratification that comes with capturing a Pokémon causes our brain to release hormones like dopamine, which can propagate an addiction [29]. Perhaps Pokémon's slogan "Gotta Catch 'Em All" has a particularly strong appeal as well. The idea of capturing all possible Pokémon and filling one's Pokédex gives players an ultimate goal to work towards. This may allow players to feel accomplished and fulfilled after each session.

Pokémon GO also serves as an escape for many people because it is easy to pull out a mobile phone and launch the app to play. Pokémon GO invokes feelings of detachment from bad surroundings or situations and can reduce stress for players [29]. Further, the game socially benefits players particularly by delivering a sense of belonging. Because specific locations are marked on all users' maps as a place to find in-game incentives, players can meet new friends as they explore similar locations in the real world (Fig. 29.1). This in-person interaction with others is something that does not happen in traditional video games. This is especially important for youth, when social development is of the utmost importance [30]. While this aspect of the game can foster fun new interactions, it is imperative that parents supervise and accompany their children while playing Pokémon GO to monitor the safety of these interactions with age-appropriate peers. Further, the physical aspect of the game that incentivizes players to move and go outdoors encourages users to engage in physical exercise, which also does not typically occur when playing traditional video games [4, 31]. However, it is worth noting that while studies have found applications like Pokémon GO to be effective in increasing short-term physical activity, further studies are needed to investigate long-term effects [32]. This is further complicated by the fact that many players are unaware of how long they plan to play the game [33].

Another important benefit that arises is that ARGs have a unique ability to engage children and adolescents' interest in different topics. Those who play Jurassic World Alive may be interested in learning about dinosaurs and basic biology. Those who find building in Minecraft Earth fun and rewarding may choose to pursue other similar creative endeavors. Children who find the AR component of Pokémon GO of particular interest may decide to pursue a career in computer science or technology. While some companies like Microsoft® have already created ARGs like Minecraft: Education Edition [34], the idea of using ARGs in the classroom to supplement teaching is in its infancy. ARGs may grow in popularity as a classroom resource as research has already suggested that video games are becoming more popular as supplements to traditional teaching [35]. See the education subsection of this chapter for further information.

Research looking into the benefits of ARGs have several key findings thus far. A recent survey of 405 players in the United States observed improved daily and psychosocial functions as a result of playing ARGs. These benefits seem to arise from general enjoyment of the game [36]. Additionally, ARGs have been shown to increase the probability of players' making purchases at retail shops, restaurants, online stores, and spending money on travel [4, 31, 36]. Another survey of 399 Pokémon GO players reaffirmed these benefits by finding overall increases in well-being defined by positive feelings of nostalgia, friendship formation and intensification, and physical exercise [37].

29.2.4 Downsides to Augmented Reality Games' Growing Popularity

While there are many upsides to ARGs' growing popularity, there are several negative consequences to consider. The ease with which players can access Pokémon GO means that it is

Fig. 29.1 Many Pokémon GO users gathered at a Pokéstop in Setagaya Park in Japan to catch Pokémon. (Photo courtesy of Brian Millar on Flickr. The license for this photo can be found at https://creativecommons.org/licenses/by/2.0/)

very easy for players to become apathetic to the real world by becoming overly immersed in the game [4]. Players can be prone to distractions and need to understand that the game should not interfere with work, school, and relationships with friends and/or family. Pokémon GO may cause children to become unmotivated towards school, extracurricular activities, and social interactions [4]. Additionally, players may look at and interact with their screen more as a result of Pokémon GO, which may lead to fatigue and exhaustion.

A major concern with Pokémon GO that was discussed heavily on mainstream media was that children would be in danger from strangers and predators as they are wandering around outside and meeting other unknown players at Pokéstops [4, 31]. One instance of this occurred in July 2016 when armed robbers lured victims to an isolated trap in Missouri using Pokémon Go [38]. Furthermore, players who are not paying attention to their physical surroundings because they are distracted by Pokémon GO can get into an accident [31]. Players have put themselves into harm's way by chasing Pokémon because they become inattentive to their surroundings and some players have even posed a physical danger to themselves and others by playing while driving [39].

Another potential area of concern arises from the fact that apps like Pokémon GO rely on GPS technology. Seeing as data collection and data security is a major topic of modern political discussion, questions and debates regarding the ethicality of ARGs like Pokémon GO will likely arise. For example, in 2016 a US senator questioned the developers of the game in regard to its data privacy protections [40].

29.2.5 Safety Precautions When Using Augmented Reality Games

It is important that awareness is raised among parents and caretakers regarding the potential safety concerns of ARG usage. Pediatricians can uniquely position themselves as bridges between parents, caretakers, and children by actively encouraging open dialogue. This open dialogue should aim to help parents and caretakers understand their children's ARG use and habits. In this way, parents can ensure that their children are not putting themselves in dangerous situations or suffering from a loss of motivation and engagement towards school as well as lacking important developmental activities like peer-socializing and extracurricular activities. This open dialogue should concurrently aim to inform children that their GPS locations can be tracked with ease, not only by parents, but also by predators and strangers who can use ARGs to take advantage of them [5, 31]. Parents and caretakers can either accompany their children to sites where ARGs like Pokémon GO encourage players to go, or they can track their children's phones so they can see their physical location [31, 41]. Pediatricians should ensure that patients do not become addicted to ARGs and that they do not inspire negative mental, physical, or emotional outcomes. Teachers, coaches, and caretakers can ensure that children are not accessing ARGs while in class, practices, rehearsals, etc. Teachers, parents, and caretakers should also stress the importance of not driving distracted; this includes texting, surfing the web, or playing ARGs while driving [4].

29.2.6 The Future of Augmented Reality Games

Much is unknown regarding the future of ARGs, but new and innovative developments are continuously being introduced. For example, the Microsoft's HoloLens® is a recent addition to the field of handsfree AR headsets. The HoloLens 2 has industry-leading resolution, an ergonomic design that allows for extended use, and holograms can be interacted with in ways that feel natural. Simply flipping the visor up allows the user to step out of mixed reality [42]. Although originally intended for medical purposes, it is currently being tested to be use for gaming as well [43]. Microsoft's HoloLens® will be further elaborated on in Sects. 29.6.5 and 29.7.2 of this chapter. Additionally, a new program and AR creation tool, Apple's Reality Composer® within Apple's ARkit, facilitates the development of AR gaming or other applications by allowing users, even those with no prior experience, to create interactive AR experiences [44].

New AR gaming applications are continuously being created, developed, and improved. Daily, the Apple App Store features a variety of new AR applications and games in its catalog. However, it is still worth noting the problem of implementation. Pokémon GO certainly entered the mainstream spotlight quickly, but many other AR applications have shown a slow uptake by the general public. As technology continues to improve, a nuanced understanding of the implications of ARGs needs to be adopted. On one hand, there is great potential for ARGs to become more accessible, especially among children. On the other hand, there is no guarantee that the general public will change their habits and implement ARGs into their everyday lives.

29.3 Augmented Reality and Social Media

Social media networks are experiencing a burgeoning wave of innovation in the field of AR. Across many popular social media platforms, such as Snapchat, Facebook, and Instagram, AR has brought about new ways for individuals to share, communicate, and create content. Using the cameras on their personal smart devices, social media users have the ability to enhance photos and videos of real-world environments with virtual components, blurring the line between imagination and reality in order to create an immersive experience that can then be shared with others. Businesses have also begun incorporating AR features into their advertisements on social media platforms, providing customers with unique and memorable virtual experiences with their products. Because AR tools are becoming more widely available across social media networks, it is not surprising that the number of people using AR on social media is increasing: an estimated 43.7 million people in the United States interacted with AR content on a social media platform at least once per month in 2020, which is 14.4% more people than in 2019 [45]. These data include people who utilized AR camera tools to enhance their own content, as well as individuals who used AR in the context of an advertisement. Considering that roughly 10% of parents with children younger than 12 years old report that their child uses social media [46], and 51% of teens ages 13 to 17 report using social media daily [47], individuals are exposed to AR content on social media from a young age. Here, we explore how major social media networks have incorporated AR into their platforms, and how AR will impact social media in the future.

29.3.1 Snapchat

Snapchat is an app that allows users to send "snaps," or photos and videos, to other users that expire after a set duration of time [48]. The Snapchat company pioneered AR in the realm of social media with its 2015 launch of "Lenses®," or AR effects and filters that can be applied to photos or videos [49]. Before taking a photo or video, Snapchat users can press down on their screens and select a filter from dozens of options that are classified by "Face" or "World," depending on what their camera is focusing on and what the user is trying to manipulate in their real-world environment. Some filters add virtual components to images and videos immediately when selected, while others require movement and emotion from the users in order to activate. New Lenses are added every day, and Snapchat even allows users to create their own AR effects in the app or online, making AR design more accessible to the general population [50].

To further encourage the use of AR functions and sharing of content, Snapchat created multi-player AR games, called Snappables®, in 2018. When playing a Snappables game, Snapchat users have to utilize touch, facial expressions, and motion to score points or complete tasks. For example , in the game "Mini Golf," players wear virtual visors while using a finger to hit a golf ball into a hole in as few hits as possible [51]. Other games, like "AR Bowling," use AR technology to depict a bowling lane in the player's surroundings while players use their fingers to roll a bowling ball [52]. After one user takes their turn playing a Snappables game, they can invite other users to play the next round with them or to beat their high score. By playing Snappables games, Snapchat users of all ages are encouraged to interact with AR tools and to share their AR entertainment experiences with others.

Businesses have also taken full advantage of Snapchat's AR Lenses for the purposes of advertising their products and services. For example, to promote the release of the movie *Terminator: Genisys*, the movie's marketing team developed an AR filter that overlaid metallic pieces of the Terminator's face, including the iconic, bright-red Terminator eye, onto users' faces [53]. Nike® also ran custom AR Lenses before the FIFA Women's World Cup® in 2019 to promote the US Soccer Women's National Team. During this campaign, Snapchat users were able to take photos wearing virtual soccer jerseys, as well as tap their screen to visit the Nike website in order to purchase team merchandise in preparation for the tournament [54]. By placing company-sponsored AR Lenses alongside Snapchat's standard AR filters, consumers are able to not only see brand advertisements but also interact with them and share them with their friends, resulting in higher levels of engagement with brands and their products.

Snapchat's continuous development and integration of creative AR experiences throughout their platform has fostered much of AR's current popularity and mainstream acceptance. This progress has been made possible by the company's Lens Studio®, a powerful and free desktop software for outside artists and developers to create augmented reality experiences for Snapchat's hundreds of millions of users [55]. The Lens Studio software gives developers unlimited access to powerful visual programming and machine learning tools, thus allowing them to design, test, and publish AR Lenses. Once published, these AR effects can be used and shared by Snapchat's over 170-million users around the world who engage with AR daily to communicate and play [55, 56].

29.3.2 Facebook, Messenger, and Instagram

After Snapchat's original success with AR, other social media platforms moved quickly to develop their own AR experiences. In 2017, Facebook added a camera to their interface that allows Facebook users to take photos or videos with AR effects similar to those found on Snapchat [57]. After applying an AR filter to a photo or video, the content can be posted on a user's profile or published to their Facebook story for all of their friends to see. AR effects can also be found in the camera on Facebook's Messenger app, thus enabling users to quickly and privately share AR content via in-app messaging. Since the Messenger app supports video chatting alongside text-based messaging, Facebook developed several multiplayer AR games that can be played in real-time during a video call. One example of these games is "Don't Smirk," a game where video chat participants stare at each other and attempt not to smile. If the camera detects a grin from a player, AR is used to stretch the smile into an exaggerated smirk while awarding other players a point for not smiling during the round [58]. In 2018, Messenger's AR capabilities expanded into the realm of business and marketing: companies like Nike, Sephora®, and Asus® began to develop AR features on Facebook's Messenger app that would allow customers to take a closer look at a pair of shoes, try on lipstick and eyeshadow, or check out the features of a new smartphone device before purchasing the product [59].

Instagram, a photo-sharing platform that is owned by the Facebook company, joined the AR community in 2018 by introducing AR filters to its in-app camera [60]. When opening the camera, Instagram users are immediately presented with several filters and effects that can be applied to photos or videos. After a user takes a picture or video with an AR effect, they can post it to their Instagram story or send it to a friend in a direct message. Those who view the content can see the name of the effect in the top left corner and, if they click on it, can try the effect out themselves. Instagram users can also create their own AR effects and share them on their profile pages. However, these user-created effects are not immediately available to everyone on the in-app camera. In order to support content creators and give them credit for their work, Instagram requires that users follow an AR content creator's account before gaining access to any of their designed AR effects [60].

Each month, over 600 million people utilize the AR features available on Facebook and Instagram [61]. To keep users engaged with creative content, the Facebook company has created a free AR design software, called Spark AR, that provides effect creators with tools that enable them to design new AR experiences for Facebook and Instagram users. Over 1.2 million AR effects have been developed using Spark AR by 400,000 creators located in 190 countries [61]. The creation tools available to developers on Spark AR do not require any prior coding experience to get started, and instructions and video tutorials are available, making AR design accessible to anyone who is interested.

29.3.3 TikTok

TikTok® is a popular social media platform that allows users to post videos 1 min or less in length and includes a plethora of content, such as comedy, travel, animals, cooking, dancing, beauty videos, and much more [62]. When filming a TikTok video, content creators can choose from hundreds of interactive face-changing or environment-altering AR effects to make their videos unique and creative. One of the most popular TikTok effects is the "Green Screen" effect, which allows TikTokers to change the background of their videos to a default image in the app or to any image in the user's camera roll, transporting video participants into a reality and location of their choosing [63]. When viewing TikTok videos on the app, individuals can click on the name of the AR effect in the video description in order to see other videos that also use the effect or to film a video using the effect themselves.

Recently, TikTok has entered a new field of AR design that utilizes the iPhone® 12 Pro's Light Detection and Ranging (LiDAR) technology to more accurately measure distances and map out three-dimensional spaces [64]. The first AR filter to incorporate this technology is a confetti effect that superimposes confetti falling on video participants. As the confetti falls, it is able to recognize surfaces, such as an arm or couch, and appears to stop when it hits these surfaces. If confetti is resting on an individual and the individual moves out of the frame, the LiDAR technology will recognize the change in the environment and will cause the confetti to fall to the ground [65]. TikTok revealed this new effect on its official Twitter account at the beginning of 2021, and the tweet promised that "more innovative effects" using LiDAR technology were on their way later that year [66].

29.3.4 Up and Coming AR: Pinterest

Some popular social media platforms, such as Pinterest, are only beginning to incorporate AR features on their digital platforms. Pinterest, a social networking website founded in 2010 that allows for photo-bookmarking and the sharing of ideas and pictures to inspire others, has over 442 million users every month [67]. In early 2020, the platform launched its first AR experience known as "Try On" [68]. This feature allows those looking for beauty inspiration to virtually try on different lipstick shades from major makeup brands before deciding on which products to purchase. Individuals can also save photos of themselves trying on different virtual lipstick shades if they cannot decide which they prefer in the moment. Pinterest plans to develop more AR experiences in the beauty department in the near future.

29.3.5 Precautions and Risks of AR Use on Social Media Among Children and Adolescents

Snapchat, Facebook, Messenger, Instagram, TikTok, and Pinterest are all popular mobile applications utilized daily by children and adolescents as young as nine years old [46, 69–72]. With such high youth engagement, there are certain usage considerations of which parents should be aware. As children and teenagers utilize camera filters and other AR features on these apps, they may become the unplanned targets of advertisements for products and services not appropriate for their age range. With AR advertisements becoming commonplace on social media networks, parents should have an open discussion with their teenagers about how to be responsible consumers of products, as well as cautious consumers of digital media. Because users may experience bullying or harassment from friends or followers on these apps based on the content they post, parents and teenagers should also be familiar with the reporting features and anti-harassment tools these platforms have to offer [73–75]. Alongside standard precautions about usage and bullying on social media, parents of TikTok users should speak to their children about being aware of their surroundings and making sure they are in a safe, hazard-free location before filming videos with AR filters. With LiDAR technology on smartphones being a new concept that will likely be explored by many social media platforms, parents should also remind their children to proceed with caution and learn about how new AR tools utilizing LiDAR technology function before they begin using them in their content.

A notable concern across all social media apps with AR capabilities is the presence of AR filters that change the shape and likeness of a child's face. Parents must be aware that these filters could have negative psychological implications and could even trigger feelings of body dysmorphia, or an obsessive focus on one of more perceived defects or flaws in appearance, in their children [76]. With thousands of AR filters to choose from, social media users are able to alter their appearance with ease in order and conform to unrealistic and often unattainable beauty standards. While filters that add flowers or dog ears to a photograph are clearly just fun adornments, other subtle filters that smooth out a user's skin or brighten their smile can promote a pressure to look a certain way and can result in children and adolescents to internalize this beauty standard [77]. If parents believe that their child's mental health or self-perception is being negatively impacted by AR filters that manipulate the body, they should speak with a health professional.

29.3.6 The Future of AR on Social Media

As innovation in the field of AR continues, AR technologies will only become more mainstream: it is estimated that by 2023, there will be almost 2.5 billion mobile AR users [78]. AR's resounding success across social media platforms thus far is due to its immersive quality and the ease in which AR experiences can be shared. By making the smartphone camera a platform for AR, social media companies have provided consumers with new ways of communicating and expressing themselves. However, social media companies are already looking to move beyond the smartphone. Facebook has announced that it will be developing "smart glasses" that will be released in 2021 [79]. Although this product will not have an interactive display and will not be classified as an AR device itself, it is part of a larger research endeavor known as

Project Aria – Facebook's new research project which aims to design the first generation of wearable AR devices that incorporate a 3D layer of useful information overlaid on top of the physical world [80]. Further analysis and discussion of current and future devices for AR can be found in other chapters in this handbook.

Although innovations like Project Aria, if successful, will change the landscape of human connection and communication, there are potential negative implications of the expansion of AR on social media. Since future wearable AR devices will likely include cameras and microphones, consumers will have to consider the risk that such devices could be recording everything the user is seeing and hearing in order to create the immersive experience that AR is known for. While social media and smartphone companies will likely try their best to make sure all data are secure and that privacy is not breached, it will be difficult to know who has access to this information and what is being done with it [81]. This is especially concerning for young users who will be less aware of these features that could be compromising their safety and privacy.

Expanding the immersive experiences of AR on social media sites in the near future could have negative mental health implications for younger populations, as well. Researchers have previously coined the term "Facebook depression" which is described by the Council on Communications and Media of the American Academy of Pediatrics (AAP) "as depression that develops when preteens and teens spend a great deal of time on social media sites, such as Facebook, and then begin to exhibit classic symptoms of depression" (p. 802) [82]. By combining reality with AR on social media sites, children and teenagers could develop a more severe form of "Facebook depression," and those who seek validation on social media or use it as a refuge from reality could experience further exacerbated negative effects on their mental health. This may also put children at a higher risk of social isolation, which is contrary to social media's goal of using AR to increase connectedness. Because children today are interacting with technology at a younger age than ever before, it is important for parents and healthcare providers to understand how children consume media and how their online interactions, which will increasingly be enhanced by AR, could be affecting their psychosocial development.

29.4 Augmented Reality in Sports

AR within the realm of sports and athletic performance holds tremendous potential, with an estimated 43% of athletes believing AR can improve performance [7]. AR can help athletes train and develop proper technique and skills and provides users with highly interactive experiences. Not only athletes learning skills can benefit from the highly interactive experiences that AR can offer, but coaches, referees, and spectators who play a more passive role in the events can too. Evidence for the benefits of using AR in sports has already been demonstrated in academic publications involving sports including speed skating, climbing, basketball, tennis, soccer, cycling, and martial arts, among others [83]. Furthermore, novel AR-based e-sports have even been developed, and further creation of new e-sports is anticipated in the future. In fact, certain AR games have demonstrated the potential to be an effective way for people to meet recommended daily/weekly physical activity levels [84]. By adapting AR technology's use to apply to younger users in athletics and new e-sports, AR may further help children live active and healthy lives, meet recommended daily physical activity levels, and have fun if used safely.

29.4.1 Enhancing Athletic Training

Understanding the fundamental skills and techniques that a sport requires is crucial for children to excel in their chosen activity. Without such skills, children may fail to succeed or become discouraged, possibly causing them to forgo the opportunity to continue playing their respective sport. Therefore, finding a way to make learning a sport easier for children is encouraged to promote continuous participation, enjoyment, and dedication, thus making AR applications to sports not only beneficial for children but potentially parents and coaches as well!

29

Learning how to play various sports is heavily dependent upon visual cues in order for athletes to execute proper movements [85]. AR commonly provides additional visual information that could aid young athletes in developing their skills. For example, Kelly and O'Connor [81] developed an AR-based tool to improve the timing, technique, and body posture of tennis players. After cameras set up around the tennis court capture an amateur athlete's positioning, this tool uses AR to superimpose a time aligned virtual avatar of an elite athlete's position over the amateur tennis player's position. This allows coaches and players to more easily compare their gameplay to a more experienced player for reference [86]. Similar use of AR has been demonstrated in archery, with archers comparing their form to a virtual expert [87]. AR is also being implemented in adult's and children's rock climbing to assist climbers by using a projector to display color-coded guides to plan routes and increase movement challenges if desired. The AR climbing technology also provides computer-generated feedback based on specific movements captured in one's performance to allow them to improve their climbing abilities [88]. It has even been suggested that in certain scenarios, feedback via AR may be faster and more accurate than reviewing video playback or instructor feedback. AR feedback is intended to guide and motivate motor learning, and its speed allows athletes to have more time for implementation, repetition, and evaluation of adjustments [88].

29.4.2 Spectator Interactions

While it is easy to focus on the athletes physically engaging with the sport, we must not forget how AR can apply to spectators as well, who themselves enjoy interacting with the games to better understand them. For example, while not a form of AR itself, the broadcasters for various sports games offer additional information for spectators to enjoy, which in turn improves their understanding of the game itself. By listening to broadcaster commentaries, spectators can then use this information to watch the game in a more active way than by watching the game with little explanation. An example of AR use for spectators is the Hawk-Eye® system for tennis [83]. The Hawk-Eye system is a camera-based system which is able to track the tennis ball's movement, allowing referees to make more accurate calls regarding whether a ball was in bounds or not [89]. The ball's projectile motion is displayed on a screen, and the display is angled such that viewers are able envision the motion as if they were directly above the court, similar to a birds-eye view. This makes it easier for the spectators and referees to understand whether the ball was in bounds and allows them to watch from previously inaccessible viewpoints [90]. For children, it is possible that gaining improved understanding of a sport, while acting as a spectator, may encourage them to start or continue playing that respective sport.

29.4.3 Limitations of Augmented Reality in Sports

Augmented reality has the ability to improve the field of athletics in numerous ways; however, the limitations must be considered. The equipment necessary for AR use in athletic settings may interfere with natural movement patterns and may be uncomfortable for athletes to wear. For example, basketball players wearing AR glasses might find that they interfere with gameplay. In this case, the relative costs of using AR technology in certain sports may outweigh the benefits. It is possible that use of AR in sports, in its current state, could also lead to injuries and accidents as it potentially does when used in video gaming [91]. Although AR provides additional information that may benefit the athlete, this additional information may prove to be a distraction. For example, the AR-based game Pokémon Go, though not a sport, may lead children to focus more on the AR environment as opposed to the actual surrounding environment, causing accidents. Additionally, the quick movements made during sports would require high level technology in order for information to be processed quickly enough to be useful [83]. This technology may also prove to be expensive.

While AR technology may increase the activity levels of certain children, it is also important to consider other ways to help children become more physically active. Recent studies have demonstrated that the majority of adolescents do not meet current physical activity guidelines, with technology suspected to be one of many factors [92]. Given that AR is based on technology, we should be careful of how we utilize it in athletics. While it may be intended to benefit children, it may actually have negative impacts on them by encouraging increased utilization of technology. Therefore, it is suggested that children are encouraged to obtain their necessary physical activity in a variety of ways.

29.4.4 The Future of Augmented Reality in Sports

While AR in sports has great potential, it is still very much a work in progress. The equipment necessary for AR use in athletic settings may interfere with natural movement patterns and may be uncomfortable for athletes to wear. For example, basketball players wearing AR glasses might find that they interfere with gameplay. It is possible that use of AR, in its current state, within the realm of sports could also lead to injuries and accidents as it does in video games [91]. Additionally, the quick movements made during sports would require high level technology in order for information to be processed quickly enough to be useful [83]. This technology may also prove to be expensive. However, with further development, AR may not only change the current sports played by many children, but it is possible that AR could also introduce new e-sports involving further interaction with virtual information. For instance, the game HADO® is an AR sport that has gained popularity in recent years, culminating in a yearly World Cup event in which participants can win over $100,000 [83]. In HADO, participants are essentially playing an augmented game that resembles dodgeball in which players can throw virtual projectiles with physical movements and can block these projectiles with a virtual shield [93].

Given the recency of AR, it is impossible to predict what kind of impact it will have in the future, but there are many possibilities. With physical activity being essential to the health of children, utilization of AR is encouraged for situations in which it may help the child to attain recommended physical activity levels. As this technology becomes more widespread and accessible, even more people can utilize this helpful technology in sports and activities.

29.5 Augmented Reality in Art

AR has already made headway in the realm of art, improving experiences of art viewers as well as creators. Within the context of the museum, where much art education occurs, additional information about artwork are provided by AR which may help viewers to more easily understand the pieces. AR is useful in art because, like guided tours or audio-taped tours, it can provide this important information in an engaging way. Museum goers also express positive feedback with regard to their AR-guided experiences when compared to traditional-guided tours [8]. In the physical classroom, children have enthusiastically embraced AR, which may lead to more enriching art-related experiences [9]. However, AR has drawn concerns that it may impede social interaction in these settings [8, 9]. With further research, we may soon understand how to best utilize AR in art settings in order to maximize its benefits while reducing its harms. The introduction of AR has potential to change and improve how we experience and interact with art.

29.5.1 Art Appreciation and Experience

When going to museums and exhibits, many people often do not know much about the artwork they are viewing. With little information, it may be difficult to fully understand the message and grasp the beauty associated with art. However, the development of AR presents a new way for museum goers to have a more interactive experience. The utilization of AR in these settings can take various forms, but they often provide additional visual or verbal information [94]. New technology such as Google Glass®, which is a wearable pair of AR glasses, enable the presentation of this information [95]. For example, AR equipment may be able to recognize a painting and provide a brief synopsis of the context of the painting to its side [94]. So far, the merging of art and AR has been well received by museum goers. A study was conducted to compare individuals' experiences on an audio-guided tour or nonguided tour to individuals' experiences on an AR-guided tour. On the AR-guided tours, museum goers utilized a mobile-AR guide whose lens recognized a picture and provided additional information about it. Those in the AR group demonstrated improved learning, stronger belief in the digestibility of information, spent a longer time looking at the artwork, and felt the AR guide device was interesting, creative, and entertaining [8].

29.5.2 Augmented Reality, Art, and Young Children

Although viewing artwork can be an enriching experience for children, creating art and interacting with one's own art can also be a very fulfilling experience. In a study of 4–5-year-old children in Hong Kong, the utilization of AR to enhance the children's artwork was enthusiastically received [9]. In this study, children were given templates of pictures which they could color. After the children finished coloring, using the mobile AR application ColAR Mix on an iPad with the camera pointed at the drawing, the application generated animated 3-D images of the drawings the children made. This iPad was connected to the classroom projector to allow the entire class to see the animated drawing. All children in the study liked playing with the AR objects very much [9]. Additionally, most children indicated that it was easy to use [9]. Most students particularly liked the use of the AR tool because it turned their images into 3-D images [9]. By making children more enthusiastic about artwork via AR, parents and principals believed the technology had the potential to enhance the child's learning interests as well as their self-confidence [9]. While this is only one example of an AR application being used in art, other similar popular art-related AR apps exist. Examples of these apps include Color Alive® [96], uMake [97], LightSpace [98], and World Brush [99], among others, all of which enable children to harness their inner creativity. These apps can work with provided drawing templates or can allow children to create directly on the app itself.

29.5.3 Limitations of Art and AR

While AR is able to benefit many children, there remain a variety of limitations to AR in the realm of art. With regard to trips to the art museum, concerns remain regarding how it may affect levels of social interaction [8, 95]. For example, in a study comparing AR-guided tours, audio-guided tours, and tours with no guide, the AR group demonstrated less interaction with peers than the other two groups [8]. Reduced social interaction makes it harder to know and understand the different perspectives that peers may have on a piece of art and may result in a more isolated experience [8]. Oftentimes, concerns are raised that children using AR technology may become more intrigued with the technology itself as opposed to their peers [10]. One study observed that when AR was used in learning environments for students, some students became so focused on the AR technology that they were more easily distracted from their real environment and even ignored it altogether [100]. Consequently, the new technology might distract them from the art itself, different learning activities, their peers [95], and may even lead to safety concerns [100].

Like most technology-based interactive tools, moderate use may be beneficial to young children, but extensive use may be harmful. With the continuous improvement of AR technology, it is important that children deliberately make time to interact with their peers in order to develop their social skills. Given the high prevalence of technology use among children and adolescents as is, the use of AR technology in art for these populations should proceed with caution. Another potential limitation to utilization of AR to benefit children's art education is the technological knowledge of guardians and teachers. According to this study, older teachers are less willing to use this newer technology, and they are less confident when using it than younger teachers [9]. Therefore, there may be many situations in which children using the AR technology may not even be able to experience its benefits due to improper instruction on how to use it appropriately.

29.5.4 Future of Art and AR

The use of AR within the realm of art holds much promise. By providing additional information and opportunities for interaction with artwork, AR may be able to improve an individual's understanding of art and provide richer cultural experiences. With that said, it is important that there be a balance, as too much technology may reduce the child's interaction among their peers. However, AR technology also appears to be extremely popular among children, so with more research into how to best use AR effectively, we may soon be able to fully harness its capabilities.

29.6 Augmented Reality in Education

Educators in classrooms from kindergarten through graduate school have developed new and exciting applications of AR technology as it continues to develop and progress. The immersive nature of AR engages students in the classroom in creative ways, ultimately resulting in improvements in learning outcomes [101]. AR can immerse students and enhance their educational experiences in various subjects as it can place them in any environment, from the solar system to the Egyptian pyramids, or even to the operating room. AR has demonstrated promising effects in the classroom, as it improves learning motivation among students [102]. Some research has even noted neurological impacts of AR on study participants, specifically noting increases in memory and engagement [103].

As AR's use in classrooms increases, many parents are supportive of efforts to incorporate AR into their child's education. Notably, a majority of parents believed video games could be beneficial and helpful in educational settings [104]. Consequently, AR is expected to continue to grow in its usages and applications in the educational setting, with Educause forecasting its adoption in most classrooms between 2021 and 2022 [105]. Equipped with AR technology, the possibilities for learning inside and outside of the classroom expand extensively, with applications through all levels of education. This technology in classrooms has increased relevance during remote and distance learning during the COVID-19 pandemic and beyond. This section will specifically examine AR's role in education at various stages, including preschool, elementary school, middle school, high school, college, graduate education, and vocational training. The segment will conclude with a brief discussion of limitations to the usage of AR in the classroom.

29.6.1 Preschool

Although younger children (3–5 years old) are considerably more limited in the range of technologies they can use, a number of interesting potential applications of AR use remain among this age demographic. As the primary educational focus for infants is language, potential applications of AR may focus on improving language learning. One such study has utilized AR to promote English learning among infants, incorporating images and audio for content learning as well as narrations and songs to introduce phonetics. Infants appeared both responsive and receptive to these efforts, with improved learning results after the intervention [106]. Other applications in the preschool setting might include Quiver's coloring pages: interactive coloring pages, where children can print and color various sheets and then bring their image to life using the QuiverVision® app [107].

Although AR can be productively used in the preschool setting for interactive activities and language learning, its use is limited by the age of its users. According to recommendations from the AAP, children under six years of age should limit their use of screen media to only 1 hour per day, so usage of AR in the classroom for longer than this recommended period is not advisable [108]. Furthermore, pre-school aged children require training to be able to effectively utilize AR, presenting another obstacle to its application for this age group [100]. Although AR may have some applications for pre-school aged children, there remain obstacles to its usage, and it can serve a greater purpose among older students.

29.6.2 Elementary School

Although elementary school students should still avoid excessive screen time, AR can have more applications in the classroom for this age group. One of the most exciting innovations in AR for elementary school students is AR books, which allow readers to experience their books alongside helpful visuals and animations, bringing the material to life. The MagicBook AR interface system allows educators to transform any book into an AR creation, adding various media alongside the text [10]. Educators can do this by developing three-dimensional computer graphics in the VRML format to accompany an illustration, which students can then view through their AR interface. Students view the graphic using a handheld AR display, similar to a set of goggles. When viewed through this AR display, the book will produce three-dimensional computer generated content that appears to spring from illustrations on the book page, and students can interact with the images by moving the book or themselves to alter their perspective [10, 109]. This technology can be applied to textbooks, transforming a geology textbook [10] or Greek history book [110] into a more engaging experience for students, immersing them in the material. AR books can also be adapted to make leisure reading more engaging among elementary school students, promoting positive attitudes towards reading.

In the elementary school classroom, AR can also be used to promote engagement with diverse components of the curriculum in addition to reading, such as art history. In one study, students used an AR application to point at paper markers containing QR codes to access three-dimensional figures of pre-Roman artworks. Students reported positive attitudes towards the application of AR, as it promoted active participation among and fostered greater enthusiasm for the subject [111]. Easy but engaging activities like this can be used in the classroom to promote positive attitudes towards learning. Furthermore, AR can be used to facilitate social learning, as well as familial and community engagement through collaborative gameplay, providing additional educational benefits [5].

As the COVID-19 pandemic ushered in an era of distance learning, AR presents another opportunity for promoting engagement with students even when they cannot be physically present. For instance, Banuba is an AR application that allows for real-time three-dimensional face filters. During the pandemic, some teachers found that utilization of creative face filters, especially among younger children, helped keep children engaged in the lesson and focused on their teacher [112]. Remote learning has also resulted in the cancellation of many field trips, and AR provides an exciting substitute, virtual field trips, where students can embark on exciting journeys without leaving their own home or neighborhood [113]. The Seek Education® app offers a number of these virtual field trips, including one through the Egyptian pyramids [114]. While the end of the COVID-19 pandemic will result in a transition back to in-person education, distance learning technology will remain relevant. Children may need to attend school remotely for a number of reasons, including but not limited to medical issues, family emergencies, and weather-related events. AR's applications for distance learning can help educators in the present and future keep elementary students engaged and excited about learning.

29.6.3 Middle School

Middle school can be a time of rapid changes in all areas for students, and it is important to promote a positive attitude towards education. Given this viewpoint, it is increasingly important to promote engagement and enthusiasm in the classroom among this demographic. One way to do this is to adapt existing social trends to the school environment. For example, Niantic's Pokémon GO is an AR-based game particularly popular among middle school students. One creative teacher has managed to incorporate Pokémon GO gameplay into mathematics and social science lessons. At each PokéStop and throughout the game, students were given mathematic tasks to complete related to the Pokémon GO game and were asked to analyze real-world objects in the vicinity [115]. By playing into current trends and gamifying the classroom, students were engaged in active learning and excited about their lessons. Using AR for gamification can extend beyond just Pokémon GO, educators can innovate, using the technology to further engage students. While AR can also be used within the classroom, its portable nature on handheld mobile devices also allows for teachers to take the classroom outdoors and to new settings, offering further opportunities for engagement. For example, they can create treasure hunts or scientific investigations for students to complete on school field trips or in their own backyard [116].

Remote learning during the COVID-19 pandemic can also benefit from various AR technologies. For instance, AR can help science students to visualize the 3D structure of DNA or bring a solar system to life in the absence of traditional classroom model using apps such as Google Expedition [117]. AR textbooks can increase engagement among students during distance learning, helping them to absorb more from their independent readings [10]. Furthermore, applications such as ARTutor aid students in participating in interactive distance learning and self-study. The application enables educators to create AR content by inserting augmentations such as audio, videos, and 3D models, linked to a "trigger image" in the given document. Students then use the ARTutor app to view, interact, and engage with these augmentations. The application even responds to voice commands, allowing students to ask questions which the application then produces answers for from the given document [118]. As remote learning has posed increasing demands on students to engage in independent learning, applications such as ARTutor serve increasingly important roles that will persist beyond the COVID-19 pandemic.

29.6.4 High School

While applications of AR for younger ages may emphasize student engagement and enthusiasm for academic content, the challenging curriculum requirements in high school create further educational applications for AR, specifically in science, technology, engineering, and mathematics (STEM). In chemistry, AR can be used to visualize complex molecules in 3D, allowing students to rotate and view molecules from various angles [119]. Meanwhile, students can learn about circuits for physics with MetaAR, an application developed by Purdue University that uses a model of a city to teach the basics of circuit boards, with wooden pieces representing "streets" and "buildings" that are used to assemble the circuit [120]. Applications like Arloon geometry and more can help students visualize complex shapes and figures, further enhancing STEM education [121].

In addition, high school classrooms have increasingly adopted the flipped classroom model, where students watch pre-recorded lectures before attending class, dedicating class time to problem-solving instead of lecture. Furthermore, the conversion to distance learning during the COVID-19 pandemic also resulted in more teachers adopting this strategy and recording their lectures. These recorded video lectures provide a new opportunity for AR technology, allowing educators to enhance their videos and make them more interactive for students [122]. AR can also be used to enhance notes accompanying lectures and independent reading material, providing further use in the flipped classroom [123]. By augmenting notes and video lectures, educators can better engage and teach students during distance learning and in the flipped classroom setting, allowing them to dedicate more attention to answering questions and working through problems.

29.6.5 College

As AR technology progresses, college classrooms have adopted its use across a variety of departments. Given the increased technical capabilities of college students, AR applications in the college classroom expand to include the development of AR itself. In fact, a number of schools have even started to establish majors in AR and VR [124]. Class projects involving students in the development of AR span a wide range of topics, including the preservation of ancient sites and reenactments of historic events. For instance, in the University of Denver Anthropology department, students and faculty have collaborated to preserve Amache, a Japanese-American internment camp, through both AR and VR [125]. At the New School, a media class recreated the War of the Worlds radio broadcast using VR and AR, allowing students and faculty to experience an event they were not alive for and otherwise could not experience [124].

While some college students' educational exposure to AR revolves around developing the technology itself, others use AR to learn their class' subject material. In theater departments, students can utilize Stagecraft, an app that allows stage designers to visualize a stage with no set coming to life with their performance, immersed in full-scale scenery [126]. Meanwhile, business courses at Fordham University have used AR to visualize multi-dimensional datasets, practice communication and collaboration, and simulate experiences outside the typical classroom environment [127]. Furthermore, AR has even been used to enhance campus life, from Pokémon GO inspired orientation scavenger hunts [128] to an AR application providing background information on school monuments [129].

The COVID-19 pandemic transitioned students of all ages to remote learning requiring the use of technology, with college students being some of the last to return to an in-person education. AR is an extremely valuable tool for distance learning, as noted in the innovations already mentioned for younger students. Many college students have complex STEM laboratories that they are required to complete; however, in-person attendance is no longer possible due to remote learning. Fortunately, a number of applications have offered new remote laboratory opportunities that use AR and VR to offer interactive and collaborative learning exercises to students [130]. Technologies such as Microsoft's HoloLens®, AR goggles that allow users to view the world through an augmented lens, have been put to use. At Imperial College London, engineers have worked to adapt this technology to allow students to perform lab-based experiments, showing promise for its use beyond the pandemic's remote learning era [131].

29.6.6 Graduate Education and Vocational Training

AR has a number of important applications for graduate education, most notably in medical school and training. In the field of medical education, AR can be used to support exploration and understanding that caters to the user's baseline knowledge level.

Medical School Students

Traditionally, medical students have learned from a mix of two-dimensional textbook images and in-person experiences with cadavers, mannequins, and live humans. The transition from reading a textbook to manipulating a three-dimensional form is often difficult because words on a page omit the visual and tactile perspectives necessary for three-dimensional medical practice. AR offers solutions to this dilemma.

Augmented reality is the next step in cutting-edge medical learning tools in that it incorporates virtual elements into real-life clinical applications. AR holds the promise of improving the quality of medical students' training, particularly in the areas of anatomy and procedures. It can combine the aforementioned traditional learning modalities by incorporating virtual images or text over real tangible or imaged structures.

Currently emerging technology such as a "Magic Mirror" and "Gunner Goggles®" have both been well received as having educational value by the vast majority of medical students who have tried them [11, 132]. These represent the tip of the iceberg regarding AR potential. "Gunner Goggles" adds links, videos, and images to traditional anatomy textbooks, thus integrating educational resources for a more comprehensive learning experience [11, 133]. "Magic Mirror" is a screen-based program which shows a real-time "reflection" computer image of a student with overlapped computed tomography (CT) images in order to demonstrate how structures may exist in a student's own body [11, 132]. AR is poised to revolutionize the anatomy lab in the not-too-distant future. For example, AR could contribute value to cadaver dissection by incorporating a projection of manipulatable virtual reference images, labeled structures, and visual demonstration of organ functionality. Spatial relationships can be explored through manipulation of real structures and, soon virtual ones, throughout the process of dissection. AR technology can also be used while observing clinical sessions or surgeries performed by physicians through smaller-scale AR devices such as goggles or glasses [11, 134].

Additionally, AR can be used as an educational tool to help students learn proper medical technique. A study showed success in using ARToolKit®, an open-sourced AR library, to teach previously untrained people how to accurately place electrocardiogram (EKG) leads. Markers were placed on the real model, which were then used to calibrate a camera attached to the user's head-mounted display. Instruction was provided through a combination of audio messages, visual text-boxes, pointers, and other features overlaid on the user's computerized field of view within the head-mounted display [135]. Though this study's success was achieved in a controlled setting, it is feasible that similar technology can soon be utilized to help teach proper suturing, splinting, or casting techniques in a functional setting. Although AR technology cannot fully replace the in-person experiences of dissecting cadavers and performing complex operations, it can serve important educational purposes and enhance learning when these in-person experiences are unavailable. While using AR in conjunction with textbook can improve book-based learning, AR can never replace learning with human subjects.

Physicians and Clinicians

Physicians hold patient lives in their hands and therefore must stay up to date with the latest standards of care. Lifelong learning is mandatory for the duration of a doctor's practice. In this way, they can also benefit from similar augmented technology opportunities as those students in medical school.

In addition to helping refresh basic medical knowledge, AR can also support training physicians to prepare for emergency medical events such as traumas and medical codes. When emergencies occur in a medical setting (most frequently a hospital), it is essential that personnel be prepared to handle the crisis. Real patients should not be an initial learning ground, which is why special computer and mannequin-based simulation programs exist to certify preparation of clinical personal. A study that examined the potential of using AR cardiopulmonary resuscitation (CPR) training system when teaching healthcare professional for emergency situations. The study concluded that it was feasible to do CPR training with AR technology and such programs were found to be favorable by study participants [136]. In the future, it may be useful to implement AR training programs for Basic Life Support [137], as well as for Advanced Cardiovascular Life Support, Pediatric Advanced Life Support, and Neonatal Resuscitation Program training. These programs, as they currently exist, typically include a computer portion which employs virtual simulation events, followed by an in-person training session with a lecture and simulation events using stationary human dummies. Though these programs do provide exposure to prepare clinicians for emergency and code events, they do not offer a realistic setting or realistic materials.

Augmented reality may be the key to enhancing realism and quality in preparedness training. In the future, AR technology may provide the ability to project a virtual interactive image of a sick patient who reacts to illness and treatment. As it is now, most simulations involve verbal walk-throughs of situations and vague miming of procedures. However, AR could provide a more lifelike experience of manipulating equipment and providing care in a code situation by practicing proper technique and allotting for the true time it takes to complete each resuscitative task. AR has the potential to continue enhancing experiential learning and clinical preparation [138].

Patients

The word "doctor" originates from the Latin term meaning "to teach." It is the duty of clinicians to not only care for their patients but also to impart the knowledge necessary for patients to understand pathology and treatment options. AR can be used to improve communication, and therefore the therapeutic alliance, between clinicians and their patients who typically face communication challenges such as having those with a non-native primary language or those with hearing impairments. Technology which can provide real-time transcription or translation of medical discussions is currently available [139, 140]. However, this technology is not recognized as the preferred solution due to refinements necessary to provide accuracy and interpretation of cultural context which, at this time, is still best interpreted by human translators.

Patients may also benefit from augmented technology as an adjunct to verbal conversation with a physician in order to visually clarify the course of medical care. AR may be applied to plastic surgeries through manipulation of the patient's current form to virtually project a three-dimensional potential postsurgical outcome [141]. "Magic Mirror" is one such application which was originally developed to assist plastic surgeons in collaborating with their patients who were considering breast augmentation [11]. Physicians in other subspecialties may use a three-dimensional model of a patient's organ as a tool for communicating abnormalities, pathology, and interventions [142]. For example, a cardiologist may show a patient or colleague the structure of anomalous arteries on a model which is customized and truly representative of a patient's individual heart structure; this can make it easier to explain or collaborate about surgical plans or goals of treatment with medication. Outside the clinic, AR games have been used to support patient education by providing positive reinforcement through points or advancement to the next gaming level. One study described an AR Android mobile game that was designed to educate children with diabetes about food carbohydrate content. The game utilized AR by having game players interact with virtual images of food overlying a real dish; players gained points and advanced through levels by correctly identifying the carbohydrate content of the imaged foods. The results of this study suggest that playing this AR educational game resulted in improved knowledge of food carbohydrate content across all age groups, including younger populations [143].

Other Vocations

Other vocations have also found innovative uses for AR to enhance and expedite job training. In the military, AR has been used to simulate dangerous situations to better prepare members for potentially real scenarios [144]. Construction sites are working to adapt their training to utilize AR to better simulate the worksite without placing untrained employees at risk [145]. AR has even been adapted for use among aircraft maintenance technician trainees, providing visual aids during their learning to prevent the need to flip through tedious handbooks while also offering feedback along the way [146]. AR serves numerous applications for vocational training and is especially important for workers completing potentially dangerous tasks. Researchers should continue to develop and progress these training models to better prepare workers for their duties and to maximize their safety.

29.6.7 Augmented Reality and the Disabled

AR can be used to augment experiences for those with different abilities [147], from providing social structure and cues for children with autism [148] to shopping help for those with physical disabilities [149], the possible uses of AR are endless. Additionally, AR can be an essential tool in teaching and reinforcing safety lessons, without being in dangerous environments [147]. The authors believe AR technology should be considered when Individualized Education Plan goals align with technologic advances. Training in the use of these programs should be offered to those currently training for careers in special education and those that have already completed their educations and could benefit from introduction and training in the uses and abilities of AR applications. As with all technologies, the abilities of the teacher will predict the opportunities afforded to the student. While AR may require educators to learn new skills, AR can offer the student unrivaled opportunities to learn and acquire new skills. It is therefore essential that AR companies have comprehensive instructions and explanations on the applications and use of their products.

29.6.8 Cautions

Although AR serves numerous valuable educational purposes, its usage still comes with limitations. Especially among school-aged children, caution must be taken to ensure users do not become so engaged with AR that they lose sight of reality, potentially risking injury [100]. Furthermore, effective usage of AR requires training of both teachers and students in its usage given its complexity [150]. Many educational institutions may lack the resources to provide this training or even to afford AR technology, slowing its uptake in the classroom [151]. Furthermore, while AR enhances opportunities for distance learning and offers its own unique benefits, it cannot fully replace the in-person classroom experience in remote settings. Parents and educators should also remain wary of encouraging excessive screen time for children and adolescents, even for educational purposes. Although many are reasonably hopeful regarding the future of AR in the classroom, this optimism must be tempered by the technology's limitations, and various challenges must be addressed in order to enable its widespread application. However, it is important to note that AR's limitations in complexity, accessibility, and training are significantly reduced compared to its more expensive and complex counterpart, VR. While AR can be accessed from goggles like Google Glass or a simple handheld device, such as a smartphone, VR generally requires more complex technology like a bulkier goggle set. Consequently, it is more difficult to involve all students in the classroom with VR, while bring-your-own-device models of AR can vastly increase accessibility. By leveraging its relative portability and ease of use, AR can serve as a relatively accessible and effective way of engaging students in exciting new worlds and bringing academic subjects to life, all from the safety of AR-based glasses or even a smartphone.

29.7 Augmented Reality in Medical Practice

AR is rapidly integrating into many aspects of life. The medical field is no exception – AR is currently part of medical training and medical practice, and it has potential to support patient therapies and improve therapeutic experiences for patients. Education of medical students, physicians, and patients has started to incorporate AR. Additionally, AR has therapeutic value in helping clinicians with communication, planning of procedures, and clinical management. It also can be used as an adjunct to patient rehabilitative therapies, psychotherapies, and act as a motivational tool to meet health/fitness goals. Epidemiologic applications of AR have not yet been fully realized within the field of public health.

29.7.1 Educational Opportunities

Please refer to Sect. 29.6.6 for details on how AR is being implemented in medical education.

29.7.2 Therapeutic Uses

Augmented reality technologies are helping move medical therapeutics forward. It can help with diagnostics, planning and execution of treatment, as well as patient rehabilitation after procedures. This blending of virtual elements with reality may serve to improve clinician and patient experience as well as possibly improve patient outcomes.

Medical Procedures

Some of the first ways AR was incorporated in medical treatment started in the hospital setting. A commonly used example of AR technology is vein finder devices such as the AccuVein®. Vein finders project a person's vascular map onto the surface of their skin to help during phlebotomy or intravenous line insertion [152, 153]. These devices are not necessary for all patients, but make a significant difference in improving success rate and decreasing patient discomfort for those with suboptimal vasculature or dehydration.

Augmented reality is frequently used to help physicians understand anatomy and prepare for surgeries. For example, cardiologists and cardiothoracic surgeons utilize AR's ability to overlay computerized tomography (CT) scans or magnetic resonance images (MRIs) of their patient's unique anatomy in order to understand pathology and carefully plan procedures [154]. Additionally, these images can be manipulated to practice an operative approach tailored to an individual's specific organ(s) structure [142]. In this way, surgeons can best plan their therapeutic approach and account for potential risks with less guesswork than basing their surgical plan on general experience and two-dimensional imaging.

Intraoperative visualization of structures and surgical manipulation are evolving uses of AR. Neurosurgical stereotactic surgery has a long history of integrating virtual images from radiographic scans to make surgical movements safer [155]. More recent breakthroughs in surgeries assisted by AR include vascular surgeries with the introduction of the HoloLens™, which is a tool that has been proven helpful to prepare for and to execute intraoperative reconstruction of subsurface vascular structures by projecting Computed Tomography Angiography images over a patient's skin thus allowing surgeons to visualize structures before the first incision [156]. Other procedures may require use of a surgical robot, and in some cases, these robots are controlled from a remote location. During robotic surgery, images of surgical tools are projected over real-time video of the surgery. Image integration in the past was done through a "head's up" method in which live video with virtual image projections were viewed on a screen. However, recent advancements in lightweight wearable AR technology now allow for some images to be projected onto the patient directly for a "see through"-like image which allows a more natural posture for the surgeon [155]. Overall, these advancements in operative approaches can serve to improve the safety of practice by refining dexterity and precision during procedures – especially microsurgeries [155, 157, 158].

Patient Communication, Management, and Experience

There has been a recent interest in advancing telemedicine and supporting integration with AR as a postoperative and clinical management tool. Telemedicine is a modality that allows clinicians to care for patients remotely by connecting virtually with patients, such as those with health or logistic limitations, who may not otherwise be able to present to an office for management [11]. In Australia, Google Glass was used in a trial partnership between Small World and the Australian Breastfeeding Association to have remote at-home lactation consultations which allowed providers to observe and adjust the mother's technique from her own perspective [159]. Similarly, Dr. Brett Ponce developed the Virtual Interactive Presence® (VIP) system which uses telemedicine hybridized with AR to guide patients through postoperative care using an app [160]. Patients and physicians found VIP to be superior to phone, text, or email correspondence. This is because VIP offers the benefit of being able to virtually see both the patient's and physician's visual perspectives and therefore interact through AR to evaluate wound healing and dressing, as well as guide medical equipment management [160]. Though VIP was found to be clinically helpful when it worked, there were significant technical issues and a learning curve for appropriateness of use [160]. Due to the 2020 Coronavirus SARS-CoV-2 pandemic, telemedicine has become a commonplace modality for doctor's appointments. However, telemedicine technical issues persist and must be improved before the full potential of telemedicine-AR integration can be realized. Augmented reality has the potential to revolutionize the patient experience as it relates to healthcare. Patients often have negative feelings associated with hospitals and procedures. This may be due to either experience or anxiety around the unknown. In both cases, AR can be utilized to help patients cope with difficult healthcare experiences. Alder Play is an example of an app designed for children which incorporates VR images of a hospital, or other potentially scary experiences, and utilizes a virtual animal to guide them through what may happen as well as some AR interactive strategies to keep calm. Additionally, children who cooperate with care earn tangible stickers which can be scanned using a device camera to get virtual rewards in the Alder Play game [12]. Some hospital systems utilize VR and AR to distract children during procedures as well. For example, Lucile Packard Children's Hospital Stanford has started using an AR program that enables patients to observe and participate in their procedure through the CHARIOT program [5, 161, 162]. Other studies have showed that AR can be used successfully to lessen children's fears and anxieties, as well as reduce reported pain severity, during otolaryngologic procedures [163] and burn wound dressing changes [164]. These interventions may have utility in populations other than children, such as those with limited intellectual ability, autism, or anxiety. For example, some studies suggest promising findings about the effectiveness of AR technologies in improving the social interaction abilities, attention skills, emotional recognition, and functional life skills of individuals with autism [165].

Nonprocedural Medical Uses

Augmented reality can be incorporated into medical therapies such as physical therapy and occupational therapy either as teletherapy or in person. AR technology can provide logistics for practitioners to observe patients and suggest corrective exercises or promote good exercise technique. An AR program called Ghostman tested the theory of AR use for telerehabilitation by allowing physical therapists to remotely instruct and adjust patient progress from the patient's point of view by overlaying a ghost image on top of the patient's egocentric view through an ocular device worn by the patient [11, 166]. In a 2014 study, patients who were taught to use chopsticks remotely by therapists through Ghostman had the same rate of learning and skill retention after 1 week, compared to patients taught by a therapist in person [166]. AR can also act as motivation to complete therapy exercises, which are sometimes uncomfortable and frustrating, by incorporating instruction into a gaming format. In 2020, Unity Technologies developed an AR smartphone game targeting 50- to 70-year-old stroke victims for use during occupational and physical therapy sessions to practice rehabilitative exercises for upper limb deficits [167]. The game used an AR headset and custom device worn on the deficit limb to track limb movement. As the patient became more active, the movement-controlled AR images of a dolphin which swam to capture fish and feed turtles. Patients found the game to be "motivating, comfortable, engaging, and tolerable" (p. 1) [167]. In the future, as therapeutic gaming improves, AR may become a motivational therapeutic tool to increase compliance for patients at home.

Mental Health

Augmented reality may also prove useful in the field of psychotherapy. There is evidence that VR and AR images of a person's fear, as presented through an ocular device, can evoke anxiety similar to that of a real-life experience [168]; therefore AR may be useful for the treatment of phobias. Exposure therapy is the primary treatment for phobias which entails discovery of the phobia, discussion of the phobia, and slow introduction of exposure to the anxiety provoking situation progressing from mental visualization to real-life confrontation of the fear. AR and VR can be utilized as a tool for this therapeutic transition and has been shown in some cases to reduce real-life avoidance behaviors [169]. The benefit of AR use over VR use is that AR can allow patients to physically practice coping techniques since AR can visually register external movements and use of external objects in their simulated field view. This can be advantageous by providing reinforcement of clinical progress before real exposure challenges [169]. VR can show realistic images of the anxiety provoking situation, but would not be able to incorporate or analyze how a patient reacts to the stimulus. Though the choice of method of phobia exposure therapy should be tailored to the individual patient, there is some evidence that similar outcomes may be achieved with AR, VR, or in person-guided treatment [170].

Health/Wellness Goals

Patients may also harness AR for psychological positive reinforcement in order to reach health goals recommended by their physicians. For example, AR can provide a motivational tool to promote weight loss through exercise. In 2016, Pokémon Go became the most rapidly downloaded application and took the world by storm as it mobilized typically sedentary gamers by overlying game images with the natural environment using a mobile phone's camera [171]. The app's interface with reality was engaging, and many people started walking outdoors due to the positive reinforcement of opportunities to virtually "catch" Pokémon. Since then, AR uses for exercise have proliferated including the now commonplace use of pedometers and fitness trackers in smartphones and watches, which give reminders and feedback about progress towards daily goals. Additionally, exercise machines have started to utilize AR. Peloton® machines, stationary exercise bikes by the Peloton company, feature VR in that they allow users to exercise in virtual spaces alone or with others in the Peloton community [172, 173]. The Mirror Home Workout System® expanded on the Peloton's VR interactive model by projecting an AR fitness instructor onto a mirror so users can compare their movements and strive towards the ideal exercise technique and form [174].

Overall, AR technologies have widespread applications in the field of healthcare. Some uses are better developed than others, yet there is no doubt that the expansion of this technology will lead to novel advancements which will further medical education, therapeutics, as well as public health initiatives.

29.8 Augmented Reality Safety Precautions

Safety measures need to be taken to reduce the risk of harm when using AR. Below are some simple measures recommended by the authors that may reduce the risks of physical harm to young users associated with AR. Throughout Sect. 29.8 on AR safety precautions, please refer to Table 29.1 at the end of this section for more detailed risks and remedies.

29.8.1 Safety Prior to AR Use

Indoor Considerations

AR use can involve active movement. Indoors, it is imperative that the floor area is clear of any hazards or impediments. Trip hazards must be eliminated as they may result in falls and even head injuries. Users should clear space of hard corner furniture and glass tops in order to prevent injuries. All electric cords should be out of the way and may require taping with painter's tape in order to prevent further tripping injuries. Lastly, special care should be taken when using AR in the kitchen as full attention may not be on appliances and sharp objects. The excitement of using AR must be tempered with common safety precautions.

Outdoor Considerations

If AR is being used outside, prior to users walking/running, careful consideration should be taken to clear all areas of obstacles. AR should be used in open areas so that large natural items such as trees can be avoided. Special consideration should be used when using AR around pools or other bodies of water to avoid drowning risks and AR water damage. Additional caution is necessary when using AR on bicycles, skateboards, rollerblades, and scooters. As these devices may increase physical momentum, AR should not be used as response time is reduced. AR use should also be avoided in severe weather or slippery conditions, such as snow and rain. AR should not be used in conjunction with sharp objects or when operating machinery.

Driving

Motor vehicle accidents are the primary cause of death in 10–24 year olds [175]. Distracted driving is one of the major causes of vehicle crashes. Distracted driving occurs when one engages in another activity while driving. This may involve mobile phone or a similar electronic device use, which can distract the driver from the primary task of driving and greatly increase the risk of motor vehicle accidents [176]. The use of smart devices and activities like texting or playing games while driving have caused driver fatalities to rise at striking levels. AR use while driving is considered a major distraction and is not recommended while driving a car in any situation. As young drivers are unexperienced, the risk of injury and fatality is heightened.

A study performed by Faccio and McConnell (2020) in Indiana examined the association between motor vehicle accidents and the use of the popular Pokémon Go game around its launch using police accidents reports. Research identified a significant increase in motor vehicle accidents causing damage, injuries, and deaths in locations where Pokémon Go play was possible. If fact, researchers calculated that in the 148 days following the release of Pokémon GO, the financial ramifications of users playing while driving cost up to $25.5 million. Nationally, this cost would be up to $7.3 billion [177].

A study conducted by Ayers et al. (2016) also analyzed the use of Pokémon GO in relation to accidents, using Twitter social media mining [178]. Researchers searched for tweets that contained references to Pokémon and driving over a 10-day period. There were over 345,433 tweets identified, and a random sample of 4000 tweets was analyzed. There were 14 discrete crashes and several other near misses reported. Eighteen percent of people posted about playing Pokémon while driving, 11% of posts indicated a passenger was playing, and 4% of posts indicated a pedestrian was distracted and was almost involved in a motor vehicle collision [178].

People need to be aware of the very real distractions and risks associated with AR while otherwise occupied in the real world. There should be more warnings associated with games that utilize AR technology. For example, a pop up could appear on the screen each time a game boots up, reminding a user not to play while driving and to be attentive when walking. AR applications should automatically shut off if connected to a car. Pokémon currently disables portions of

Table 29.1 AR safety precautions and remedies for children and young adults

Risk	Remedy
Indoor considerations	
With eyesight obstructed due to AR technology, hazard on the floor may cause injuries	Prior to AR use, clear floor space of any obstructions including furniture, wires, and appliances
For small children, tablecloths provide a unique hazard as the loose fabric can get caught on active young players, causing objects on the table to fall off and cause injury	Do not allow small players to use AR in spaces with tablecloths. If it is necessary to play in these locations, remove tablecloths
Low hanging wires or strings can cause "clothesline" injuries	Tape low hanging wires and strings prior to AR use
Outdoor considerations	
Crashes can result from using AR on scooters, bicycles, skateboards, and rollerblades due to obstructed vision	Do not use AR while riding on recreational equipment
Hazardous weather increases risk of physical injury due to unforeseen conditions, and in conjunction with AR use these conditions can be more dangerous	Do not use AR outdoors in hazardous weather such as during rainstorms, lightning storms, and blizzards
Using AR as a driver in a moving vehicle enables distracted driving which can be dangerous to the diver, passengers, and others on the road	Do not use AR when operating a motor vehicle
Other considerations	
Prolonged periods of sitting can cause physical injury while using AR	Remove headwear and take frequent movement breaks. This risk is increased for younger children still developing, and they should take breaks every 15 min for at least 10 min
Prolonged periods of digital screen use can cause eyestrain	Take a break from looking at screens every 20 min that lasts at least 20 s while looking 20 feet away
Repetitive motion caused by repeatedly pressing buttons and moving levers can cause carpal tunnel syndrome, tendonitis, and synovitis	Ensure players take repeated breaks and young players use appropriately sized gaming devices
Gaming disorder and other mental health conditions can occur in players using AR	Parents should ensure children have limited game time and ensure they have access to mental health services and evaluation if necessary. Additionally, parents should organize exercise and other social and physical activities outside of AR use for children

their game when a user is determined to be going above a certain speed [179]. Other companies should consider providing these advances.

Distracted Driver Prevention AR Applications

One organization, the PEERS Foundation®, is addressing this issue in a different way. Using AR, they created a hands-on experience to prevent distracted driving. PEERS uses an Augmented Reality Distracted Driving Education Simulator (ARDDES) to create this experience [180]. ARDDES uses a real vehicle in which distractions and real-world environments are projected onto the windows in order to simulate distracted driving. In this way, users can experience the real-life dangers and consequences of distracted driving in a completely safe environment. Users can text, scroll through social media, listen the radio, and navigate while in the vehicle to replicate distractions and to experience the change in reaction time to potential danger [180]. This type of program may be very useful for teens who may not understand risks associated with distracted driving. This hands-on experience may help young adults be more conscious about the risks of distracted driving and avoid playing games while driving or as a pedestrian. The ARDDES program should be made available in high schools and college, and this may help decrease the incidence of distracted driving.

29.8.2 Safety Measures During AR Use

Injuries and Risks from Spatial Ware and Body Position

The safety of prolonged use of spatial ware or headsets that produce AR images has not been extensively studied. Industry experts believe that using AR devices for an hour or two is not harmful; however, they remain uncertain about the effects associated with prolonged use of AR devices [181]. Dangers associated with the use of AR technology include headaches and neck strain due to the weight of the technology. Additional concerns include exposure to optical radiation [181]. There should be standardized regulations for the production of such headsets in order to ensure the safety of all users. Future studies should explore the maximum time that AR users should use devices and gear without injury in order to make firm recommendations on AR usage.

Users should take frequent breaks where headsets are removed. The AAP recommends that children have limited use of screen time. For children between the ages of 2 and 5 years, they should only watch 1 hour of high-quality programming per day, while children under the age of two should have no screen time unless video chatting with relatives [182]. Older children should have limits on their amount of screen time set by parents [108]. These limits can help allow for breaks when using AR technology to allow for movement. Additionally, the authors of this chapter include developmental and behavioral pediatricians; they recommend that for younger children using AR, breaks should be taken at least every 15 min and should last at least 10 min. Children should perform gentle neck range of motion exercises. If any pain is experienced, AR use should be discontinued until the child is pain free. If discomfort continues, the child should be assessed by a medical professional. It is imperative that children maintain proper posture in a relaxed position with feet placed flat on the ground when playing games. If one does not maintain proper posture when sitting, it increases the risk for several conditions. These medical conditions will be further discussed below. Movement breaks are recommended every 15 mins, with the movement of arms and legs by either walking or performing simple exercises.

Eye Strain

Eye strain can be caused by prolonged time looking at digital devices [183]. It is imperative that children keep track of time spent on devices as AR devices may be exciting, and children may be prone to spending more time on the screen than originally planned. AR incorporates elements of the real world and digital, potentially making it easier for children to not realize the time they have spent playing.

Eye strain, also known as asthenopia, occurs when one's eyes spend a long time focusing on various tasks [184]. The stress placed on the muscles of the eye from focusing for extended periods of time can result in a headache, double vision, or blurred vision. Dry and irritated eyes can happen when people spend too long staring at a screen and not blinking [183]. Care must be taken to prevent eye strain and its complications from occurring. The American Academy of Ophthalmologists recommends the 20-20-20 rule when using screens. Every 20 min, screen users should take a 20 s break to look 20 feet away to reduce eye strain [183]. This is difficult to remember especially when someone is immersed in a game, so timers can be set as reminders.

If a young user experiences eye pain, blurred vision, or double vision, AR use should be discontinued. If these symptoms continue, the AR user should consult a medical professional as soon as possible for further guidance.

Deep Vein Thrombosis

Deep venous thrombosis (DVT) is a concern for persons who spend prolonged periods of time in non-movement situations. DVT occurs when the blood pools in the lower extremities and a clot occurs. This clot can dislodge and travel dangerously to the heart, brain, or lungs. While traveling, the clot can stop moving and cause painful swelling. If the clot travels to any organs, this can become a medical emergency and even lead to death. Obesity may increase the risk of DVT even further [185]. People often get very engrossed in AR games and may play for hours at a time. Though some types of ARGs may involve movement and even walking, several do not. It is important that frequent movement and activity breaks be taken when playing games for a long period of time.

Injuries from Repetitive Movements

Overuse injuries can occur with overuse of the joints of the hands and wrists. Repetitive motion caused by repeatedly pressing buttons and moving levers can cause carpal tunnel syndrome, tendonitis, and synovitis. This can cause weakness and pain in the wrists and hands, and significantly interfere with childhood tasks needed for school such as writing, typing, and drawing [186].

Game pads or controllers are made to fit adult hands; however, children have smaller hands, but still use the same size game pad and mobile devices. A comparison of a standard game pad in an adult's hand and a seven year old's hands is seen in Fig. 29.2. There is an obvious difference in sizing and how the game pad will be held and manipulated while playing. This can increase their risk of such injuries. There are smaller size game pads on the market that are often sized for adolescents (Fig. 29.3).

One should take breaks while playing games and perform hand stretching and strengthening exercises to decrease the chances of developing carpal tunnel syndrome.

While repetitive hand injuries like carpal tunnel syndrome have been extensively documented, not as much research has been done on the injuries that may be unique to AR gaming. A recent study by the University of Oregon examined common movements in both VR and AR headset play. Findings indicated that in as little as 3 min, shoulder discomfort can occur with extension of the arms [187]. Gorilla arm syndrome and rotator cuff injuries may also occur with prolonged use [188].

Video game makers must address some of these issues to make games safer for children. However, in the interim, frequent breaks need to be taken while playing these games and positions frequently readjusted.

29.8.3 Mental Health

According to a self-report study, AR games were used by participants during the COVID-19 pandemic to reduce stress, relax, get exercise, and stay socially connected. The study suggested that improved physical and mental health was the result of stress relieving AR game play that provided structure and routine, as well as an escape from the overwhelming nature of the pandemic [189].

While some aspects of AR can be positive during pandemics, other complications may arise from overuse. Vul-

Fig. 29.2 (**a**) Adult holding a regular-sized game pad. (**b**) Child holding a regular-sized game pad. (Photo captured by and courtesy of the author, Yaa Asante)

nerable individuals, including children, may be at risk for future diagnosis for gaming disorder secondary to the highly addictive nature of these play for reward games [190]. The World Health Organization reinforces that gaming disorder as defined in the 11th Revision of the International Classification of Diseases is "a pattern of gaming behavior ("digital-gaming" or "video-gaming") characterized by impaired control over gaming, increasing priority given to gaming over other activities to the extent that gaming takes precedence over other interests and daily activities, and continuation or escalation of gaming despite the occurrence of negative consequences" [191]. Gaming disorder will require consultation and treatment with mental health professionals. Teenagers and young adults, especially those involved in gaming, need to be made aware of the risks to their mental health with prolonged gaming and have easy access to mental health services if needed. Pediatricians should discuss gaming with their young patients during routine well check visits in order to monitor for symptoms of gaming disorder. Therefore, it is essential that parents closely monitor the extent of children's AR use. Gaming disorder varies with individuals, but impaired control over gaming is the key symptom. This can occur with short game time and cannot be ruled out if a child is exhibiting key symptoms but does not have prolonged play. Especially with AR, which may offer an escape from the real world, parents must mediate and monitor use to protect against significant impairment in socialization, education, or other areas of functioning. Parents should outline clear rules for the use of AR, including times for stopping and starting play, game content, player language, and stranger safety [192]. Asking children to play with the sound on in order to have parents listen in on gaming occasionally will allow parents to be actively involved. Children who consistently wear headsets make it difficult for parents to overhear, and this may interfere with direct supervision.

Parents need to ensure that their children maintain consistent sleep times to allow children to get necessary rest they need in order to properly function in all aspects of their lives. Game play should stop at least an hour prior to bedtime and should not start until at least an hour from rising. Parents should be aware that children often will "sneak time" in early morning or late evening, and devices may need to be confiscated in order to maintain sleep routine.

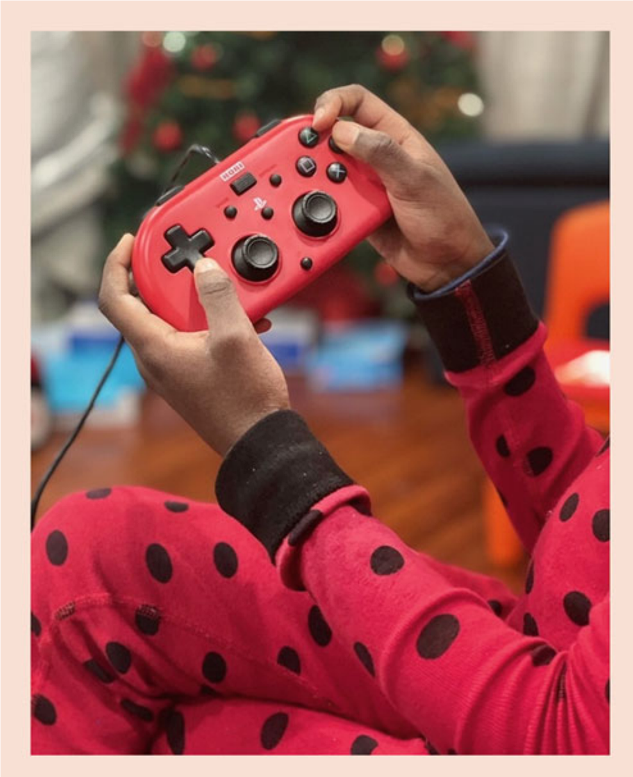


Fig. 29.3 A youth-sized game pad which is about 40% smaller than the adult controller. (Photo captured by and courtesy of the author, Yaa Asante)

Parents should be encouraged to experience AR with their children as the shared time together can be a nidus for discussion and family fun. Parental involvement in AR will encourage parents' familiarity with games and applications that their children are using and enjoying. This familiarity can help parents to make appropriate decisions based on direct opinion and their child's maturity. While there are health and emotional concerns, AR can safely be used in most situations for the enjoyment of all, with proper precautions taken.

29.9 Considerations and Causes for Concern

29.9.1 Age Considerations for Augmented Reality

While AR technology can be enticing for both parents and children as it offers a variety of learning and entertainment possibilities, recommendations on use of this technology vary with age. Researchers have explained the dearth of research on children using AR, citing numerous difficulties obtaining permission and consent for AR exposure studies with younger children. Although research in the field remains paltry, reasons for concern exist related to younger children using AR technology, nonetheless. Taking specific cautions may be necessary.

Piaget Theory Age-Related Concerns

Object permanence is the understanding that physical objects continue to exist even when they are out of one's sight. In his child development theory, Piaget states that object permanence slowly develops during the sensorimotor stage (0–2 years old). It is not until later in the sensorimotor stage, beginning at 16 to 18 months of age, that a child understands object permanence [193]. Over the following 6 months, until the age of two, this concept becomes more concrete as the child enters the preoperational stage (2–7 years old) and begins to develop language, memory, and imagination skills. It is in this critical phase that children begin extraordinary brain development that requires both sensory input and educational experiences [193]. While AR can supplement this development, it is essential that it does not interfere with the natural understanding as a result of merging the real world with the augmented world. Additionally, it is theorized by the authors that AR may influence development as it may decrease the use of conventional imagination when digital stimuli can provide preconstructed altered realities. The developmental outcomes of using AR at this age are yet to be determined. Therefore, AR use should be discouraged for children under the age of two, and then great care should be used when first exposing children to AR. While some small children may delight in the augmented images of glasses and dog tongues found in filters, others may be deeply disturbed by the notion. Prior experiences with digital media may not prepare a child or predict what child will be disturbed by AR.

While direct use of AR by infants is discouraged, AR technology can be used to help infants in many new and emerging ways. A promising application of AR is being developed to be used by pediatric physical therapists in the treatment of torticollis, a rare condition involving the muscles of the neck. Franklin et al. (2020) have developed AR technology that uses markers on clothing to measure and display the angle and rotation of the head and shoulders and track progress over time on a mobile app [194]. These supportive systems could help therapists gauge the success of recommended treatment strategies and demonstrate significant changes to parents, thereby encouraging adherence and participation in these therapies. Programs of this nature allow therapists to clearly see a path to success and share therapy progress and results with physicians, educational agencies, and additional therapists [194]. Another new AR technology that will benefit infants is a program designed to help parents visualize their infant's world and therefore provide infant-eyed view of the play area in order to baby proof the space

[195]. The app uses the phone's camera to detect markers indicating locations of physical objects, and then the app overlaps possible scenarios of how a child could interact with the space. Emerging AR programs created and published by Nishisaki are designed to help parents see developmental opportunities while decreasing accident risk [195].

AAP Recommendations and AR-Specific Supplementation by Authors

It is understandable that medical organizations recommend caution when using all digital devices, including but not specifying AR devices. The AAP recommendations include no digital device usage except for video-calling of family members for children under the age of two, and carefully constructed media diets for all children over the age of two. Media diets should include limits on how long a child spends on a certain game/activity as well as what type of game activity should be allowed [182]. Emphasis is placed on quality and quantity concepts, with younger children being shielded from all forms of "video violence." Additionally, parental supervision is recommended [196].

Many concerns exist with AR's digital cousin, virtual reality (VR). The majority of VR technology recommends a firm 12–13 year lower age limit with The Oculus Rift® and Samsung's Gear® VR headsets being advised for use by ages 13 years and older, with Sony's PlayStation close behind with age recommendations of 12 years and above [197]. HTC's Vive® was not built for and should not be used by children without specific age speciation. Even Google® states that children should only use its Cardboard® headset when adult supervision is present [198]. Similar formal guidelines do not exist for AR technology. However, age-related precautions may be necessary.

AR technology does not rely on clunky headsets, like VR, that can injure small head and neck muscles of younger children. Even with this advantage, it is still unclear how the realistic images produced by grafting augmented objects into real-time settings in AR programs, such as Google Animal AR and Pokémon GO, will be interpreted by younger children [199]. Indeed, child entertainment giants such as Disney® and National Geographic® have been using AR technology for over 10 years. Disney has even introduced the DisneyNow app which utilizes AR to graft popular characters into games based on rides at the amusement parks themselves, except they appear in your own location or home virtually [200]. Additionally, Disney has had large scale Times Square displays in New York City of AR characters interacting with tourists in 2011 [201]! National Geographic grafted extinct species into shopping malls [202]. With large companies using AR technologies targeted at younger audiences, there still exists a surprisingly small amount of research that has been conducted on the safety of these AR concepts on young children's growing minds. Therefore, while the AAP does not specifically provide guidelines for children's use of Augmented Reality, this type of guidance is important. AR should not be used with children under the age of two in agreement with AAP guidelines for digital usage [182] and previously mentioned Piaget's developmental theory's concerns [193]. Beyond the age of two, AR should be limited to less than a half an hour a day with the supervision of a responsible adult. As children this age intrinsically struggle with concepts of reality and unreality, AR may provide significant psychological challenges as they struggle to understand that objects that appear on screen are not, in truth, real. It should be stressed that AR usage should be delayed until a child acquires this developmental concept; this concept is the important factor, not age. At no time should this technology be used to scare or surprise children for the amusement of others. Seemingly harmless pranks can have long-lasting psychological implications, and some AR images can be horrifying to unsuspecting individuals of all ages.

From ages 4 to 10, the authors recommend that AR should be used as an adjunct to learning and entertainment programs to augment positive experiences. While limited experiences, especially with parent supervision can be highly enjoyable, immersive prolonged experiences are not recommended. It is essential that children spend a significant amount of time in the "real world" without digital interference. While special considerations and exceptions have been made during pandemic times, it is essential that limits be readjusted after health concerns have passed.

While time limits are loosened as children grow [182], setting limits on digital and AR technologies remains essential for teens and young adults [41, 108, 203]. As children mature into adults, the important concept of being able to set limits on one's own entertainment practices increases as supervision decreases. This concept, if learned as young children, will help older children to self-regulate and separate from AR and electronics in general.

Discussions about media responsibility with older teens and dependent young adults is important prior to the purchase of AR technology. Teens leaving for college and or embarking on adult life experiences should understand the concepts of implementing self-limitations on AR and technology usage. Considerations of physical, mental, and ethical issues prior to unsupervised and unlimited usage of AR technology are essential for older teens in order to prevent negative outcomes such as digital addiction, obesity, bullying, and persuasive usage [41].

29.9.2 Gender Considerations

Gender differences in the digital community have been a rapidly changing front. As the number of female gamers continues to increase, the industry, including AR companies and programs, have taken considerable steps to further capture girls and women as customers. Literature reviews suggest that women still feel that female characters in video games are over-sexualized and that they face more online harassment than their male counterparts. While conventional video games do not target female audiences, AR games, as they are based in reality with a virtual overlay, strive to be more inclusive of all genders [204]. Girls and young women should be encouraged to safely explore AR options that are presented to them, especially in educational and instructional environments. Female use of and involvement in the fields of creative content and technical creation of AR will help to expand the field to all genders.

Video games and digital media are not the only areas where AR technology should be targeted to both genders. Workers' training and other tasks can be effectively taught to both genders using AR as a superior method. A research article by Hou et al. (2013) emphasizes additional ways that AR can be beneficial to both genders, other than in gaming. The researchers found that AR educational programming helps trainees of both genders more quickly learn and remember assembly routines and was more effective for both males and females than using 3D manuals. Males were more effective using 3D manuals than females, exposing another benefit of AR regarding gender equality [205]. These unique learning opportunities should be available to all, and continued efforts should be made to explore the concepts of teaching adults with AR.

AR programs are currently being used to combat gender inequality. Different gender roles can be explored by using VR and AR tools such as BeAnotherLab, where a person can experience what it is like to be another person in a situation using multiple cameras and headsets [206, 207]. As understanding an issue can often be an impetus to change, this ability to understand another's point of view can help those of all genders to look at situations in a different light. Ironically, some programs that were designed to help combat inequalities may appeal more to males, as men seem to have more of an affinity towards video game and digital devices [204]. It is unknown whether this is a natural ability in this area or an acquired ability. Conversely, research by Dirin et al. (2019) on AR and VR showed that when presented with novel programming, women were more enthusiastic about using AR technologies and applications than men [208]. Clearly, this is a research avenue that should be explored. Both the differences in learning and preferences should be considered as AR programming and product development progresses. In addition, a study on AR technology and gender differences in pilots concluded that both gender groups stated they enjoyed receiving positive feedback when using AR games. However, there were differences in the satisfaction of in-other in game experiences. A higher percentage of women reported more satisfaction compared to men regarding in-game experiences. The findings of this research could be used to further explore and promote gender diversity in educational opportunities that incorporate the use of new technologies [209].

29.9.3 Pandemic Considerations

Video games and electronic devices have always been viewed as non-preferred alternatives to physical activity and real-time interactions. However, during the 2020 COVID-19 pandemic many children were unable to socialize conventionally and participate in group physical activity experiences. AR can be used to augment childhood experiences when others are deemed medically unsafe. While stressing the importance of real-life experiences, AR can provide virtual companionship in closed environments, as well as fun obstacles, and augmented programs for those of all ages.

29.9.4 Accessibility for All

As with all digital advances, AR technology can be expensive and therefore limit accessibility to all. As AR advances, it is essential that this technology be available to all levels of socio-economic strata. Public institutions such as libraries and schools should purchase and make this technology available as part of educational and recreational programs. As the COVID-19 pandemic has illuminated and emphasized the importance of accessibility to technology, it is imperative that this programming is available to all that are interested in exploring the capabilities of this exciting digital media. Additionally, companies that produce this technology should consider donations of equipment and programs to public institutions in order to encourage equal availability.

29.10 Conclusion

As AR technology develops, advances, and becomes more accessible, it may be incorporated into even more aspects of child and young adult life. It is important that parents, educators, pediatricians, and others who work with children understand how AR can enhance and improve various experiences for children such as learning, play, and social interaction. Although an AR experience cannot fully replace an in-person or hands-on activity, it can be used to supplement these experiences when they are not possible in-person or when this technology can offer additional enhancement. For example, during the COVID-19 pandemic when in-person education was limited and taking field trips was not possible, AR could be used to teach chemistry students ionic theory or to bring students to an art museum or even different planets in the solar system [117]. AR can also supplement experiences, such as sports training by providing more accurate assessments and feedback on an athlete's techniques. AR can even be used to make hospital visits for children less intimidating by providing an animated character to help guide them through their experience [12]. AR can aid and enhance a variety of experiences for children, giving them valuable knowledge, skills, and experiences in gaming, education, social media, art, sports, and medicine.

Although AR offers many unique and beneficial opportunities for children, it is important to understand its drawbacks. This technology is often expensive, making it inaccessible for all who want it, further driving gaps in learning opportunities and experiences for children of different socioeconomic backgrounds. Furthermore, while some AR reaches a large audience quickly such as Pokémon GO, many therapeutic and other AR applications show very slow uptake in care and application. Additionally, the AAP recommends limits on children's use of screens and media [108, 182], limiting AR's applications in the lives of children. Caregivers whose children utilize AR technology should also be aware of body dysmorphia concerns that can arise as a result of youth using AR to distort their appearance. Furthermore, caregivers should be mindful when children are using games that require children to go outside and meet other players, as this can lead to interactions with predators, among other dangers [4]. Dangers can also arise as a result of children becoming distracted and immersed in their AR experiences. Caution should be taken when using AR technology and adult supervision is necessary. If children become injured when using AR or experience negative emotional or psychological ramifications of this technology, immediate consultation with a medical professional is important. Additionally, future research is necessary to have a comprehensive understanding of all benefits and risks associated with AR use in a younger age group.

With these considerations in mind, in a rapidly developing digital society, our communities' youngest members have the ability to utilize AR to communicate, socialize, learn, and express themselves. As the technology becomes more accessible and widespread, AR holds the potential to help children in all aspects of their lives – improve their academic performance and understanding, stay in touch with friends, improve their athletic and artistic abilities, help them gain comfort, and understanding during medical experiences. If necessary safety precautions are in place, innovations in the field of AR will allow children and young adults to reap the full benefits of this technology, allowing children to have a new perspective on all they experience.

References

1. Arvanitis, T.N., Petrou, A., Knight, J.F., Savas, S., Sotiriou, S., Gargalakos, M., Gialouri, E.: Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities. Pers. Ubiquitous Comput. **13**, 243–250 (2009). https://doi.org/10.1007/s00779-007-0187-7
2. Carmigniani, J.: Augmented reality: an overview. In: Handbook of Augmented Reality, (eds) by B. Furht Springer New York, (2011) pp. 3–46
3. Liu, S.: Augmented reality market size worldwide 2017–2025. https://www.statista.com/statistics/897587/world-augmented-reality-market-value/ (2020)
4. Serino, M., Cordrey, K., McLaughlin, L., Milanaik, R.L.: Pokémon Go and augmented virtual reality games: a cautionary commentary for parents and pediatricians. Curr. Opin. Pediatr. **28**(5), 673–677 (2016). https://doi.org/10.1097/MOP.0000000000000409
5. Jimenez, D.E., Shah, J., Das, P., Milanaik, R.L.: Health implications of augmented reality games on children and adolescents. In: Augmented Reality Games I, pp. 149–176. Springer International Publishing (2019)
6. Bullock, L.: AR and social media: is augmented reality the future of social media? https://www.forbes.com/sites/lilachbullock/2018/11/16/ar-and-social-media-isaugmented-reality-the-future-of-social-media/?sh=ec2194de141e (2018)
7. Gradl, S., Eskofier, B., Eskofier, D., Mutschler, C., Otto, S.: Virtual and augmented reality in sports: an overview and acceptance study. In: Presented at: 2016 ACM International Joint Conference (2016). https://doi.org/10.1145/2968219.2968572
8. Chang, K.E., Chang, C.T., Hou, H.T., Sung, Y.T., Chao, H.L., Lee, C.M.: Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. Comput. Educ. **71**, 185–197 (2014). https://doi.org/10.1016/j.compedu.2013.09.022
9. Huang, Y., Li, H., Fong, R.: Using augmented reality in early art education: a case study in Hong Kong kindergarten. Early Child Dev. Care. **186**, 1–16 (2015). https://doi.org/10.1080/03004430.2015.1067888
10. Yuen, S.C.-Y., Yaoyuneyong, G., Johnson, E.: Augmented reality: an overview and five directions for AR in education. J. Educ. Technol. Dev. Exch. **4**(1) (2011). https://doi.org/10.18785/jetde.0401.10
11. Campisi, C.A., Li, E.H., Jimenez, D.E., Milanaik, R.L.: Augmented reality in medical education and training: from physicians to patients. In: Geroimenko, V. (ed.) Augmented Reality in Education. Springer, Cham (2020)
12. Debczak, M.: This augmented-reality app makes the hospital experience less scary for Kids | Mental Floss. https://www.mentalfloss.com/article/521854/augmented-realityapp-makes-hospital-experience-less-scary-kids (2017)
13. Bogost: The Tragedy of Pokémon Go. https://www.theatlantic.com/technology/archive/2016/07/the-tragedy-of-pokemongo/490793/ (2016)
14. Piekarski, W., Thomas, B.: ARQuake: the outdoor augmented reality gaming system. Commun. ACM. **45**, 36–38 (2002). https://doi.org/10.1145/502269.502291
15. Thomas, B., Close, B., Donoghue, J., Squires, J., De Bondi, P., Morris, M., Piekarski, W.: ARQuake: an outdoor/indoor augmented reality first person application. In: Fourth International Symposium on Wearable Computers - ISWC, pp. 139–146. IEEE Comput. Soc (2000)
16. Steinbock, D.: The Nokia Revolution: The Story of an Extraordinary Company that Transformed an Industry. AMACOM, New York (2001)
17. Henrysson, A., Billinghurst, M., Ollila, M.: AR tennis. In: ACM SIGGRAPH 2006 Emerging Technologies, p. 1. ACM Press (2006)
18. Bala, S.: The history of augmented reality. https://community.arm.com/developer/toolssoftware/graphics/b/blog/posts/the-history-of-augmented-reality (2018)
19. O'Dea, S.: Smartphones – Statistics & Facts. https://www.statista.com/topics/840/smartphones/ (2020)

20. App Store. https://www.apple.com/app-store/ (2021)
21. Google Play. https://play.google.com/store?hl=en&gl=US (2021)
22. Minecraft Official Site. https://www.minecraft.net/en-us/ (2021)
23. Jurassic World Alive. https://apps.apple.com/us/app/jurassic-world-alive/id1231085864 (2021)
24. Pokémon Go tops Twitter's daily users, sees more engagement than Facebook. https://social.techcrunch.com/2016/07/13/pokemon-go-tops-twitters-daily-users-seesmore-engagement-than-facebook/ (2021)
25. Pokémon GO Revenue and Usage Statistics (2020). https://www.businessofapps.com/data/pokemon-go-statistics/ (2020)
26. Nelson, R.: Mobile users are spending more time in Pokémon GO than Facebook. https://sensortower.com/blog/pokemon-go-usage-data (2016)
27. Pokémon GO. https://apps.apple.com/us/app/pokémon-go/id1094591345 (2021)
28. Pokemon | Description, History, & Facts. https://www.britannica.com/topic/Pokemonelectronic-game (2021)
29. Snyder, C.: A psychologist explains why we need to 'catch "em all" in "Pokémon GO". https://www.businessinsider.com/psychologist-explains-why-we-catch-em-all-pokemongo-2016-8 (2016)
30. Adolescent Social Development. https://www.mentalhelp.net/adolescentdevelopment/social/ (2020)
31. Ge, W., McLaughlin, L., Milanaik, R.: Augmented reality video games: a commentary on what all parents and doctors should know. J. Child Adolesc. Behav. **4**(6) (2016). https://doi.org/10.4172/2375-4494.1000321
32. Althoff, T., White, R.W., Horvitz, E.: Influence of Pokémon Go on physical activity: study and implications. J. Med. Internet Res. **18**, e315 (2016). https://doi.org/10.2196/jmir.6759
33. Wong, F.Y.: Influence of Pokémon Go on physical activity levels of university players: a cross-sectional study. Int. J. Health Geogr. **16** (2017). https://doi.org/10.1186/s12942-017-0080-1
34. Homepage | Minecraft: Education Edition. https://education.minecraft.net/ (2021)
35. Kirriemuir, J., McFarlane, A.: Use of computer and video games in the classroom. In: Presented at: Digital Games Research Conference, University of Utrecht, The Netherlands (2003)
36. Zach, F.J., Tussyadiah, I.P.: To catch them all–the (Un)intended consequences of Pokémon GO on mobility, consumption, and wellbeing. In: Schegg, R., Stangl, B. (eds.) Information and Communication Technologies in Tourism 2017, pp. 217–227. Springer International Publishing, Cham (2017)
37. Bonus, J.A., Peebles, A., Mares, M.-L., Sarmiento, I.G.: Look on the bright side (of media effects): Pokémon Go as a catalyst for positive life experiences. Media Psychol. **21**(2), 263–287 (2018). https://doi.org/10.1080/15213269.2017.1305280
38. Yuhas, A.: Pokémon Go: armed robbers use mobile game to lure players into trap. http://www.theguardian.com/technology/2016/jul/10/pokemon-go-armed-robbers-deadbody (2016)
39. Chong, Y., Sethi, D.K., Loh, C.H.Y., Lateef, F.: Going forward with Pokemon Go. J. Emerg. Trauma. Shock. **11**(4), 243–246 (2018). https://doi.org/10.4103/JETS.JETS_87_17
40. Carsten, P.: In China, fears that Pokemon GO may aid locating military bases. https://www.reuters.com/article/us-nintendo-pokemon-china-idUSKCN0ZV0YD (2016)
41. Das, P., Zhu, M., McLaughlin, L., Bilgrami, Z., Milanaik, R.L.: Augmented reality video games: new possibilities and implications for children and adolescents. Multimodal Technol. Interact. **1**(2), 8 (2017). https://doi.org/10.3390/mti1020008
42. HoloLens: 2–overview, features, and specs | Microsoft HoloLens. https://www.microsoft.com/en-us/hololens/hardware (2021)
43. Hardawar, D.: "Pokémon Go" on HoloLens 2 is a glimpse at the future of AR gaming. https://www.engadget.com/pokemon-go-hololens-2-microsoft-mesh-demo-171537512.html (2021)
44. Augmented Reality. https://developer.apple.com/augmented-reality/ (2021)
45. Williamson, D.A.: Augmented reality in social media – it's growing up, and the pandemic is playing a role. https://www.emarketer.com/content/augmented-reality-insocial-media (2020)
46. Auxier, B., Anderson, M., Perrin, A., Turner, E.: Children's engagement with digital devices, screen time. https://www.pewresearch.org/internet/2020/07/28/childrensengagement-with-digital-devices-screen-time/ (2020)
47. Social Media and Teens. https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFFGuide/Social-Media-and-Teens-100.aspx (2018)
48. Elgersma, C.: Everything you need to know about Snapchat. https://phys.org/news/2018-06-snapchat.html (2018)
49. Snapchat's History: Evolution of Snapchat and timeline. https://www.buycustomgeofilters.com/blog/snapchat-history-and-updated-timeline (2020)
50. Create Your Own Filters & Lenses: • Snapchat. https://www.snapchat.com/create (2021)
51. Lenses for Snap Staff: Mini Golf V2 Snapchat Lens & Filter. https://lensesforsnap.com/mini-golf-v2-snapchat-lens-filter/ (2020)
52. Lenses for Snap Staff: Bowling Snapchat Lens & Filter. https://lensesforsnap.com/bowling-snapchat-lens-filter/ (2019)
53. Doyle, B.: 27 Creative Snapchat campaigns to inspire your snap strategy. https://wallaroomedia.com/blog/snapchat/25-snapchat-campaigns-to-inspire-snapstrategy/ (2019)
54. Williams, R.: Nike runs shoppable Snapchat lens to support women's soccer. https://www.marketingdive.com/news/nike-runs-shoppable-snapchat-lens-to-supportwomens-soccer/554506/ (2019)
55. Lens Studio – Lens Studio by Snap Inc. https://lensstudio.snapchat.com/ (2021)
56. Snap Team: News – Snap Partner Summit: Snap's Augmented Reality Platform – Snap Inc. https://snap.com/en-US/news/post/snap-augmented-reality-platform (2020)
57. Heathman, A.: Facebook's Snapchat-style stories and effects are live. Here's how to use them. https://www.wired.co.uk/article/facebook-filters-like-snapchat (2017)
58. Constine, J.: Facebook builds its own AR games for Messenger video chat. https://social.techcrunch.com/2018/08/08/facebook-messenger-ar-games/ (2018)
59. Johnson, K.: Facebook brings augmented reality to Messenger bots. https://venturebeat.com/2018/05/01/facebook-brings-augmented-reality-to-messengerbots/ (2018)
60. Lee, D.: Facebook now lets anyone make an Instagram face filter. https://www.theverge.com/2019/8/13/20802620/facebook-spark-ar-instagram-story-facefilters (2019)
61. Spark, A.R.: Studio – Create Augmented Reality Experiences | Spark AR Studio. https://sparkar.facebook.com/ar-studio/ (2021)
62. Kriegel, E.R., Lazarevic, B., Athanasian, C.E., Milanaik, R.L.: TikTok, Tide Pods and Tiger King: health implications of trends taking over pediatric populations. Curr. Opin. Pediatr. **33**(1), 170–177 (2021). https://doi.org/10.1097/MOP.0000000000000989
63. Marvellous: The 15 TikTok filters and effects you want to use. https://blog.iconosquare.com/the-15-tiktok-filters-and-effects-you-want-to-use/ (2020)
64. Whitney, L.: The Best LiDAR Apps for Your iPhone 12 Pro or iPad Pro | PCMag. https://www.pcmag.com/how-to/the-best-lidar-apps-for-your-iphone-12-pro-or-ipad-pro (2021)
65. Gartenberg, C.: TikTok debuts first augmented reality filter that uses iPhone 12 Pro's LIDAR camera. https://www.theverge.com/2021/1/7/22217465/tiktok-augmented-realityfilter-iphone-12-pros-lidar-camera (2021)
66. @TikTokComms: TikTokComms on Twitter. https://twitter.com/TikTokComms/status/1346828242506969089 (2021)

67. Find Your Audience | Pinterest Business. https://business.pinterest.com/audience/ (2021)
68. Pinterest's new AR tech lets you try on makeup before you buy. https://newsroom.pinterest.com/en/post/pinterests-new-ar-tech-lets-you-try-on-makeupbefore-you-buy (2020)
69. Anderson, M., Jiang, J.: Teens, social media & technology 2018. https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ (2018)
70. Auxier, B.: 8 facts about Americans and Instagram. https://www.pewresearch.org/facttank/2020/10/21/8-facts-about-americans-and-instagram/ (2020)
71. Clement, J.: U.S. TikTok users by age 2020. https://www.statista.com/statistics/1095186/tiktok-us-users-age/ (2020)
72. Clement, J.: Social networks used by teens in U.S. https://www.statista.com/statistics/250172/social-network-usage-of-us-teens-and-youngadults/ (2020)
73. What should I do if I'm being bullied, harassed or attacked by someone on Facebook? | Facebook Help Center. https://www.facebook.com/help/116326365118751 (2021)
74. Reporting harassment or bullying on Instagram | Instagram Help Center. https://www.facebook.com/help/instagram/547601325292351 (2021)
75. Report Abuse on Snapchat. https://support.snapchat.com/en-US/a/report-abuse-in-app (2021)
76. Body dysmorphic disorder. https://www.mayoclinic.org/diseases-conditions/bodydysmorphic-disorder/symptoms-causes/syc-20353938 (2019)
77. Rajanala, S., Maymone, M.B.C., Vashi, N.A.: Selfies-living in the era of filtered photographs. JAMA Facial Plast. Surg. **20**, 443–444 (2018). https://doi.org/10.1001/jamafacial.2018.0486
78. Muhammad, Z.: The success of augmented reality in social media. https://www.digitalinformationworld.com/2019/05/ar-in-social-media.html (2019)
79. Statt, N.: Facebook's first 'smart glasses' will be Ray-Bans. Coming next year. https://www.theverge.com/2020/9/16/21439929/facebook-ar-smart-glasses-ray-banannouncement (2020)
80. Announcing Project Aria: A research project on the future of wearable AR. https://about.fb.com/news/2020/09/announcing-project-aria-a-research-project-on-thefuture-of-wearable-ar/ (2020)
81. Marr, B.: The important risks and dangers of virtual and augmented reality. https://www.forbes.com/sites/bernardmarr/2019/07/17/the-important-risks-and-dangersof-virtual-and-augmented-reality/ (2019)
82. O'Keeffe, G.S., Clarke-Pearson, K.: Council on communications and media: the impact of social media on children. Adolesc. Fam. Pediatr. **127**(4), 800–804 (2011). https://doi.org/10.1542/peds.2011-0054
83. Soltani, P., Morice, A.H.P.: Augmented reality tools for sports education and training. Comput. Educ. **155**, 103923 (2020). https://doi.org/10.1016/j.compedu.2020.103923
84. Fountaine, C.J., Springer, E.J., Sward, J.R.: A descriptive study of objectively measured Pokémon GO playtime in college students. Int. J. Exerc. Sci. **11**(7), 526–532 (2018)
85. Fery, Y.A., Ponserre, S.: Enhancing the control of force in putting by video game training. Ergonomics. **44**, 1025–1037 (2001). https://doi.org/10.1080/00140130110084773
86. Kelly, P., O'Connor, N.E.: Visualisation of tennis swings for coaching. In: 2012 13th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), pp. 1–4 (2012)
87. Soga, M., Nishino, T., Taki, H.: Proposal and development of motion navigator enabling learners to observe expert's motion from expert's viewpoint by augmented reality. In: Presented at: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (2016). https://doi.org/10.1007/978-3-642-23854-3_5
88. Kajastila, R., Holsti, L., Hämäläinen, P.: The augmented climbing wall: high-exertion proximity interaction on a wall-sized interactive surface. In: Presented at: 2016 CHI Conference on Human Factors in Computing Systems (2016). https://doi.org/10.1145/2858036.2858450
89. McIlroy, P.: Hawk-Eye: augmented reality in sports broadcasting and officiating. In: 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. xiv–xiv (2008). https://doi.org/10.1109/ISMAR.2008.4637309
90. Hawk-Eye. https://www.hawkeyeinnovations.com/ (2018)
91. Escarvajal-Rodríguez, J.C.: Pokémon GO and its influence on Spanish Facebook users. Educ. Fis. Deport. **3**, 38–49 (2018). https://doi.org/10.5672/apunts.2014-0983.es.(2018/3).133.03
92. Guthold, R., Stevens, G.A., Riley, L.M., Bull, F.C.: Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. Lancet. Child Adolesc. Heal. **4**, 23–35 (2020). https://doi.org/10.1016/S2352-4642(19)30323-2
93. Araki, H., Fukuda, H., Motoki, T., Takeuchi, T., Ohta, N., Adachi, R., Masuda, H., Kado, Y., Mita, Y., Mizukami, D., Kakeya, N.: "HADO" as Techno Sports was born by the fusion of IT technology and sports. In: EPiC Series in Engineering, p. 40. EasyChair (2018)
94. He, Z., Wu, L., Li, X.: When art meets tech: the role of augmented reality in enhancing museum experiences and purchase intentions. Tour. Manag. **68**, 127–139 (2018). https://doi.org/10.1016/j.tourman.2018.03.003
95. Dieck, M.C.T., Jung, T.H., Dieck, D.T.: Enhancing art gallery visitors' learning experience using wearable augmented reality: generic learning outcomes perspective. Curr. Issues Tour. **21**, 2014–2034 (2018). https://doi.org/10.1080/13683500.2016.1224818
96. Color Alive by Crayola | Action Coloring Pages. https://www.crayola.com/application/coloralive/panels.html (2021)
97. uMake – 3D Modeling CAD for the iPad and iPhone. https://www.umake.xyz/ (2020)
98. Light Space – Logical Animal. https://logicalanimal.com/lightspaceapp (2020)
99. World Brush. https://worldbrush.net/ (2017)
100. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. J. Sci. Educ. Technol. **18**, 7–22 (2009). https://doi.org/10.1007/s10956-008-9119-1
101. Georgiou, Y., Kyza, E.A.: Relations between student motivation, immersion and learning outcomes in location-based augmented reality settings. Comput. Human Behav. **89**, 173–181 (2018). https://doi.org/10.1016/j.chb.2018.08.011
102. Khan, T., Johnston, K., Ophoff, J.: The impact of an augmented reality application on learning motivation of students. Adv. Human-Computer Interact. (2019). https://doi.org/10.1155/2019/7208494
103. Thomas, B.H.: The future of augmented reality gaming. Emerg. Technol. Augment. Real. Interfaces Des, 367–382 (2006). https://doi.org/10.4018/978-1-59904-066-0.ch018
104. Brand, J.E., Todhunter, S., Jervis, J.: Digital Australia Report 2018. **43** (2017)
105. EDUCAUSE: 2019 Horizon Report. https://library.educause.edu/resources/2019/4/2019-horizon-report (2019)
106. Martínez, A.A., Benito, J.R.L., González, E.A., Ajuria, E.B.: An experience of the application of augmented reality to learn English in infant education. In: 2017 International Symposium on Computers in Education (SIIE), pp. 1–6, Lisbon (2017)
107. QuiverVision. https://quivervision.com (2020)
108. Hill, D., Ameenuddin, N., Chassiakos, Y.R., Cross, C., Radesky, J., Hutchinson, J., Boyd, R., Mendelson, R., Moreno, M.A., Smith, J., Swanson, W.S.: Media use in school-aged children and ado-

lescents. Pediatrics. **138**(5), e20162592 (2016b). https://doi.org/10.1542/peds.2016-2592
109. Billinghurst, M., Kato, H., Poupyrev, I.: The MagicBook: a transitional AR interface. Comput. Graph. **25**(5), 745–753 (2001). https://doi.org/10.1016/S0097-8493(01)00117-0
110. Karamanoli, P., Tsinakos, A.: A mobile augmented reality application for primary school's history. IOSR J. Res. Method Educ. **6**(6), 56–65 (2016)
111. Saez-Lopez, J.M., Cozar-Gutierrez, R.: Iberian cultures and augmented reality: studies in elementary school education and initial teacher training. In: Geroimenko, V. (ed.) Augmented Reality in Education. Springer, Cham (2020)
112. Surakitbanharn, C.: Augmented reality to engage remote learning students with Banuba. https://dschool.stanford.edu/news-events/augmented-reality-to-engage-remote-learningstudents (2020)
113. Bradbury, J.: 3 ways to use augmented reality in hybrid learning. https://www.teachercast.net/3-ways-to-use-augmented-reality-in-hybrid-learning/ (2020)
114. Seek Education | Interactive & Immersive Learning Through App-less Augmented Reality. https://education.seekxr.com/ (2020)
115. Mozelius, P., Jaldemark, J., Eriksson Bergström, S., Sundgren, M.: The concept of 'Bringing your own device' in scaffolded and augmented education. In: Geroimenko, V. (ed.) Augmented Reality in Education. Springer, Cham (2020)
116. Kohen-Vacs, D., Ronen, M., Cohen, S.: Mobile treasure hunt games for outdoor learning. Bull. Tech. Comm. Learn. Technol. **14**(4), 24–26 (2012)
117. Bring your lessons to life with Expeditions | Google for Education. https://edu.google.com/products/vr-ar/expeditions/ (2021)
118. Lytridis, C., Tsinakos, A., Kazanidis, I.: ARTutor–An augmented reality platform for interactive distance learning. Educ. Sci. **8**(1), 6 (2018). https://doi.org/10.3390/educsci8010006
119. Fjeld, M., Fredriksson, J., Ejdestig, M., Duca, F., Bötschi, K., Voegtli, B., Juchli, P.: Tangible user interface for chemistry education: comparative evaluation and re-design. In: Conference on Human Factors in Computing Systems – Proceedings, pp. 805–808 (2007)
120. Hands-on with augmented reality in remote classrooms. https://www.purdue.edu/newsroom/releases/2020/Q2/hands-on-with-augmented-realityin-remote-classrooms.html (2020)
121. ARLOON: Apps Educativas con Realidad Aumentada. http://www.arloon.com/en/ (2019)
122. Coffin, C., Bostandjiev, S., Ford, J., Hollerer, T.: Enhancing classroom and distance learning through augmented reality. In: Herrington, J., Montgomerie, C. (eds.) Proceedings of EdMedia + Innovate Learning 2010, pp. 1140–1147. Association for the Advancement of Computing in Education (AACE), Toronto (2010)
123. Ferrer-Torregrosa, J., Jiménez-Rodríguez, M.Á., Torralba-Estelles, J., Garzón-Farinós, F., Pérez-Bermejo, M., Fernández-Ehrling, N.: Distance learning acts and flipped classroom in the anatomy learning: comparative study of the use of augmented reality, video and notes. BMC Med. Educ. **16**, 230 (2016). https://doi.org/10.1186/s12909-016-0757-3
124. Schaffhauser, D.: 9 Amazing Uses for VR and AR in College Classrooms. https://campustechnology.com/articles/2019/05/15/9-amazing-uses-for-vr-and-ar-incollege-classrooms.aspx (2019)
125. Contreras, R.: Drones help map Japanese internment camp in eastern Colorado. https://durangoherald.com/articles/275397 (2019)
126. Husson University to Demonstrate New Augmented Reality App for Theatrical Set Designers. https://www.husson.edu/news/2019/04/husson-university-to-demonstratenew-augmented-reality-app-for-theatrical-set-designers (2019)
127. AR/VR Experiences at the Gabelli School. https://www.fordham.edu/info/23906/fulltime_mba/11931/arvr_experiences_at_the_gabelli_school (2020)
128. Smith, K., Hottinger, P.: Gotta catch 'em all. In: A Case Study About Cal Poly Pomona's Pokemon GO AR Orientation. Rowman & Littlefield (2018)
129. The Penn State Obelisk. https://sites.psu.edu/obelisk/ (2018)
130. MarketScale: Augmented reality R&D surges during remote learning. https://marketscale.com/industries/education-technology/augmented-reality-rd/ (2020)
131. Stott, L.: Augmented reality remote learning with Hololens2. https://techcommunity.microsoft.com/t5/educator-developer-blog/augmented-realityremote-learning-with-hololens2/ba-p/1608687 (2020)
132. Ma, M., Fallavollita, P., Seelbach, I., Von Der Heide, A.M., Euler, E., Waschke, J., Navab, N.: Personalized augmented reality for anatomy education. Clin. Anat. **29**(4), 446–453 (2016). https://doi.org/10.1002/ca.22675
133. Gunner Goggles. https://www.us.elsevierhealth.com/gunnergoggles (2021)
134. Pantelidis, P., Chorti, A., Papagiouvanni, I., Paparoidamis, G., Drosos, C., Panagiotakopoulos, T., Lales, G., Sideris, M.: Virtual and augmented reality in medical education. Med. Surg. Educ. – Past Present Futur. (2017). https://doi.org/10.5772/intechopen.71963
135. Bifulco, P., Narducci, F., Vertucci, R., Ambruosi, P., Cesarelli, M., Romano, M.: Telemedicine supported by augmented reality: an interactive guide for untrained people in performing an ECG test. Biomed. Eng. Online. **13**, 153 (2014). https://doi.org/10.1186/1475-925X-13-153
136. Balian, S., McGovern, S.K., Abella, B.S., Blewer, A.L., Leary, M.: Feasibility of an augmented reality cardiopulmonary resuscitation training system for health care providers. Heliyon. **5**, e02205 (2019). https://doi.org/10.1016/j.heliyon.2019.e02205
137. Ingrassia, P.L., Mormando, G., Giudici, E., Strada, F., Carfagna, F., Lamberti, F., Bottino, A.: Augmented reality learning environment for basic life support and defibrillation training: usability study. J. Med. Internet Res. **22**, e14910 (2020). https://doi.org/10.2196/14910
138. Munzer, B.W., Khan, M.M., Shipman, B., Mahajan, P.: Augmented reality in emergency medicine: a scoping review. J. Med. Internet Res. **21**(4), e12368 (2019). https://doi.org/10.2196/12368
139. Lyall, F.C., Clamp, P.J., Hajioff, D.: Smartphone speech-to-text applications for communication with profoundly deaf patients. J. Laryngol. Otol. **130**(1), 104–106 (2016). https://doi.org/10.1017/S0022215115003151
140. Franz, A.M., Horiguchi, K.: United States Patent 6,282,507: method and apparatus for interactive source language expression recognition and alternative hypothesis presentation and selection (2001)
141. Winner, E., Platt, P.: Augmented reality imaging system for cosmetic surgical procedures. https://patentimages.storage.googleapis.com/d0/aa/3f/712786666aef4c/US20170119471A1.pdf (2017)
142. Marbury, D.: What does the future hold for AR and VR in healthcare? https://healthtechmagazine.net/article/2020/11/what-does-future-hold-ar-and-vrhealthcare (2020)
143. Calle-Bustos, A.-M., Juan, M.-C., García-García, I., Abad, F.: An augmented reality game to support therapeutic education for children with diabetes. PLoS One. **12**, e0184645 (2017). https://doi.org/10.1371/journal.pone.0184645
144. Livingston, M.A., Rosenblum, L.J., Brown, D.G., Schmidt, G.S., Julier, S.J., Baillot, Y., Swan, J.E., Ai, Z., Maassel, P.: Military applications of augmented reality. In: Furht, B. (ed.) Handbook of Augmented Reality, pp. 671–706. Springer, New York (2011)
145. Wang, X., Dunston, P.S.: Design, strategies, and issues towards an augmented reality based construction training platform. Electron. J. Inf. Technol. Constr. **12**, 363–380 (2007)

146. Haritos, T., Macchiarella, N.D.: A mobile application of augmented reality for aerospace maintenance training. In: Presented at: 24th Digital Avionics Systems Conference, p. 5.B.3-5.1 (2005)
147. Gonzalez, M.: Augmented and virtual reality will benefit the disabled people. https://www.thegenius.ca/augmented-virtual-reality-benefit-disabled-people/ (2017)
148. Chen, C.H., Lee, I.J., Lin, L.Y.: Augmented reality-based video-modeling storybook of nonverbal facial cues for children with autism spectrum disorder to improve their perceptions and judgments of facial expressions and emotions. Comput. Human Behav. **55**, 477–485 (2016). https://doi.org/10.1016/j.chb.2015.09.033
149. Rashid, Z., Melià-Seguí, J., Pous, R., Peig, E.: Using augmented reality and internet of things to improve accessibility of people with motor disabilities in the context of smart cities. Futur. Gener. Comput. Syst. **76**, 248–261 (2017). https://doi.org/10.1016/j.future.2016.11.030
150. Wu, H.-K., Lee, S.W.-Y., Chang, H.-Y., Liang, J.-C.: Current status, opportunities and challenges of augmented reality in education. Comput. Educ. **62**, 41–49 (2013). https://doi.org/10.1016/j.compedu.2012.10.024
151. Tinti-Kane, H., Vahey, P.: xR in EDU Survey: Benchmarking Adoption Trends in K12 and Higher Education October 2018 EdTech Times (2018)
152. Edge, R., Butcher, R.: Vein Illumination Devices in Long-Term and Acute Care Settings: A Review of Clinical Effectiveness, Cost-Effectiveness and Guidelines. Canadian Agency for Drugs and Technologies in Health Rapid Response Report (2020)
153. Park, J.M., Kim, M.J., Yim, H.W., Lee, W.C., Jeong, H., Kim, N.J.: Utility of near-infrared light devices for pediatric peripheral intravenous cannulation: a systematic review and meta-analysis. Eur. J. Pediatr. **175**, 1975–1988 (2016). https://doi.org/10.1007/s00431-016-2796-5
154. Sadeghi, A.H., el Mathari, S., Abjigitova, D., Maat, A.P.W.M., Taverne, Y.J.H.J., Bogers, A.J.J.C., Mahtab, E.A.F.: Current and future applications of virtual, augmented, and mixed reality in cardiothoracic surgery. Ann. Thorac. Surg. (2020). https://doi.org/10.1016/j.athoracsur.2020.11.030
155. Khor, W.S., Baker, B., Amin, K., Chan, A., Patel, K., Wong, J.: Augmented and virtual reality in surgery–the digital surgical environment: applications, limitations and legal pitfalls. Ann. Transl. Med. **4** (2016). https://doi.org/10.21037/atm.2016.12.23
156. Pratt, P., Ives, M., Lawton, G., Simmons, J., Radev, N., Spyropoulou, L., Amiras, D.: Through the HoloLensTM looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. Eur. Radiol. Exp. **2**, 2 (2018). https://doi.org/10.1186/s41747-017-0033-2
157. Healthysimulation: Augmented Reality in Medicine | Healthcare Simulation. https://www.healthysimulation.com/augmented-reality-in-medicine/ (2021)
158. Al Omran, Y., Abdall-Razak, A., Sohrabi, C., Borg, T.-M., Nadama, H., Ghassemi, N., Oo, K., Ghanem, A.: Use of augmented reality in reconstructive microsurgery: a systematic review and development of the augmented reality microsurgery score. J. Reconstr. Microsurg. **36** (2019). https://doi.org/10.1055/s-0039-3401832
159. Przystalski, K.: How is augmented reality used in healthcare? | Codete Blog. https://codete.com/blog/augmented-reality-in-healthcare/ (2020)
160. Ponce, B.A., Brabston, E.W., Zu, S., Watson, S.L., Baker, D., Winn, D., Guthrie, B.L., Shenai, M.B.: Telemedicine with mobile devices and augmented reality for early postoperative care. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4411–4414 (2016)
161. DeTrempe, K.: Virtual reality alleviates pain, anxiety for pediatric patients. http://med.stanford.edu/news/all-news/2017/09/virtual-reality-alleviates-pain-anxiety-forpediatric-patients.html (2017)
162. CHARIOT Program – Stanford Children's Health. https://www.stanfordchildrens.org/en/innovation/chariot (2021)
163. Caruso, T.J., Madill, M., Sidell, D., Meister, K., Wang, E., Menendez, M., Kist, M.N., Rodriguez, S.: Using augmented reality to reduce fear and promote cooperation during pediatric otolaryngologic procedures. Laryngoscope. **131**, E1342–E1344 (2021). https://doi.org/10.1002/lary.29098
164. Mott, J., Bucolo, S., Cuttle, L., Mill, J., Hilder, M., Miller, K., Kimble, R.M.: The efficacy of an augmented virtual reality system to alleviate pain in children undergoing burns dressing changes: a randomised controlled trial. Burns. **34**, 803–808 (2008). https://doi.org/10.1016/j.burns.2007.10.010
165. Marto, A., Almeida, H.A., Gonçalves, A.: Using Augmented Reality in Patients with Autism: A Systematic Review, vol. 34, pp. 454–463. Springer Netherlands (2019). https://doi.org/10.1007/978-3-030-32040-9_46
166. Chinthammit, W., Merritt, T., Pedersen, S., Williams, A., Visentin, D., Rowe, R., Furness, T.: Ghostman: augmented reality application for telerehabilitation and remote instruction of a novel motor skill. https://www.hindawi.com/journals/bmri/2014/646347/ (2014)
167. LaPiana, N., Duong, A., Lee, A., Alschitz, L., Silva, R.M.L., Early, J., Bunnell, A., Mourad, P.: Acceptability of a mobile phone–based augmented reality game for rehabilitation of patients with upper limb deficits from stroke: case study. JMIR Rehabil. Assist. Technol. **7**, e17822 (2020). https://doi.org/10.2196/17822
168. De Witte, N.A.J., Scheveneels, S., Sels, R., Debard, G., Hermans, D., Van Daele, T.: Augmenting exposure therapy: mobile augmented reality for specific phobia. Front. Virtual Real. **1** (2020). https://doi.org/10.3389/frvir.2020.00008
169. Juan, M.C., Alcaniz, M., Monserrat, C., Botella, C., Banos, R.M., Guerrero, B.: Using augmented reality to treat phobias. IEEE Comput. Graph. Appl. **25**, 31–37 (2005). https://doi.org/10.1109/MCG.2005.143
170. Suso-Ribera, C., Fernández-Álvarez, J., García-Palacios, A., Hoffman, H.G., Bretón- López, J., Baños, R.M., Quero, S., Botella, C.: Virtual reality, augmented reality, and in vivo exposure therapy: a preliminary comparison of treatment efficacy in small animal phobia. Cyberpsychol. Behav. Soc. Netw. **22**, 31–38 (2019). https://doi.org/10.1089/cyber.2017.0672
171. Boghossian, J., Armstrong, D.G., Joseph, B.: Dosing distraction in the world of augmented reality. https://blog.oup.com/2016/10/distraction-safety-augmented-reality/ (2016)
172. Carraro, G.U., Cortes, M., Edmark, J.T., Ensor, J.R.: The peloton bicycling simulator. In: Proceedings of the Third Symposium on Virtual Reality Modeling Language, pp. 63–70. ACM Press (1998)
173. Peloton® | Workouts Streamed Live & On-Demand. https://www.onepeloton.com/ (2021)
174. MIRROR | How It Works. https://www.mirror.co/workouts?utm_source=google_ads&utm_medium=search_cpc&utm_campaign=brand_mirror&utm_content=mirror_workout&utm_term=mirror workout (2021)
175. WISQARS: Leading causes of death visualization tool. https://wisqarsviz. cdc.gov:8006/lcd/home (2018)
176. National Highway Traffic Safety Administration: Distraction | National Highway Traffic Safety Administration (NHTSA). https://one.nhtsa.gov/Research/Human-Factors/Distraction (2021)
177. Faccio, M., McConnell, J.J.: Death by Pokémon GO: The economic and human cost of using apps while driving. J. Risk Insur. **87**, 815–849 (2020). https://doi.org/10.1111/jori.12301

178. Ayers, J.W., Leas, E.C., Dredze, M., Allem, J.P., Grabowski, J.G., Hill, L.: Pokémon GO-A new distraction for drivers and pedestrians. JAMA Intern Med. **176**(12), 1865–1866 (2016). https://doi.org/10.1001/jamainternmed.2016.6274
179. Hurst, W.: Why doesn't Pokemon go work when I'm driving anymore? https://www.hg.org/legal-articles/why-doesn-t-pokemon-go-work-when-i-m-drivinganymore-40923#::text=Manyareasking why%3F,greaterthanapproximately25MPH (2021)
180. ARDDES (Distracted Driving Prevention) | PEERS Foundation. https://peersfoundation.org/programs/arddes-distracted-driving-prevention/ (2021)
181. Industry Experts Concerned About Prolonged Use of AR/VR Technology | UL. https://www.ul.com/news/augmented-reality-technology (2020)
182. Hill, D., Ameenuddin, N., Chassiakos, Y.R., Cross, C., Radesky, J., Hutchinson, J., Boyd, R., Mendelson, R., Moreno, M.A., Smith, J., Swanson, W.S.: Media and young minds. Pediatrics. **138**(5), e20162591 (2016). https://doi.org/10.1542/peds.2016-2591
183. Porter, D.: Blue light and digital eye strain – American Academy of Ophthalmology. https://www.aao.org/eye-health/tips-prevention/blue-light-digital-eye-strain (2020)
184. Bedinghaus, T.: Asthenopia (Eye Strain): symptoms, causes, and treatment. https://www.verywellhealth.com/do-you-suffer-from-asthenopia-or-tired-eyes-3421982 (2020)
185. Bauer, K.A., Lip, G.Y.: Overview of the causes of venous thrombosis – UpToDate. https://www.uptodate.com/contents/overview-of-the-causes-of-venous-thrombosis (2021)
186. Staff Writer: Video gaming injuries are on the rise – Health Nexus. https://thehealthnexus.org/video-gaming-injuries-are-on-the-rise/ (2020)
187. Penumudi, S.A., Kuppam, V.A., Kim, J.H., Hwang, J.: The effects of target location on musculoskeletal load, task performance, and subjective discomfort during virtual reality interactions. Appl. Ergon. **84**, 103010 (2020). https://doi.org/10.1016/j.apergo.2019.103010
188. The Register-Guard: OSU study assesses injury risk due to use of VR devices. https://www.registerguard.com/news/20200107/osu-study-assesses-injury-risk-due-touse-of-vr-devices (2020)
189. Ellis, L.A., Lee, M.D., Ijaz, K., Smith, J., Braithwaite, J., Yin, K.: COVID-19 as 'Game changer' for the physical activity and mental well-being of augmented reality game players during the pandemic: mixed methods survey study. J. Med. Internet Res. **22**(12), e25117 (2020). https://doi.org/10.2196/25117
190. Evren, C.: A brief update of disordered gaming. J. Psychiatry Neurol. Sci. **3**(33), 223–227 (2020). https://doi.org/10.14744/DAJPNS.2020.00085
191. World Health Organization: Addictive behaviours: gaming disorder. https://www.who.int/news-room/q-a-detail/addictive-behaviours-gaming-disorder (2018)
192. Coyne, S.M., Radesky, J., Collier, K.M., Gentile, D.A., Linder, J.R., Nathanson, A.I., Rasmussen, E.E., Reich, S.M., Rogers, J.: Parenting and digital media. Pediatrics. **140**(Supplement 2), S112–S116 (2017). https://doi.org/10.1542/peds.2016-1758N
193. Piaget, J.: The development of object concept. In: Cook, M. (ed.) The Construction of Reality in the Child, pp. 3–96. Basic Books, New York (1954)
194. Franklin, D.M., Castle, K., Walton-Mouw, R.: Early-detection and treatment of torticollis in infants using augmented reality. In: Ahram, T., Taiar, R., Gremeaux-Bader, V., Aminian, K. (eds.) Advances in Intelligent Systems and Computing, vol. 1152, pp. 340–346. Springer (2020)
195. Nishizaki, M.: Visualization of infants' behavior at home using augmented reality for promotion of development and prevention of accidents. In: Cocchiarella, L. (ed.) Advances in Intelligent Systems and Computing, vol. 809, pp. 930–941. Springer (2019)
196. Christakis, D.: Virtual violence. Pediatrics. **138**(2), e20161298 (2016). https://doi.org/10.1542/peds.2016-1298
197. Sheikh, K.: Beyond gaming: 10 other fascinating uses for virtual-reality. https://www.livescience.com/53392-virtual-reality-tech-uses-beyond-gaming.html (2016)
198. Gent, E.: Are virtual reality headsets safe for children? https://www.scientificamerican.com/article/are-virtual-reality-headsets-safe-forchildren/#::text=The Oculus Rift and Samsung's,use the headset at all (2016)
199. Stein, S.: Google 3D animals: how to use the cool AR feature at home. https://www.cnet.com/how-to/google-3d-animals-how-to-use-cool-ar-feature-at-homelist-of-objects/ (2020)
200. Stein, S.: AR Mickey Mouse brings a bit of Disney World to your home. https://www.cnet.com/news/ar-mickey-mouse-brings-a-bit-of-disney-world-to-yourhome/ (2020)
201. Smith, T.: Disney characters make virtual memories in Times Square. https://disneyparks.disney.go.com/blog/2011/11/disney-characters-make-virtualmemories-in-times-square/ (2011)
202. Kiefaber, D.: National Geographic lets you pet Dinosaurs at the Mall. https://www.adweek.com/creativity/national-geographic-lets-you-pet-dinosaurs-mall-136591/ (2011)
203. Gentile, D.A., Reimer, R.A., Nathanson, A.I., Walsh, D.A., Eisenmann, J.C.: Protective effects of parental monitoring of children's media use a prospective study. JAMA Pediatr. **168**(5), 479–484 (2014). https://doi.org/10.1001/jamapediatrics.2014.146
204. Lopez-Fernandez, O., Williams, A.J., Griffiths, M.D., Kuss, D.J.: Female gaming, gaming addiction, and the role of women within gaming culture: a narrative literature review. Front. Psychiatry. **10**, 454 (2019). https://doi.org/10.3389/fpsyt.2019.00454
205. Hou, L., Wang, X.: A study on the benefits of augmented reality in retaining working memory in assembly tasks: a focus on differences in gender. Autom. Constr. **32**, 38–45 (2013). https://doi.org/10.1016/j.autcon.2012.12.007
206. Mileva, G.: How virtual reality promotes gender equality. https://arpost.co/2018/09/07/how-virtual-reality-promotes-gender-equality/ (2018)
207. The Machine to be Another – BeAnotherLab. http://beanotherlab.org/home/work/tmtba/ (2020)
208. Dirin, A., Alamäki, A., Suomala, J.: Gender differences in perceptions of conventional video, virtual reality and augmented reality. Int. J. Interact. Mob. Technol. **13**, 93–103 (2019). https://doi.org/10.3991/ijim.v13i06.10487
209. Schaffernak, H., Moesl, B., Vorraber, W., Koglbauer, I.V.: Potential augmented reality application areas for pilot education: an exploratory study. Educ. Sci. **10**(4), 86 (2020). https://doi.org/10.3390/educsci10040086

Elana Kriegel received her Bachelor of Science degree from Cornell University in 2020. She is currently working as a research assistant in the Division of Developmental and Behavioral Pediatrics at Cohen Children's Medical Center in New York. She plans to attend medical school.

Bojan Lazarevic graduated from Princeton University in 2020 with a Bachelor of Arts in molecular biology. He currently works as a research assistant in the Developmental and Behavioral Pediatrics Division of Cohen Children's Medical Center in Lake Success, NY. He plans on attending medical school.

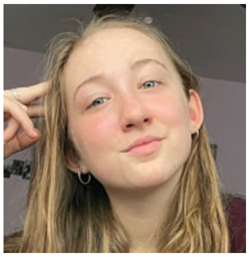
Deborah Feifer is an undergraduate Biology major and Philosophy minor at Dartmouth College. Her research with the Dartmouth Institute for Health Policy and Clinical Practice focuses on physicians' advance care planning habits and ventilator allocation policies during the COVID-19 pandemic. At Cohen Children's Medical Center, her research focuses on the TikTok mobile application and sources of support for custodial grandparents.

Christian Athanasian earned a degree in Human Evolutionary Biology from Harvard College in 2020. He is currently performing research within the Division of Developmental and Behavioral Pediatrics at Cohen Children's Medical Center, and is planning to attend medical school. His research interests include the intersection between everyday lifestyle choices and medicine.

Nelson Chow is currently a sophomore at Princeton University concentrating in Molecular Biology. He started as a research intern in the Division of Developmental and Behavioral Pediatrics at Cohen Children's Medical Center in May 2020.

Jamie P. Sklar has received her MD from SUNY Buffalo School of Medicine in 2016. She completed her residency in General Pediatrics at Cohen Children's Medical Center in 2019. She is currently a fellow in Developmental and Behavioral Pediatrics at Cohen Children's Medical Center and is expected to graduate in 2022.

Yaa Asante received her MD from the University of Ghana School of Medicine and Dentistry at Accra, Ghana, in 2012. She worked as a general practitioner in Ghana before completing her Pediatric residency in Harlem Hospital Center, NY. She is currently pursuing a fellowship in Developmental and Behavioral Pediatrics in Cohen Children's Medical Center following her passion for child development.



Clara Goldman has served as a young adult focus group leader for multiple research projects involving children and high school students. Her current focuses include art and youth during the COVID-19 pandemic and technology for young adult use.

Ruth Milanaik, D.O., graduated from the New York College of Osteopathic Medicine in 1997. She completed her residency at Winthrop University Hospital, and her fellowship at Schneider's Children's Hospital. She is board certified in general pediatrics, and developmental and behavioral pediatrics. She is the director of the neonatal follow-up program and associate professor at the Hofstra Zucker School of Medicine.

Augmented Reality-Assisted Healthcare Exercising Systems

30

Soh Khim Ong, M. Y. Zhao, and Andrew Yeh Ching Nee

Contents

30.1 Augmented Reality-Assisted Healthcare Exercising Systems 743

30.2 Related Works 744
30.2.1 Upper-Extremity Healthcare Exercising Systems 744
30.2.2 Upper-Extremity Movement Tracking Methodologies 747
30.2.3 Assessment of User's Upper-Extremity Functions Recover 748
30.2.4 Discussion 749

30.3 System Overview 749

30.4 Monitoring Module 750
30.4.1 Motion Tracking 751
30.4.2 AR-Based Data Glove 751
30.4.3 Vibration Wrist Band 751

30.5 Exercise Module and Scoring Module 753
30.5.1 Flaccid Stage 753
30.5.2 Synergy Stage 754
30.5.3 Activities of Daily Living Stage 755

30.6 Usability Experiment 758
30.6.1 Method 758
30.6.2 Participants 758
30.6.3 Experiment Procedure 758
30.6.4 Design of Pre-test and Post-test 759
30.6.5 Results and Discussion 759

30.7 Conclusion and Future Work 760

References 760

S. K. Ong (✉) · A. Y. C. Nee
Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore
e-mail: mpeongsk@nus.edu.sg; mpeneeyc@nus.edu.sg

M. Y. Zhao
Mechanical Engineering Department, National University of Singapore, Singapore, Singapore
e-mail: a0086310@u.nus.edu

Abstract

The number of people with low upper-extremity functions is increasing due to the sedentary lifestyle, muscular disuse, and aging of population. Therefore, healthcare exercising systems that aim to enhance upper-extremity skills are desirable. The improvement of motor functions is an ordered process, and hence, the development of an upper-extremity training plan with stages with respect to the capability of the users is an important issue. Augmented Reality (AR) -assisted motor-skills training applications have been proven to be effective. This chapter discusses the importance of providing AR-assisted healthcare exercises in stages. The chapter reviews the current AR-assisted healthcare exercising systems and makes a comparison with virtual reality-based systems as well as conventional systems. A novel AR-assisted Three-stage Healthcare Exercising system (ARTHE) is presented to demonstrate stage-based AR-assisted systems for training activities of daily living.

Keywords

Augmented reality · Upper-extremity · Healthcare · Activities of daily living · Motor ability improvement stage

30.1 Augmented Reality-Assisted Healthcare Exercising Systems

The term "upper-extremity" refers to the region from the deltoid to the hand, including shoulder, arm, and hand. The upper-extremity plays an important role in daily life. However, upper-extremity functions may be reduced by a sedentary lifestyle, muscular disuse, bedridden status, neurological diseases, and some other factors [1]. Moreover, as the amount of muscle mass and voluntary strength decrease when an individual ages [2], it is foreseeable that the number of people

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_30

with decreased upper-extremity functions will be rising continuously. Many healthcare exercises have been developed to enhance upper-extremity functions. For people who have low upper-extremity functions, moderate exercises will promote the enhancement of motor functions. Thus, effective healthcare exercising systems that can assist people to enhance the upper-extremity motor skills are desirable. Conventional exercises often require ample facilities and supervision by healthcare workers to reduce the risk of overuse injuries [3]. Due to a worldwide shortage of qualified personnel, there is a constant requirement for low-cost and home-based technologies that can assist users to enhance their motor functions.

Recovery of upper-extremity motor functions, e.g., stroke patients, can be divided into several stages, including the flaccid stage and the synergy stage. The first stage is the flaccid stage. The symptoms in this stage involve reduced muscle flexibility, low muscle strength, and limited Range of Motion (ROM) in their joints. The second stage is the synergy stage, in which the major symptoms are voluntary motion with synergistic and increased muscle strength. The focus of the treatment in the flaccid stage includes appropriate positioning of the weak limb, maintaining the range of movement for joints and enhancing muscle flexibility. In the synergy stage, the goal of recovery training is inhibiting abnormal muscle tone and facilitating functional motion. The improvement of motor functions is an ordered and stepwise process, as described using several commonly administered assessment scales [4]. Cheng et al. [1] and Naghdi et al. [5] summarized the enhancement of limb movement into three stages. According to the description of Twitchell [6], the improvement of motor skills in the arm begins with the initial stage of flaccid paralysis. Next, synergy of voluntary movements develops until normal voluntary movement occurs with normal speed. Activities of Daily Living (ADL) training can begin when the muscle strength has increased. It has been documented that challenging tasks that are neither too difficult nor too easy are likely to enhance motivation and elicit motor learning [7]. Hence, the difficulty of training exercises should correspond with the users' motor skills [8]. For healthcare systems that are developed for users with numerous movement abilities, multiple levels of exercises should be provided to fulfill individual requirements, and these exercises should be customized according to the motor recovery process.

Augmented Reality (AR) evolves from Virtual Reality (VR) technology but is closer to the real environment. It is a technology that enables human-machine interaction via superimposing computer-generated information upon real-world environment [9]. AR can provide a realistic feeling and intuitive interaction for the users, and support a controllable environment with real-time feedback at the same time [9]. Selection interfaces that use hands motion and gestures have been reported which allow users to interact with virtual interfaces to select options [8]. AR and VR games have been developed where users interact with these games using dynamic hands/feet gestures in front of a camera, and these gestures trigger the interaction events to interact with the virtual objects in the games [8, 10, 11].

AR-assisted healthcare systems that integrate AR with sensors have been reported to achieve multi-modal interaction space in a smart home and a clinic. AR-assisted healthcare systems are capable of providing cognitive and psychological stimuli [12]. The majority of AR-assisted upper-extremity exercises enhance motor skills using advanced technologies or devices. They aim for minor impairment or a single stage of the improvement of motor functions [13], which means that the target users is narrow. As mentioned above, healthcare systems should be developed following the process of the improvement of motor skills. Although some systems contain exercises that are appropriate for people who are in different motor improvement stages or different motor skills [13], they do not arrange these exercises according to motor progress. In addition, none of them discussed the significance of providing healthcare exercises in stages.

This chapter presents a review of reported healthcare exercising systems that aim to enhance and recover upper-extremity functions. The systems surveyed include conventional interventions, VR-based systems, haptic-based systems, and AR-based systems. An AR-assisted Three-stage Healthcare Exercising (ARTHE) system, which has been developed by the authors, that provides exercises to enhance upper-extremity motor functions is presented to demonstrate the use of AR in stage-based healthcare and rehabilitation systems.

30.2 Related Works

This section presents a survey of the literature pertinent to the studies on upper-extremity exercising systems, algorithms of arm-movement tracking, and methods and devices for performance evaluation. It begins with a review of systems that aim to enhance upper-extremity functions, including conventional methods, VR-based systems, haptic-based systems, and AR-based systems. Next, as an important issue to be considered in AR-based healthcare applications, sensor-based tracking and bare-hand tracking methods are introduced. Section 30.2.3 presents methods and devices for the assessment of upper-extremity movements recovery. Finally, a brief discussion of the reviews in this section is presented.

30.2.1 Upper-Extremity Healthcare Exercising Systems

Numerous approaches on the improvement of upper-extremity functions have been reported. These approaches are designed for one or multiple specific motor functions.

Some commonly used upper-extremity exercising applications are introduced in this section.

Conventional Methods

Range of Motion (ROM) exercises are classical exercises designed to preserve flexibility and mobility of joints [14]. ROM refers to the amount of movements of each joint. ROM exercises for the upper-extremity involve elbow flexion and extension (bend and straighten the arm), shoulder flexion and extension (raise the arm over the head), as well as finger and wrist flexion and extension. If a user cannot perform these exercises with the more-affected arm, he could use the healthier limb for assistance. These exercises are suitable for the flaccid stage.

The more-affected arm exhibits increased muscle strength and voluntary motion in the synergy stage. Hence, exercises designed for this stage should concentrate on the facilitation of movement. Repetitive task training is one of the interventions developed for the synergy stage. It increases the training intensity of the target muscle or joint by repetitive motor practice. This intervention has shown improvement in the lower extremity [15]. For the upper extremity, repetitive training on simple movement is beneficial for maintaining arm functions. However, repetitive training of complex movements is not more beneficial than classical interventions [16]. One possible cause of this situation may be that for people with enhanced muscle strength, rapid movement is more beneficial than precise movement [17].

Another commonly practiced method is the motor relearning technique [18], which emphasizes on functional motor tasks for regaining motor functions for the enhancement of activities of daily living [19]. This intervention technique consists of four steps, namely, find the user's missing function, train with remedial exercises, train with task-oriented exercises, and transfer skills between the remedial and functional tasks. Functional tasks involved in the motor relearning method include buttoning, grooming, dressing upper garment, etc.

The Proprioceptive Neuromuscular Facilitation (PNF) is the most effective stretching technique for optimizing motor performance [20]. This exercise increases the ROM by stretching the target muscle. The ROM of the shoulder, elbow, wrist, and fingers will be increased after the PNF intervention. The upper extremity patterns include diagonal, spiral, and total patterns. Each diagonal pattern consists of flexion and extension.

Exercises applied in these interventions can be classified into the reach-to-grasp movement group, strength training group, and task-oriented group. Some exercises belong to more than one group. The advantages and disadvantages of each group should be taken into consideration when designing healthcare exercises.

The reach-to-grasp movement is the most common arm movement in daily life. This action consists of several components of forelimb movements. The "reach" movement is composed of the rotation of the shoulder joint and the extension of the elbow, and the "grasp" involves flexion of fingers and wrist [21]. In this case, doing reach-to-grasp exercises could enhance the strength of different muscles and the ROM of multiple joints.

Strength training means doing exercises with resistance. According to a meta-analysis of randomized controlled trials [22], strength training is able to enhance upper extremity strength without increasing tone or pain.

Functional training represents a type of exercises that involves tasks that are meaningful to the users, such as exercises that mimic daily living activities [23]. Real-life tasks, e.g., opening drawers, writing, dressing, etc., are commonly used functional training exercises. However, this type of exercises is only suitable for users with sufficient muscle strength [24].

Typically, formal interventions are designed for users who are suffering from stroke, hemiplegia, or other diseases which lead to arm weakness. These applications are administered in hospitals by occupational therapists and physical therapists. It is expensive to set up the equipment and train the therapists. Moreover, users can become bored during exercises.

VR-Based Systems

Besides conventional interventions, VR-based, haptic-based, and AR-based training systems have been reported and used to translate conventional boring interventions into entertaining ones. This section presents a brief overview of VR-based upper-extremity exercising systems.

It has been documented that three vital elements for motor functions recovery are early intervention, task-oriented training, and repetition [25]. In order to make it less boring to perform the same exercise repeatedly, users should be motivated in the process of training. VR enables the training process to be task-oriented and entertaining via visual and auditory motivational components.

The application of VR for motor functions recovery has grown dramatically in the recent years. Inman et al. [26] utilized a virtual environment (VE) to train children with disability to control motorized wheelchairs. VR has been used to render appropriately spaced virtual cues for helping patients who have been inflicted by Parkinson's disease to regain walking ability [27]. Several approaches on improving the upper extremity functions using VR have been explored. A low-cost VR rehabilitation system was designed by Crosbie et al. [28] to help stroke patients recover upper limb strength. This system provides several functional tasks that guide users to reach, grasp, release, and manipulate components. The representation of the user's upper-extremity and the VE are rendered on a Head-Mounted Display (HMD), which is cumbersome and not easily accepted socially. Ustinova et al. [29] developed a "reaching flower" game using VR to study the relationship between viewing angles and the reach distance. Images with viewing angles ranging from 0° (behind) to 90° (overhead) created using VR are displayed on the screen. This research is useful for setting the environment for motor recovery.

Most of VR-assisted upper-extremity exercises are designed for users who suffer from cardiovascular diseases. Only a few applications focus on elderly users. Steffen et al. [3] proposed a VR-assisted personalized exercise trainer for elderly users to train biceps strength. A virtual avatar is rendered on the television and guides the user to receive strength training. This work focuses on exercise monitoring instead of providing suitable exercises for different motor ability.

However, VR-based systems have some disadvantages. The first disadvantage is that body movements and avatar movements are in independent coordinate references [9]. Ma et al. [30] compared the trunk-arm coordination between VR and physical reality during a "catch object" task. Test results showed that the movement patterns induced by VR are different from physical reality because of insufficient depth perception. In addition, VR can only provide a limited feeling of "presence" without providing the feeling of touch.

Haptic-Based Systems

The term "haptic" means "adding the sense of force in human-computer interaction" [31]. Some researchers proposed that incorporating haptic technology into VR could ameliorate the effect caused by the limited feeling of "presence." Boian et al. [32] and Jack et al. [33] created a VR system, which is one of the earliest attempts, using one CyberGlove and one Rutgers Master II-ND (RMII) force feedback glove as input devices. RMII glove applies force to the fingertips to stimulate nerve and practice muscle strength. Subsequently, numerous haptic-based healthcare systems have been reported.

Several approaches have been studied on enhancing motor functions of upper extremities, including hand [34–38], arm [39, 40], and wrist [37, 41]. Shakra et al. [35] developed a haptic-based hand rehabilitation system which utilizes the CyberGrasp system to provide tactile and kinesthetic stimuli. The CyberGrasp system simulates the feeling of object weight and measures hand movement, respectively. This group built three tasks using VR to train users to grasp objects. This system was improved by Alamri et al. [36] and Kayyali et al. [42]. They implemented the Jebsen Test of Hand Function and the Box and Block Test [31] into the task-oriented exercises. This system aims to enhance the hand functions and finger dexterities through engaging exercises.

Early haptic-incorporated systems applied end-effector robots that combine the movement of the user at one point [43]. Examples of this type of devices include the PHANToM Omni [40] and the HapticMaster [44, 45]. Reference [40] presented an application called "The labyrinth" for motivating stroke patients' arm to move via an immersive workbench and the PHANToM Omni. The PHANToM Omni is a mechanical arm composed of five joints and one handle. By operating this device, the patient could navigate a virtual stylus to go through a virtual maze without touching the wall. PHANToM provides force feedback when the patient hits the wall. HapticMaster [45] is a three degree-of-freedom (DOF) haptic interface applied in Gentle/s [44], a 3-year project to develop a machine-mediated therapy system for neurorehabilitation of stroke patients. A gimbal is equipped on HapticMaster which allows the users to put their hands on it and move it during exercises. According to a pilot study carried out from 2001 to 2002, this system is able to motivate users to participate in recovery sessions and provide positive development.

Generally, end-effector robots are easy to fabricate and suitable for users with different arm lengths. However, the output only reflects a mixture of movements from all the joints. New devices should be created to present isolated movement for each single joint. As a result, an exoskeleton was created to measure a single joint movement. Exoskeletons wrap around the arm, measure, as well as control the joint torque separately [43]. An exoskeleton can target specific muscle for training and provide a larger motion range. Gupta et al. [39] developed a five DOF haptic arm exoskeleton for training in a VE. This device can provide kinetic feedback to the lower limb and wrist. It has been developed based on the Mechatronics and Haptic Interface Lab (MAHI) exoskeleton through reducing the device weight and decreasing the device mobility to make the workspace comparable to the human arm workspace. This device can constrain the forearm rotation without affecting the other joints, which is beneficial for wrist rehabilitation.

According to the literature review, it can be seen that incorporating haptic-based healthcare system with VR technology allows users to manipulate and feel virtual objects during exercises [46]. However, most of the haptic devices used in these systems were bulky, and sometimes painful and difficult to wear even with the help of healthy persons. In addition, some devices require calibration to select parameters, such as time delay, angle, zoom, etc., for tracking [42]. This process often takes time and requires users with weak arm to follow various gestures. Moreover, arm fatigue has been observed in some exercises with heavy equipment. These defects should be overcome before introducing haptic devices into public use.

AR-Based Systems

In an AR environment, both the movements of virtual objects and the user's physical movements are measured in the same coordination reference, which overcomes the limitation of the insufficient depth perception that exists in the VR environment. Table 30.1 shows the comparison between AR-based healthcare systems and other interventions.

Table 30.1 Comparison among different upper-extremity interventions

	Advantage	Disadvantage
Conventional interventions	Administered by experienced therapists or caregivers; proven to be safe and effective	Users feel bored easily; costly
VR-based interventions	Make exercises more entertaining; save cost and space; more controllable	Insufficient depth perception; limited feeling of "presence"
Haptic-based interventions	Add the sense of force in human-computer interaction	Often use bulky or heavy devices that might hurt users
AR-based interventions	Immersing virtual objects into real environment; augment multiple types of senses: sight, touch, and hearing	The training plan cannot be customized and adjusted in accordance with the improvement of motor functions

AR application in healthcare is still at an early stage. Some approaches aim to help users regain lower limb functions, such as facilitating normal gait pattern [47–49]. In these approaches, AR is used to present augmented cues on the ground to guide the users. On the other hand, some systems have been developed for upper limb. The research by Luo et al. [50] for hand recovery integrates AR with assistive devices. This system is relatively low cost and small in size, and therefore it is suitable for use at home. AR elements are utilized to present weightless virtual objects on a HMD for reach-to-grasp tasks. Assistive gloves are incorporated into this system to help finger extension. Dynamic feedback of subject performance, including visual, audio, and force feedback, is monitored during training sessions. Lee et al. [51] designed three games for exercising the hand, upper extremity, and the whole body, respectively. A Wii remote device is integrated here for measuring three-axis acceleration. Users are asked to catch, move, rotate, and lay down virtual objects rendered on AR markers.

As discussed earlier, touching and feeling real objects are essential for regaining motor functions. Combining AR with haptic devices will increase the realism of the exercises. In addition, physiological data extracted from haptic devices provides a deeper understanding of the user's performance. Some approaches have been proposed to provide real feeling by adding entity or equipping force feedback sensors. Wang et al. [52] utilized an air pressure detecting device in an AR-based gaming system. In the training session, a user plays a fish tank game by squeezing a rubber ball which generates force feedback; the system measures the force exerted by the user through an embedded pressure sensing device. Some fishes would be displayed on the screen for the user to "catch" by squeezing the ball at the right time. This system enhances the immersive feelings during the training session.

The improvement of motor functions is an orderly process, and healthcare exercises should be developed following the motor recovery process. However, most AR-assisted upper-extremity healthcare applications only provide exercises for a specific motor recovery stage. Some applications focus on individual joint movement. Wang et al. [52] and Shen et al. [12] proposed healthcare games for finger dexterity. Aung et al. [13] provided a series of exercises, and most of them were designed to enhance the ROM of the shoulder. On the other hand, some systems enhance upper-extremity functions using ADL exercises. Ghostman [53] trained a user to use chopsticks by mimicking a therapist's gesture which is rendered on the user's HMD. Among all the AR-assisted upper-extremity exercises, the reach-to-grasp exercise is used most commonly, e.g., web-based serious games [54] and exercises designed in the TheraMem system [55]. Few systems propose upper-extremity exercises that belong to two or more recovery stages. The Augmented Reality-based REHABilitation (AR-REHAB) system [56, 57] provides task-oriented exercises and exercises derived from daily activities to increase a patient's involvement in the training process. However, these exercises are not arranged to follow the motor recovery progress.

From these reported studies, it can be seen that new technologies are crucial for the development of upper-extremity exercising systems. However, most systems focus on enhancing the training effect by introducing innovative devices or technologies without realizing the importance of training the users in stages. At present, none of the existing VR/haptic/AR-based healthcare systems provide exercises in stages.

30.2.2 Upper-Extremity Movement Tracking Methodologies

This section presents some technical issues that are vital in the implementation of an AR-based healthcare exercising system. Tracking is one of these issues and is essential in developing a successful AR-based application for training and recovery of the upper-extremity motor functions. This section discusses some common tracking methods for upper-extremity exercises.

Sensor-Based Tracking

Hand tracking is a crucial requirement in upper-extremity exercises. An earlier strategy is to measure hand movements through hand-held devices, including robot end-effectors and glove-based devices [43]. Some common robot end-effectors, such as PHANToM Omni [40] and HapticMaster [45], have been introduced. These devices combine the whole arm movement into a single point movement, which eases the tracking process but decreases the tracking accuracy. Data gloves which involve wearable sensors have also been studied extensively. The CyberGlove, 5DT DataGlove, and Humanglove™ are commercially available gloves. CyberGlove is able to read a user's spatial characteristics through flex sensors. Hand motion data will be transmitted to a nearby computer wirelessly using WiFi communication. This device has been used in VR-based [33, 32] and AR-based [56, 57] healthcare applications. However, most users cannot afford these gloves due to their high price. In addition, wearing gloves without assistance from caregivers is an arduous task for users with weak limbs.

Bare-Hand Tracking
Computer vision-based bare-hand tracking methods can overcome these shortcomings. Several bare-hand tracking methods have been proposed. Depending on the devices used, bare-hand tracking methods can be classified as depth sensor-based and RGB sensor-based methods. The depth sensor-based tracking methods often utilize Kinect to obtain depth information, such as the application developed by Lange et al. [58] and the Virtual Exertion [59]. Kinect is a simple device which generates real depth maps for pose estimation. In contrast with Kinect, a web camera, which uses RGB sensors, is cheaper and more appealing to everyday users. The NeuroR system [60] uses only one web camera to track the arm movement. This marker-less tracking algorithm begins with face detection and searches for the contour of the shoulder and the upper part of the arm. To achieve an accurate segmentation result, this algorithm needs adequate lighting and a white background.

Most bare-hand tracking methods detect the movement of larger organs, such as hands and arms, instead of fingers. Only a few approaches detect finger bending angles. Watanabe et al. [61] tracked the fingers and estimated bending angles using a single camera. This algorithm assumes no occlusion of fingers occurs during tracking, such that it is only suitable for a certain range of degrees. Metcalf et al. [62] estimated bending degrees through determining the positions of fingertips and convexity defect points measured using Kinect. This is a novel method that goes beyond gesture recognition and lays the foundation for home-based rehabilitation. For home-used finger training exercises, they should be built using cheap devices and able to detect a whole range of angles. Moreover, these exercises ought to be suitable for users with various hand dexterities, including individuals with hand spasticity and contractures. This work discusses the possibility of using low-cost devices, i.e., web cameras, to fulfill these requirements.

30.2.3 Assessment of User's Upper-Extremity Functions Recover

Healthcare systems often contain an assessment module to evaluate the user's performance in response to his/her physical conditions and recovering progress, in order to keep track of the user's progress so as to provide more effective training. This section starts with a brief introduction of the popular outcome measurements for upper-extremity assessment. Next, glove-based devices for motor functions assessment are presented.

Outcome Measurements
The Fugl-Meyer Assessment (FMA) [63] is a reliable and valid assessment tool for describing motor recovery and helping therapists plan treatment. This assessment contains 33 upper extremity items for evaluating the movement of the shoulder, elbow, and forearm. Generally, the FMA evaluates a user's ability to fully flex or extend joints and measures the ROM of joints and movement velocity.

The Wolf Motor Function Test (WMFT) is a widely used outcome measure for evaluating upper-extremity motor abilities via single-joint or multiple-joint movements [64, 65]. The WMFT has many versions, and the most widely used version contains 17 tasks which involve timed functional tasks and strength-based tasks. This outcome measure sets a time limit, i.e., 120 s, for completing each task.

The Action Research Arm Test (ARAT) is a method to assess the functional changes of the upper extremity via observational tests [66]. This method has been developed based on the Upper Extremity Function Test (UEFT) and consists of 19 items which have been categorized into four subgroups, namely, grasp, grip, pinch, and gross movement [67]. The ARAT also incorporates time limits into the scoring definition, e.g., the user will receive the full mark if he performs the test in 5 s [68].

Glove-Based Devices for Motor Functions Assessment
According to the literature review presented in the previous section, the upper-extremity functions can be evaluated using various indicators, including ROM of joints, movement velocity, and grip force. As wearable sensors are capable of providing quantitative motion sensing, they have opened the possibility of evaluating treatment efficacy and quality of life [69]. In comparison with traditional observational assessments in which accuracy and consistency vary among clinicians, physiological sensors could provide more stable measures. In order to measure hand movement for assessment, data gloves which involve wearable sensors have been widely used.

There are some commercially available gloves that have been designed to track finger movements, e.g., finger's bending degrees. The bending degrees are an essential indicator to describe the user's motor functions. Some research-based devices have also been developed. Williams et al. [70] designed a goniometric glove which enables finger joint assessment in a clinical environment. This glove is able to measure angles of multiple fingers simultaneously. Carbon ink bend sensor, one of the resistance tracking sensors, is used here because of its ease in adaptation in length. Apart from resistance tracking sensors, optical tracking sensors and magnetic tracking sensors are also commonly used sensors. A magnetic sensor reports its location and orientation with respect to a specific magnetic field. Fahn et al. [71] presented a data glove which tracks the fingertip position and orientation using attached magnetic sensors. The advantages of magnetic sensor are high accuracy and long life span. The shortcoming of magnetic sensor is that it may be affected by metallic objects appearing in the magnetic field. Optical tracking sensors were embedded in a silicone prosthetic glove [72] to detect object slip. LED is used as a light source and sends light to a surface which will reflect light to a CMOS sensor. Patterns in the surface images contain information about the amount of object slip. Compared with other sensors, optical sensors are typically larger and depend on lighting condition.

Grip force is an important indicator for describing upper-extremity functions. In comparison with normal people, users with weak upper-limb show delayed force formulation and variable force application. Force sensors are widely used in healthcare devices to measure grip force. Numerous applications equip pressure sensors or force transducers on the object that will be grasped by users [73–75]. A large number of pressure sensors will be required to be used to assess the user's upper-extremity functions when he performs daily living activities. Therefore, embedding force sensors in data glove is a relatively feasible and low-cost way to measure grip force.

Accelerometers have been utilized in several data gloves to detect muscle activity. Accelerometers can provide information about rotation and ambulatory movements [76] that are consequential in describing the mechanical properties of the muscle. Accelerometers have been placed on the affected arm to detect the severity of dyskinesia [77], Parkinson's disease [78], stroke [79], etc. Uswatte et al. [80] showed that accelerometer data could provide more objective information about the motor function of the arm than other sensors.

Recently some researchers turn their attention to gloves with multiple sensors. The eGlove which is employed in the Ubi-REHAB [81] measures hand motion using bending sensors and three-axis accelerometers. Combining multiple sensors in one device allows the capture of the various physiological factors simultaneously, which will produce a more reliable assessment outcome.

30.2.4 Discussion

This section provides a review of conventional and new interventions for enhancing upper-extremity functions, upper-extremity movement tracking algorithms, as well as methods and devices for assessment. It has been documented that to build an effective healthcare system, exercises provided by this healthcare system should be developed following the motor recovery process [1, 82]. Although AR-based upper-extremity healthcare applications gather all the advantages of former approaches, including VR and haptic-based systems, no system provides exercises following the motor progress. In addition, to realize the objective of introducing a healthcare exercising system for people in different motor recovery stages, the tracking methods applied in this system should be suitable for individuals with various upper-extremity statuses. Therefore, for individuals with hand spasticity and contractures, a low-cost and accurate bare-hand tracking method is more appropriate. Moreover, evaluating the user's performance is crucial in exercising systems. Thus, a device that detects multiple types of physiological data for assessment is required.

30.3 System Overview

The ARTHE system presented in this chapter is designed and developed as a home-based three-stage healthcare upper-extremity exercising system for people with low upper-extremity functions. The framework of this system is shown in Fig. 30.1.

The system is composed of three modules, i.e., a monitoring module, an exercise module, and an assessment module. The monitoring module detects a user's movement and physiological data through a web camera, the OptiTrack system, and an AR-based data glove. Data collected from the monitoring module are used in the exercise module. The exercise module involves seven AR-assisted exercises that have been categorized into three stages. The assessment module consists of a scoring module which is designed to evaluate the user's performance. User information, exercise results, training plans, and physiological data are stored in the database that can be further accessed by the users, caregivers, etc.

For the exercises module, both virtual and real objects are used to train users. Virtual objects offer an alternative for users who do not have sufficient strength to perform daily living activities. Augmented visual feedback is useful to enhance brain excitability and modulate specific regions of the brain [83]. Tangible objects are adopted to add resistance so as to improve the muscle strength. Allowing users to interact with tangible objects would allow the successful transfer of skills learned from the virtual environment to reality.

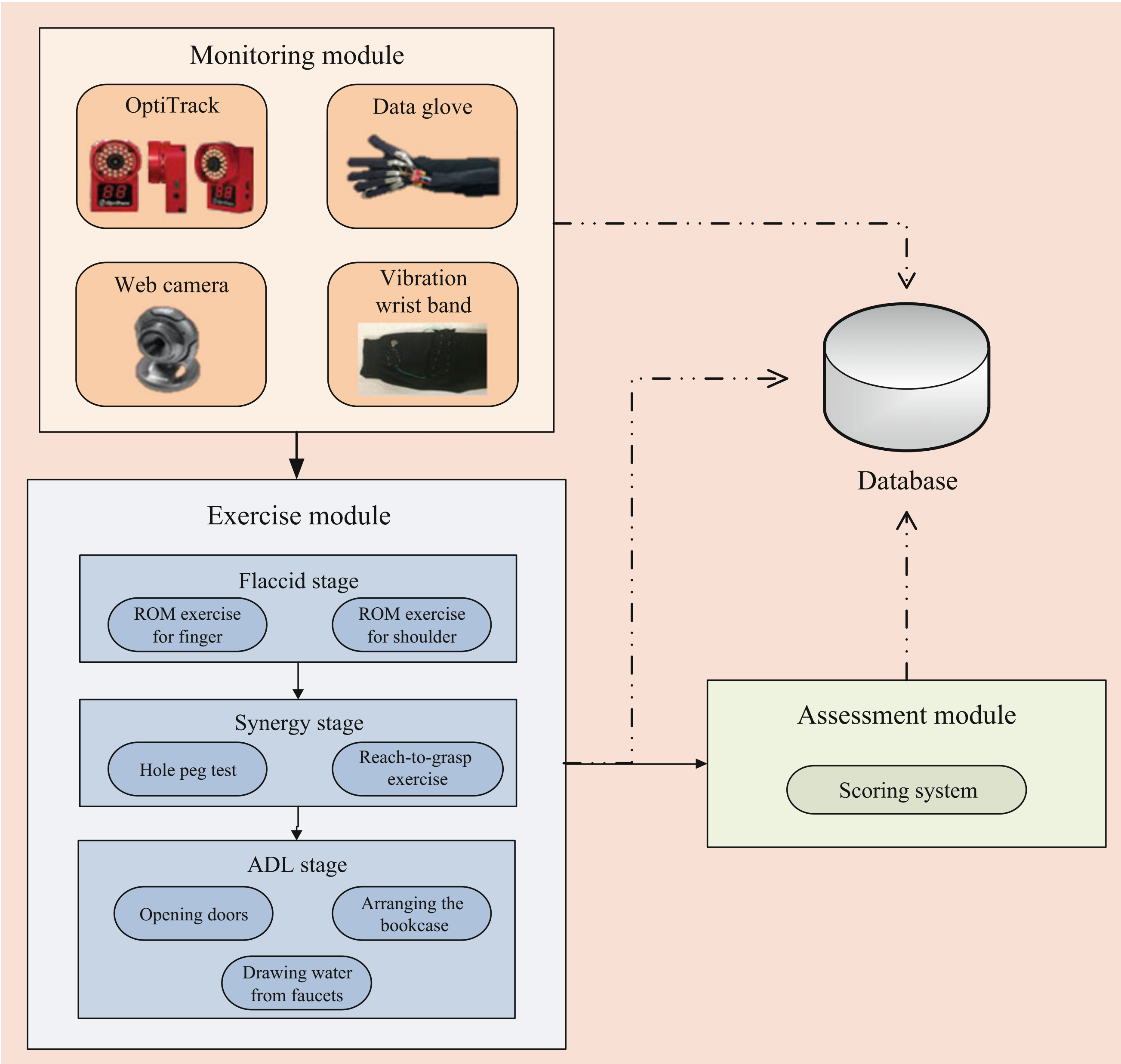


Fig. 30.1 Framework of ARTHE

The system is developed using C++, and the User Interface (UI) is built using MFC. Exercises are built using the ARToolKitPlus. 3D objects created from SolidWorks and video frames are rendered using OpenCV and OpenGL. The open source database software MySQL is chosen to store data in the ARTHE system.

Figure 30.2 shows the hardware setup of the ARTHE system. Three OptiTrack cameras, V100:R2, are positioned in front of the user to define the tracking volume. Visual information is displayed on the computer screen which is located in front of the user. One web camera is placed on the screen, facing the user's upper extremity. The user wears the data glove on the more-affected arm and put this arm on the table top during exercises. Few reflective balls are attached on the user's upper extremity for positioning.

30.4 Monitoring Module

This section introduces the motion tracking methods and the AR-based data glove used for monitoring the users.

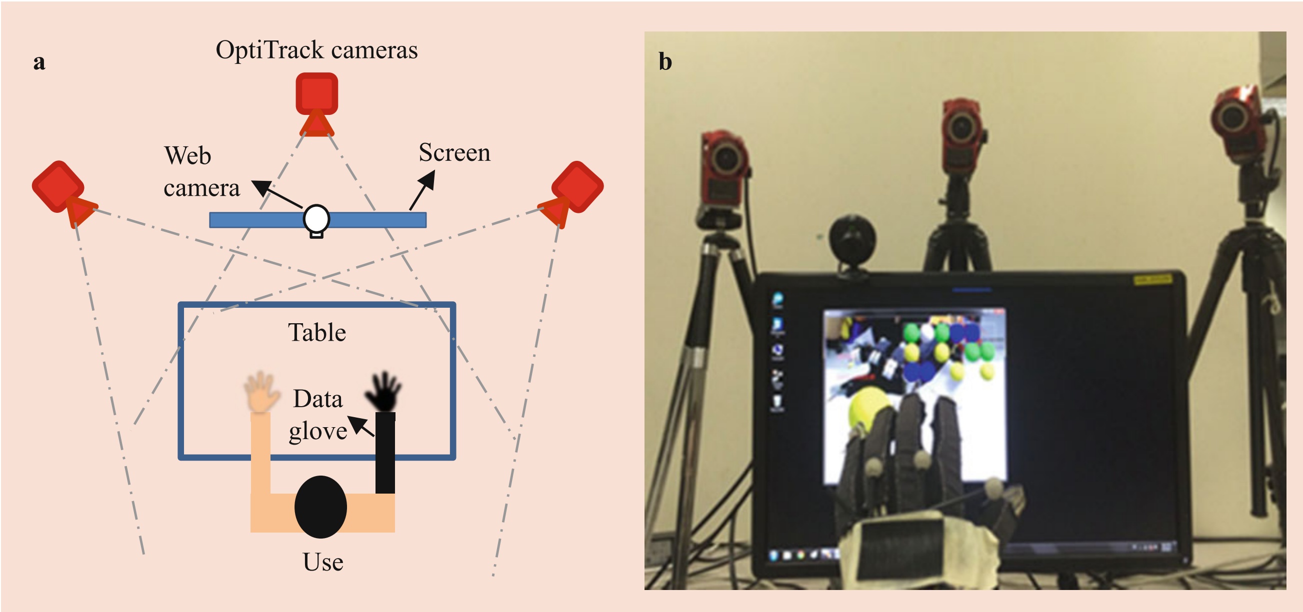


Fig. 30.2 Setup of ARTHE: (**a**) setup diagram; (**b**) snapshot of the actual setup

30.4.1 Motion Tracking

The OptiTrack motion system is used for motion tracking in this system. The coordinate system of the tracking volume is defined by the OptiTrack system after camera calibration. As this coordinate system is different from the one defined by the web camera, the extrinsic parameters, which are represented by the transformation matrix, need to be determined. The positions of the OptiTrack cameras and the web camera are fixed. Hence, this transformation matrix is constant and stored in the system configuration file that allows information about this system to be saved.

A rigid body which is composed of at least four reflective markers is used to obtain the transformation matrix. Equation (30.1) represents the relationship among the tracking volume coordinate system, the camera coordinate system, and the camera screen coordinate system. The coordinates of one marker in the tracking volume coordinate system is (X, Y, Z), which can be obtained using the OptiTrack system automatically. (X_C, Y_C, Z_C) is the location of the same point in the camera coordinate system, and (u, v) is the position of this point in the camera screen coordinate system. (u, v) is obtained through image processing. The intrinsic parameters are represented by a 3×3 matrix, $\mathbf{M}$. $\mathbf{P}$ is the transformation matrix.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{M} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \mathbf{M} \cdot \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{30.1}$$

30.4.2 AR-Based Data Glove

It is costly to have a professional expert to monitor the training process and evaluate a user's performance. Commercially available data gloves are expensive for home-used rehabilitation and assessment. In order to provide inexpensive rehabilitation and obtain a reliable assessment outcome economically, an AR-based data glove is designed to measure the bending degrees of fingers, grip forces, and accelerations of the arm simultaneously. Five flex sensors, five Flexiforce pressure sensors, and two 3-axis accelerometers MMA7361 are positioned on an elastic long-sleeve glove. The Arduino Pro Mini 328 is utilized to receive data from multiple sensors, translate them into digital format, and transmit them to the computer. In order to transmit data over a short distance wirelessly, the BlueSMiRF Gold is chosen to work with the specific Arduino. The Mini Bluetooth Adapter is utilized to pair the Bluetooth modem to computer. Figure 30.3 shows the structure of the AR-based data glove. Both the force sensors and flex sensors are positioned on the dorsal surfaces of fingers. Unit#1 in Fig. 30.3 is the processing unit that handles the data from the accelerometers. Unit#2 and unit#3 in Fig. 30.3 handle the data transmitted from the flex sensors and force sensors, respectively. The accelerometers are installed at the ends of the upper and lower arm.

30.4.3 Vibration Wrist Band

It has been documented that practices that are rich in salient feedbacks show more positive outcomes for the enhancement

Fig. 30.3 Structure of the AR-based data glove

of motor functions. A vibration wrist band was created to generate tactile feelings according to the user's performance. Figure 30.4 shows the vibration wrist band. One vibration motor was embedded in a wrist band. The Arduino Pro Mini 328 is utilized to receive data from the computer and control the vibration motor. In order to transmit data over a short distance wirelessly, the BlueSMiRF Gold is chosen to work with the specific Arduino. The Mini Bluetooth Adapter is used to pair the Bluetooth modem to the computer. The vibration motor was placed on the dorsal side of the forearm, and the other sensors were located on the palmar side.

The vibration wrist band provides three types of vibration modes. The first mode is vibration for 2s with a 1s pause in between the vibrations. When finger bending degrees are not large enough to grasp the virtual object and the user stops bending for more than 1 s, the first vibration mode will be activated repeatedly until the user continues to bend his fingers. For the second mode, the vibration motor vibrates for 0.2s and stops for 0.2s. This process will be repeated four

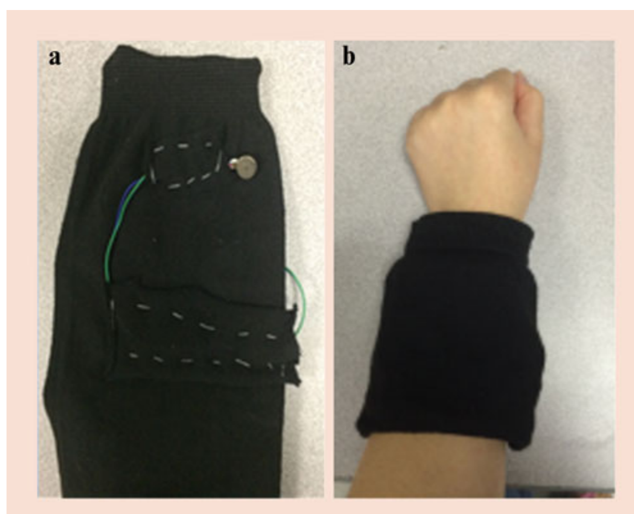


Fig. 30.4 Vibration wrist band. (**a**) The vibration motor that is installed inside the vibration wrist band; (**b**) appearance of the vibration wrist band

times. The second mode will be utilized when the user bends his fingers to specific angles successfully. The third mode will be activated when the user moves his hand on a virtual object. In this mode, the motor vibrates for 1s once.

30.5 Exercise Module and Scoring Module

The exercise module consists of seven exercises that have been categorized into three stages. The scoring module involves seven subsystems.

30.5.1 Flaccid Stage

In the flaccid stage, the focus of the training is regaining the ROM of the upper-extremity joints, e.g., fingers and shoulder. In this training stage, two tasks are provided for increasing the ROM in the shoulder and finger.

ROM Exercise for Finger

Inspired by the famous game Puzzle Bobble, a ROM game is designed to enhance the mobility of users' fingers. This exercise allows the user to interact with 3D objects of different shapes that are rendered on a computer screen. The right side of the screening is for the rendering of 3D objects, spaces that are waiting to be filled by objects and a blinking white ball that indicates the location where the next object should be located (Fig. 30.5a). The left side of the screening displays a 3D object which changes its color and shape at regular intervals.

When the user bends his fingers to a specific degree, the blinking white ball will be replaced by an object displayed on the left side. Three adjacent objects will disappear if they have the same shape and color. Table 30.2 presents the differences among three tasks. The thresholds of bending degrees have been selected based on the "boundary" of low and high function [84] and data collected from healthy participants.

As the objective of this exercise is to eliminate 3D objects that are rendered on the right side of the screen, the user's performance is quantified based on the number of virtual objects that remain on the screen, as shown in Eq. (30.2). N represents the total number of objects, i.e., the number of spaces waiting to be filled plus the number of objects that are shown on the screen at the start.

$$\text{Score}=\left(\frac{1-\text{Number of objects stay on the screen}}{N}\right)\times 100 \tag{30.2}$$

ROM Exercise for Shoulder

The ROM exercise for the shoulder is designed in accordance with the ROM exercise [14], as shown in Fig. 30.5b. In this exercise, a user sits in front of the camera, moves the affected

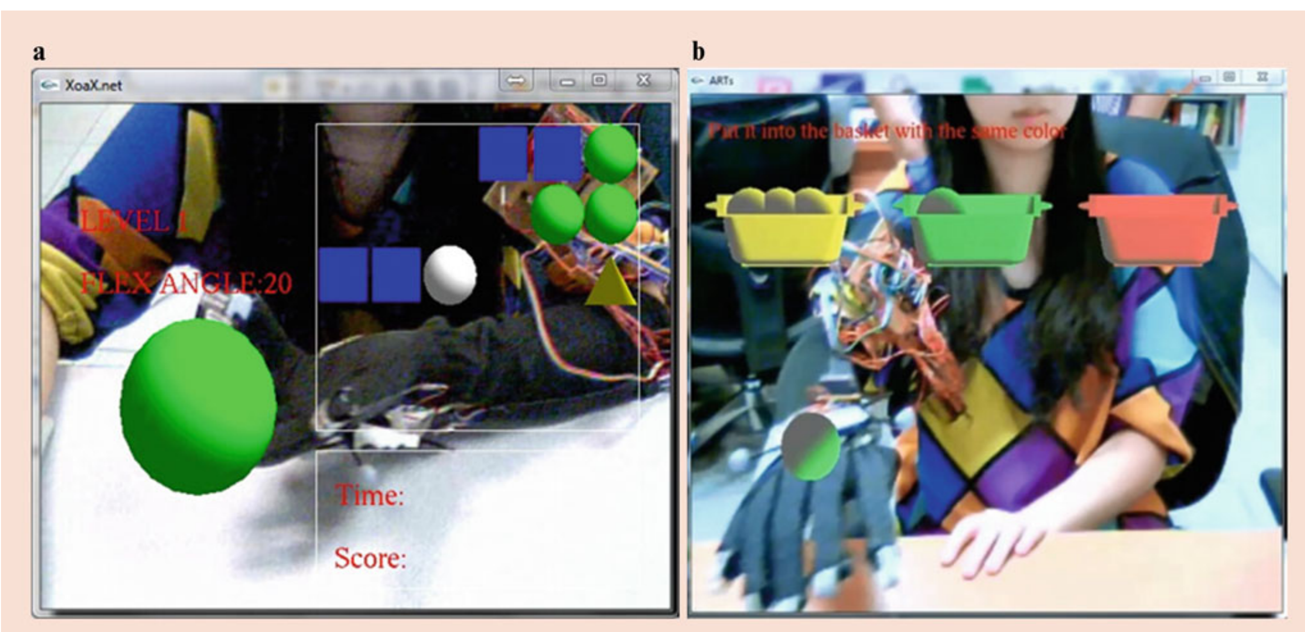


Fig. 30.5 (**a**) ROM exercise for fingers; (**b**) ROM exercise for shoulder

Table 30.2 Differences among the three levels

	Task 1	Task 2	Task 3
Number of spaces that are waiting to be filled by new objects	6	13	19
Number of object's type	3	4	5
Duration of the object that has been displayed on the left side of the screen	10s	7s	5s
The bending degree that has been selected as a threshold of drawing the object on the main interface	20°	40°	60°

Table 30.3 Rating scale for the ROM exercise for shoulder

TCT (s)	<5	5–60	60–120	>120
Score	100	70	40	0

hand behind the ball which is displayed on the left bottom of the window, and places it into the corresponding basket that has identical color with the ball. The target basket blinks to attract the user's attention. This exercise contains three tasks. The height of the baskets and the number of baskets enhanced with the progress.

The scoring module for this exercise is designed on the basis of the Action Research Arm Test (ARAT) [66], where only the Task-Completion-Time (TCT) is taken into consideration. The user is scored according to the rating scale presented in Table 30.3.

30.5.2 Synergy Stage

The most essential issues in the synergy stage are the facilitating movement and inhibiting abnormal muscle strength. Mass movement patterns which are composed of different joint movements should be integrated according to developmental sequences. Two exercises are developed in this stage. In this stage, users are required to interact with virtual objects before moving on to training with real objects that can help users regain muscle strength.

Hole Peg Test

The Hole Peg Test (HPT) is designed based on the classical Nine Hole Peg Test (NHPT) [85] which has been created to measure manual dexterity of people who have mobility impairment. The objective of the HPT is to enhance the upper-extremity functions and manual dexterity.

In the virtual object training phase, users are required to place virtual objects into appropriate holes. A virtual board with seven holes is augmented in the user's view (Fig. 30.6a). When the task begins, the user moves the affected hand to the left bottom to "grasp" an object. Next, the user rotates his wrist to adjust the object's orientation and places this object into the corresponding hole. Once the object has been placed into the hole, a new object would appear in the left bottom of the interface.

In the real object training phase, the user would need to place real objects onto a real board (Fig. 30.6b). The real objects are placed on the left hand side, and the board is placed in front of the user. The time spent completing this task is shown on the left bottom of the screen.

Reach-to-Grasp Exercise

Reach-to-grasp is one of the most common arm movements in daily life. The objective of the reach-to-grasp exercise is to enhance the dexterity and strength of the shoulder, elbow, and fingers.

The virtual object training phase allows the user to interact with a virtual ball and a virtual cup. In this phase, the user's affected hand reaches for the virtual object until a collision is detected. Next, the object will be rendered on the affected hand (Fig. 30.6c). The user should bring the object to the destination. For the virtual cup, it requires the user to maintain affected hand in an upright orientation during the moving process.

The real object training phase uses one tennis ball and one cup (Fig. 30.6d). The objective of this phase is to move the real object to the destination in the shortest time. To reduce the difficulty, the user will start with moving the tennis ball, which is lighter in weight.

Scoring Module for Synergy Stage

The two exercises in the synergy stage share the same scoring module. The user's score was calculated using the TCT, finger's extension ability, and upper-extremity's acceleration and grip stability. The TCT section is built based on the ARAT, as described in Table 30.3.

The finger's extension ability is evaluated based on the flexion-extension range of each finger and the number of fingers which have extension ability. The user can get full mark if at least three fingers can bend to at least 20° [84]. The user will obtain one third of the full mark if no more than two fingers can provide measurable bending degrees.

Hester et al. [86] predicted Fugl-Meyer score using accelerations. The ARTHE system modifies this model slightly, as shown in Eq. (30.3). F_x represents the acceleration of forearm x-axis, and F_y represents the acceleration of forearm y-axis. U_x is the acceleration of the upper arm x-axis, and U_y is the acceleration of the upper arm y-axis.

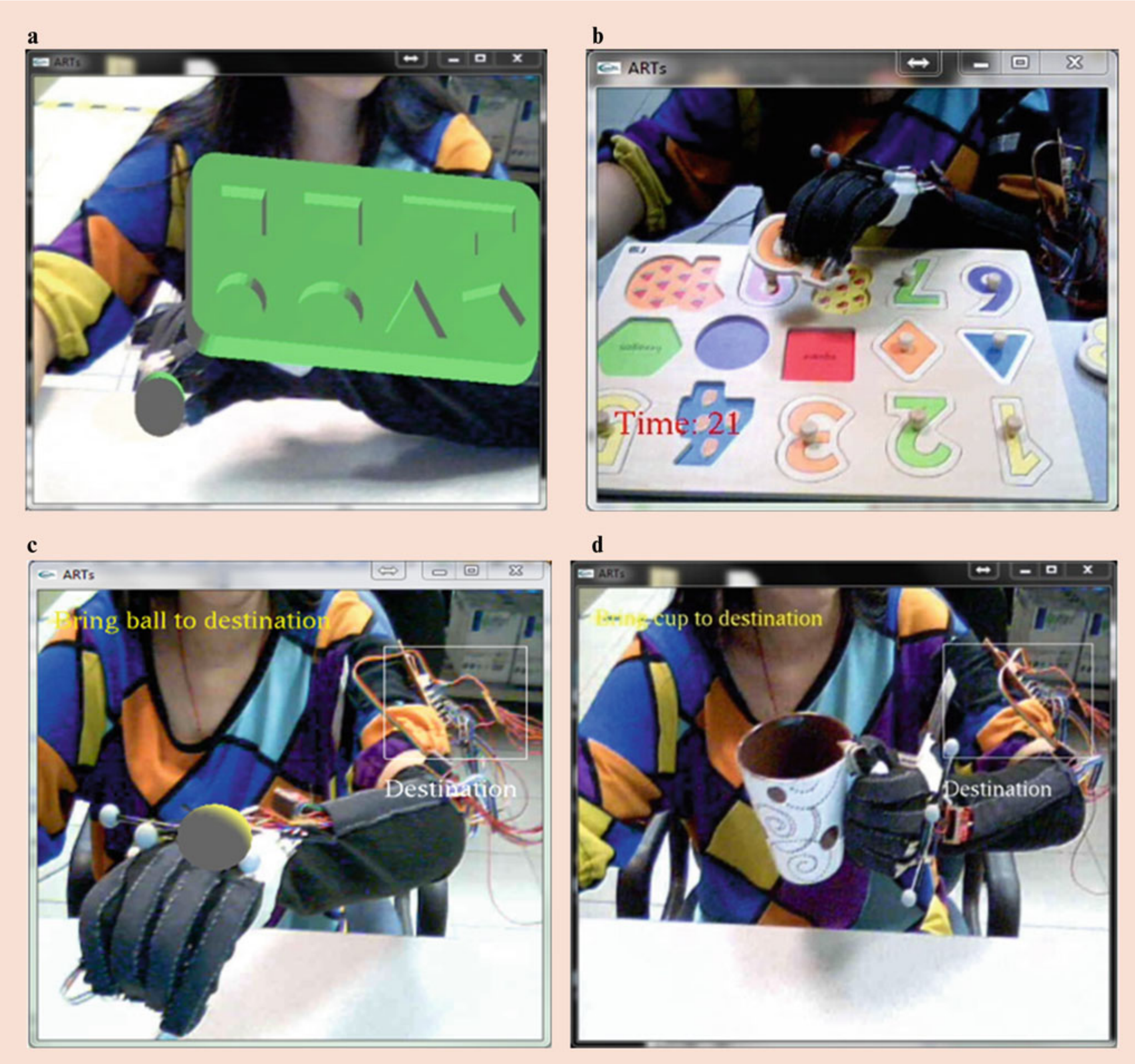


Fig. 30.6 (**a**, **b**) Virtual and real object training in HPT; (**c**, **d**) virtual and real object training in reach-to-grasp exercise

$$\begin{aligned}\text{Score} &= 1.18 \times \mathrm{F}_x + 1.81 \times \mathrm{F}_y + 0.65 \times \mathrm{U}_x + 1.5 \times \mathrm{U}_y \\ &\quad + 0.99 \times \mathrm{RMS}\,(\mathrm{F}_x) + 1.66 \times \mathrm{RMS}\left(\mathrm{F}_y\right) \\ &\quad + 0.31 \times \mathrm{RMS}\,(\mathrm{U}_x) - 1.31 \times \mathrm{RMS}\left(\mathrm{U}_y\right) - 0.86\end{aligned} \tag{30.3}$$

Grip stability is represented by the Standard Deviation (SD), which is expressed as the percentage of the mean. The scales of the marks for grip stability are defined using data measured from healthy participants and the work reported in [74], as shown in Table 30.4.

Table 30.4 Scoring of the grip stability

Grip stability	0–15%	16–30%	31–40%	Above 40%
Sub score	10	6	3	0

30.5.3 Activities of Daily Living Stage

The final stage of the ARTHE system is to train users to perform activities of daily living. Exercises in this stage are composed of the virtual and real object training phases. The

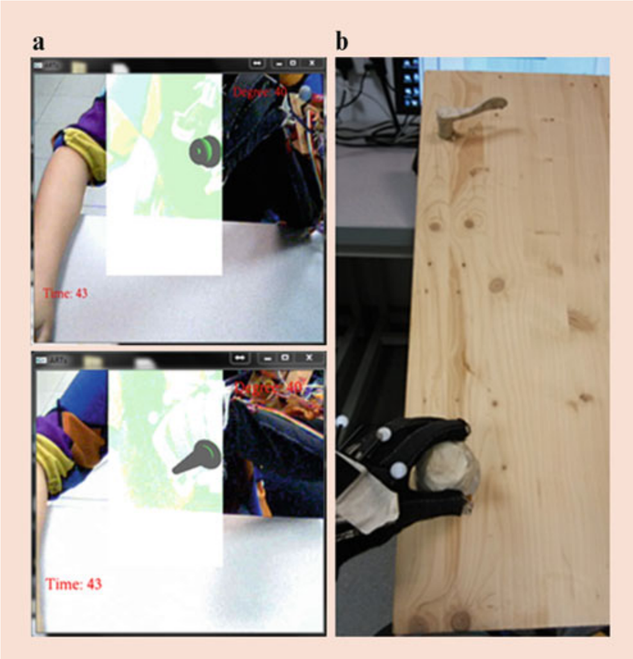


Fig. 30.7 (**a**) Virtual object training phase; (**b**) real object training phase of opening doors

details of arranging the bookcase exercise and the drawing water from faucet exercise have been introduced in a work reported earlier [87].

Opening Doors

This test simulates one of the most common ADLs, namely, opening doors. Two most commonly found doorknobs are used in this exercise.

In the virtual object training phase, a virtual door with a virtual doorknob is displayed on the screen, as shown in the left part of Fig. 30.7. The user has to grasp the doorknob and turn it 60° (measured from real doorknob). After completing this task, a different doorknob will be rendered on the door. Once this task is completed, the score, the total time, and an encouraging message will be displayed on the screen.

In the real object training phase, users will be tasked to turn real doorknobs which provide a spring force against the users' action (Fig. 30.7b). Actual doorknobs that are of the same types as the virtual ones are installed on a wooden plank, simulating a real door. The real object training phase shares the same procedure with the virtual phase.

Arranging the Bookcase

The objective of this exercise is to enhance the users' shoulder, elbow, and finger strength. The virtual object training phase is composed of two tasks. In the first task, the user would need to grasp the inverted book and turn 180° anticlockwise. In the second task, the user "grasps" the book that lies on the table, rotates it 90° to keep the book upright, and places it on the bookcase (Fig. 30.8a).

The equipment used in the real object training phase includes a real bookcase, a speaker, and real books. The user is required to walk around to complete this task, which enhances the lower-extremity strength, back muscle strength, and balance [1]. This section involves three tasks, i.e., put a dictionary on the top shelf (Fig. 30.8b), reposition an inverted folder, and place three books according to their serial numbers. Audio instructions are provided to the user in this training phase.

Drawing Water from Faucets

This exercise is designed to enhance the user's arm strength, wrist function, and finger dexterity. The virtual object training phase involves two tasks as two different faucets are used. Both tasks involve three stages. The first stage is to place a virtual cup under a virtual faucet (Fig. 30.8c). The second stage is to "grasp" the target faucet and rotate it. In the third stage, water "flows" from the water outlet, and the user only needs to wave the affected hand to proceed onto the next task.

In the real object training phase, a speaker and a wooden board installed with two faucets is used (Fig. 30.8d). This exercise begins with placing the cup under the faucet. Next, the system "tells" the user to turn on the faucet, release the faucet, and move the cup under the next faucet. After turning on the second faucet, the system announces the final score.

Scoring Module for the Daily Activities

The scoring module for the third stage is designed through comparison with healthy users. This scoring module involves four indicators, namely, TCT, bending degrees, movement stability, and grip force. The score for each indicator in each training phase is calculated using Eq. (30.4). Thresholds represent the minimum scores indicating healthy movements. A full score means the maximum score for the specific task. The number of tasks in each training phase is represented by n.

$$\text{Score of each indicator} = \sum_{i=1}^{n} \text{Full score} \times \left(\frac{\text{Threshold}}{\text{Data recorded}} \right) \tag{30.4}$$

The TCT refers to the time spent on each task. The bending degree is the flex angle required to grasp virtual or real objects. Movement stability and grip stability are represented by SD. Movement stability is evaluated using data collected from the accelerometers positioned on the lower and upper arms. The grip forces measured from the thumb, index finger, and middle finger are used to calculate the grip stability. Different indicators are used for the exercises in the ADL stage to evaluate the user's performance, as shown in Table 30.5.

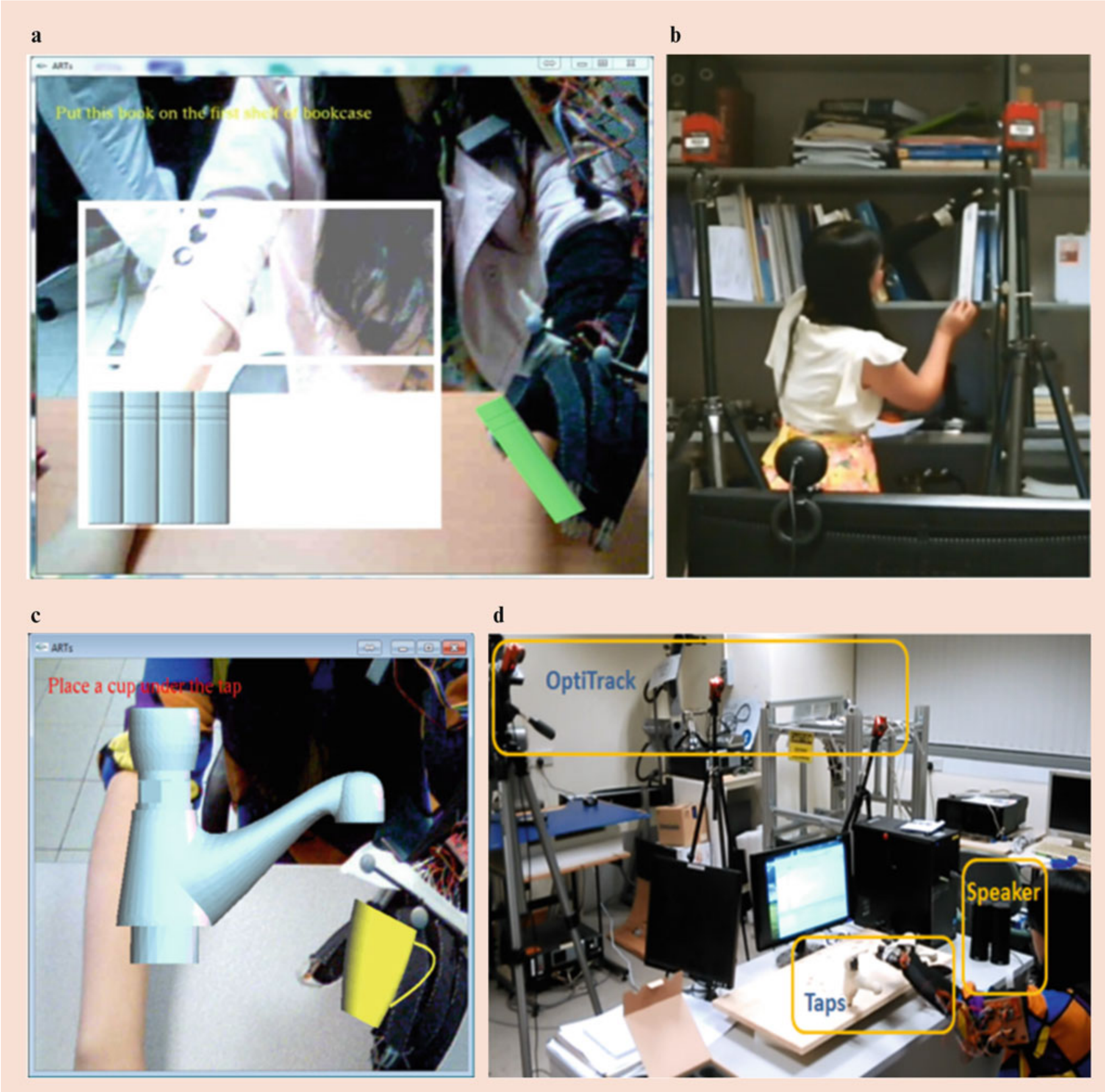


Fig. 30.8 Exercises of (**a**, **b**) arranging the books and (**c**, **d**) drawing water from faucets

Table 30.5 Selection of indicators among exercises of daily activities

	TCT	Bending degree	Movement stability	Grip stability
Opening virtual doors	✓	✓	×	×
Opening real doors	✓	✓	✓	✓
Drawing water from virtual faucets	✓	✓	✓	×
Drawing water from real faucets	✓	✓	×	✓
Arranging a virtual bookcase	✓	✓	✓	×
Arranging a real bookcase	✓	×	✓	✓

30.6 Usability Experiment

To verify that the three-stage training strategy can accelerate motor learning beyond traditional non-stage-based AR-assisted exercise, a user experiment was proposed. This experiment consists of an experimental group and a control group. Participants in the experimental group were trained using ARTHE while participants in the control group were trained using the AR-based virtual piano game [88]. The AR-based virtual piano game allows users to improve hand functions through synergy movements, e.g., move his hand to a specific location and flex fingers. Although this game is composed of multiple difficulty levels, it is not a stage-based approach which has been introduced in this chapter. This is because the difficulty levels are not designed following the process of the improvement of motor skills. The hardware used in the AR-based virtual piano game includes a standard computer, a web camera, and a self-designed data glove which detects bending angles of the fingers. A virtual piano and a virtual hand are located by two ARToolkit markers (Fig. 30.9a). During the training session, the participant wore the data glove on the left hand and sat in front of a computer screen. On the screen, a green arrow is rendered on the target key. When the target key is pressed, the sound of the pressed key is played.

30.6.1 Method

In this experiment, healthy right-handed participants were asked to use their non-dominant hands, i.e., left hands. The reason for using the non-dominant hands is that the non-dominant hands have lower ability in motor skills, strength and coordination, and cognitive planning [89]. As the objective of this experiment is to make preliminary observations on the system's capability of accelerating motor skill learning, perform tests and training on non-dominant hands is acceptable.

30.6.2 Participants

Ten participants were invited for this test and their ages range from 20 to 32. These participants include six males and four females, and all participants exhibited right-handed dominance. The participants were assigned to two groups randomly, and each group is composed of three males and two females. As this experiment is designed as a single-blind experiment, participants did not know which group they belonged to.

30.6.3 Experiment Procedure

This experiment has been developed based on the method proposed by Kaber et al. [90]. It is composed of three sessions, namely, a pre-test session, a training session, and a post-test session. The duration of the experiment is 10 days, and the schedule is presented in Table 30.6. On the first day, all participants were involved in the pre-test. Next, participants in the experimental group were trained using the proposed three-stage healthcare system for 9 days, i.e., 3 days for each stage. For the participants in the control group, they were trained using the AR-based virtual piano game for 9 days. The length of the training for each day is decided after conducting a pilot test. In the pilot test, participants were trained using ARTHE, and signs of fatigue were observed after 30 min of training. Therefore, to mitigate fatigue effects, all ten participants were trained for 20 min every day.

During the training session, all the visual information is displayed on the computer screen. Hence, no see-through

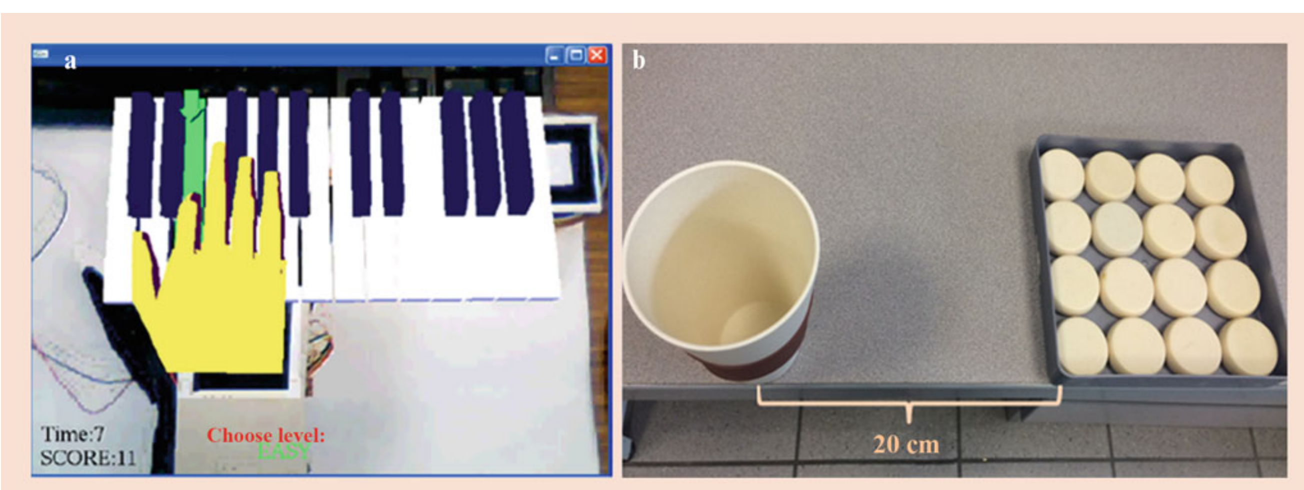


Fig. 30.9 Exercises used in the usability experiment: (**a**) virtual piano game; (**b**) setup of the trial for pre-test and post-test

Table 30.6 Experiment schedule

Day	Duration of training	Task	
		Experimental group	Control group
Day 1	10 min	Pre-test	Pre-test
Day 1–3	20 min	Stage 1: ROM exercise for finger, ROM exercise for shoulder	Virtual piano exercise
Day 4–6	20 min	Stage 2: Reach-to-grasp exercise HPT	Virtual piano exercise
Day 7–9	20 min	Stage 3: Open doors Drawing water from faucets Arrange the bookcase	Virtual piano exercise
Day 10	10 min	Post-test	Post-test

device was used in this experiment. The light intensity ranged from 260 lux to 420 lux. In the control group, participants' movements were tracked using the web camera and the self-designed data glove. For the experimental group, movement data were captured using OptiTrack, the AR-based data glove, and a web camera. Data measured using gloves were transmitted to PC wirelessly, with several Bluetooth modems.

Evaluation results are utilized to validate two hypotheses. The first one is that participants trained using the ARTHE system will spend less time on post-test, which means that their upper-extremity functions will be improved. The second hypothesis is that the ARTHE system can bring better motor enhancement effect than the virtual piano game.

30.6.4 Design of Pre-test and Post-test

As the virtual piano exercise only trains the capabilities of translating the arm and bending fingers, the trial used in the pre- and post-test should focus on these two capabilities. A trial which was also inspired by the NHPT was applied for the pre- and post-test. Figure 30.9b shows the setup of this trial, including 16 round blocks and a cup. The diameter of one block is 4.3 cm and its weight is 31.3 g. All the blocks were tiled on the bottom of a square box, and the cup is placed on the left, about 20 cm from the box. In this trial, a participant was required to bend his left hand to grasp one block, move it above the cup, drop it into the cup, move back, and grasp another block as quickly as they could. These blocks were grasped from top to bottom and left to right. The scores of the pre- and post-test are calculated using the time spent on completing these tests. Before the test, participants were allowed to practice with this device once. When participants were familiar with this device, the participant sits in front of the box to perform the experiment.

30.6.5 Results and Discussion

To evaluate the difference between the experimental group and the control group before the training session, a one-way Analysis of Variance (ANOVA) tests were performed between the pre-test scores of these two groups, and the result is presented as Test 1 in Table 30.7. In statistics, p-value and F-value are used together to assess the validity of a null hypothesis. The p-value tests a null hypothesis that samples from all groups are drawn from the same population. The significant level of these ANOVA tests is set to be 0.05. The F-value is the ratio of the mean square value between groups to the mean square value within groups. If the p-value is substantially larger than 0.05 and the F-value is close to 1.0, it demonstrates that no significant difference exists among the groups, and the observed differences are caused by random sampling. If the p-value is much smaller than 0.05 and the F-value is quite large, it is unlikely that the samples are drawn from the same population [91]. The p-value of Test 1 is 0.67 and the F-value is 0.019 when the two degrees of freedom are 1 and 8, which means that no significant difference exists between the participants in the two groups before training.

The first hypothesis is that ARTHE can improve users' upper-extremity functions. A significant increase can be observed between the pre- and post-test scores for the experimental group, which are shown as Test 2 in Table 30.7. A one-way repeated measures ANOVA (rmANOVA) test was performed to compare the pre- and post-test scores of the experimental group. The reason of choosing rmANOVA is that the data for these two groups in this test were data measured from the same members over different time periods. The result of this ANOVA test is dramatically significant ($F(1,4) = 362.604$, $p < 0.0001$). According to the above results, it can be concluded that a certain degree of motor skill development can be achieved after training using the ARTHE system.

The second hypothesis is that ARTHE can provide better training effect than other non-stage-based AR-assisted healthcare systems. Table 30.7 presents the pre- and post-test scores of the control group (Test 3). The difference between the pre- and post-test scores of the control group can be attributed to the training using the AR-based virtual piano game. The result of Test 3 indicates that a significant difference exists between the pre- and post-test scores of the

Table 30.7 ANOVA tests

Test	Group 1		Group 2		p-value[a]	F-value
	Group name	Mean ± SD	Group name	Mean ± SD		
One-way ANOVA test (Test 1)	Pre-test score of the experimental group	21.08 ± 1.699s	Pre-test score of the control group	21.26 ± 1.964s	0.894	$F(1,8) = 0.019$
One-way rmANOVA test (Test 2)	Pre-test score of the experimental group	21.08 ± 1.699s	Post-test score of the experimental group	14.01 ± 1.63s	<0.0001	$F(1,4) = 362.604$
One-way rmANOVA test (Test 3)	Pre-test score of the control group	21.26 ± 1.964s	Post-test score of the control group	16.59 ± 1.889s	<0.0001	$F(1,4) = 350.641$
One-way ANOVA test (Test 4)	Improvement of the experimental group	33.7 ± 3.5%	Improvement of the control group	22.1 ± 2.7%	0.0007	$F(1,8) = 27.905$

[a]Significant at $p < 0.05$ level

control group. These results demonstrate the AR-based virtual piano game is also capable of improving upper-extremity functions. However, the control group spent more time on the post-test. The improvements of the experimental group and the control group were calculated using the pre- and post-test scores given by Eq. (30.5). The improvement of the experimental group is 33.7 ± 3.5%, which is larger than the improvement of the control group, i.e., 22.1 ± 2.7%. Another One-way ANOVA test (Test 4) was conducted to compare the improvement between these two groups. The result of this test reveals that a significant difference exists between these two groups ($F(1,8) = 27.905$, $p = 0.0007$). These findings support the second hypothesis. Therefore, it can be concluded that as compared with non-stage-based AR-assisted exercising systems, e.g., the AR-based virtual piano game, the ARTHE system can enhance motor skill relearning.

$$\text{Improvement} = \frac{(\text{Pre-test score} - \text{Post-test score})}{\text{Pre-test score} \times 100} \quad (30.5)$$

In conclusion, this preliminary usability experiment has verified that the ARTHE system is a reliable tool for motor ability enhancement, and providing healthcare exercises in stages is more conducive to users than traditional non-stage-based AR-assisted exercise.

30.7 Conclusion and Future Work

This chapter presents a novel AR-assisted Three-stage Healthcare Exercising (ARTHE) system that enables people low upper-extremity functions to enhance upper-extremity functions through seven healthcare exercises. To follow the motor enhancement process, these exercises are categorized into the flaccid stage , the synergy stage, and the activities of daily living stage. Exercises in the synergy stage and the activities of daily living stage are composed of a virtual and a real object training phase. The OptiTrack motion tracking system, a web camera, and an AR-based data glove are used in this system for motion tracking and performance measurement. Evaluation methods have been formulated for these exercises. An experiment is performed to validate the usability of this system. The evaluation results reveal that this system is capable of improving motor skills of upper extremity. The comparison with the AR-based virtual piano game demonstrated that the ARTHE system is a more effective approach than the traditional non-stage-based AR-assisted exercising systems.

Some limitations still exist in this project. For users with low finger dexterity, wearing the data glove can be difficult or impossible. To solve this problem, bare hand tracking can be used in this system. In addition, the tracking volume created by three OptiTrack cameras limits the area of exercise. More cameras can be involved to build a bigger and more stable tracking volume. Another work that remains to be developed in further research is applying this system in the standard rehabilitation protocols. As ARTHE is developed following the process of the improvement of motor functions, it has the potential of assisting patients with damaged upper-extremity functions to recover motor skills. However, exercises in ARTHE should be modified to be targeted at specific patients. A long-period user study on related patients should be conducted in the next step.

References

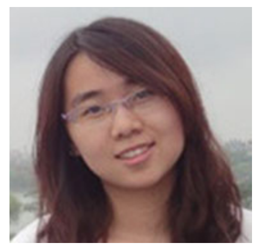
1. Cheng, Y.Y., Hsieh, W.L., Kao, C.L., Chan, R.C.: Principles of rehabilitation for common chronic neurologic diseases in the elderly. J. Clin. Gerontol. Geriatr. **3**(1), 5–13 (2012)
2. Siparsky, P.N., Kirkendall, D.T., Garrett, W.E.: Muscle changes in aging: understanding sarcopenia. Sports Health. **6**(1), 36–40 (2014)
3. Steffen, D., Bleser, G., Weber, M., Stricker, D., Fradet, L., Marin, F.: A personalized exercise trainer for elderly. In: Proceedings of the 5th International Conference on Pervasive Computing Technologies for Healthcare, Dublin, pp. 24–31 (2011)
4. Crow, J.L., Harmeling-van der Wel, B.C.: Hierarchical properties of the motor function sections of the Fugl-Meyer assessment scale for people after stroke: a retrospective study. Phys. Ther. **88**, 1554–1567 (2008)
5. Naghdi, S., Ansari, N.N., Mansouri, K., Hasson, S.: A neurophysiological and clinical study of Brunnstrom recovery stages in the upper limb following stroke. Brain Inj. **24**, 1372–1378 (2010)
6. Twitchell, T.E.: The restoration of motor function following hemiplegia in man. Brain. **74**(4), 443–480 (1951)

7. Morone, G., Spitoni, G.F., De Bartolo, D., Ghooshchy, S.G., Di Iulio, F., Paolucci, S., Zoccolotti, P., Iosa, M.: Rehabilitative devices for a top-down approach. Expert Rev. Med. Devices. **16**(3), 187–195 (2019)
8. Foulds, R.A., Saxe, D.M., Joyce, A.W., Adamovich, S.: Sensory-motor enhancement in a virtual therapeutic environment. Virtual Real. **12**(2), 87–97 (2008)
9. Zhao, M.Y., Ong, S.K., Nee, A.Y.C.: An augmented reality-assisted therapeutic healthcare exercise system based on bare-hand interaction. Int. J. Human Comput. Interact. **32**(9), 708–721 (2016)
10. Crocetta, T.B., de Araújo, L.V., Guarnieri, R., Massetti, T., Ferreira, F.H.I.B., de Abreu, L.C., de Mello Monteiro, C.B.: Virtual reality software package for implementing motor learning and rehabilitation experiments. Virtual Real. **22**(3), 199–209 (2018)
11. Ma, S., Varley, M., Shark, L.K., Richards, J.: Overcoming the information overload problem in a multiform feedback-based virtual reality system for hand motion rehabilitation: healthy subject case study. Virtual Real. **16**(4), 325–334 (2012)
12. Shen, Y., Gu, P., Ong, S.K., Nee, A.Y.C.: A novel approach in rehabilitation of hand-eye coordination and finger dexterity. Virtual Real. **16**, 161–171 (2012)
13. Aung, Y.M., Al-Jumaily, A.: AR based upper limb rehabilitation system in biomedical robotics and biomechatronics (BioRob). In: Proceedings of the 4th IEEE RAS & EMBS International Conference, Rome, pp. 213–218 (2012)
14. ALS worldwide. Range of Motion Exercises. https://alsworldwide.org/assets/misc/RANGE_OF_MOTION_EXERCISES_WITH_PHOTOS_copy.pdf . Accessed 6 Feb 2020
15. French, B., Thomas, L., Leathley, M., Sutton, C., McAdam, J., Forster, A., Watkins, C.: Does repetitive task training improve functional activity after stroke? A Cochrane systematic review and meta-analysis. J. Rehabil. Med. **42**(1), 9–15 (2010)
16. Woldag, H., Stupka, K., Hummelsheim, H.: Repetitive training of complex hand and arm movements with shaping is beneficial for motor improvement in patients after stroke. J. Rehabil. Med. **42**(6), 582–587 (2010)
17. Muellbacher, W., Ziemann, U., Boroojerdi, B., Cohen, L., Hallett, M.: Role of the human motor cortex in rapid motor learning. Exp. Brain Res. **136**(4), 431–438 (2001)
18. Carr, J.H., Shepherd, R.B.: Neurological Rehabilitation, 2nd edn. Elsevier/Churchill Livingstone, London (2010)
19. Chan, D.Y., Chan, C.C., Au, D.K.: Motor relearning programme for stroke patients: a randomized controlled trial. Clin. Rehabil. **20**, 191–200 (2006)
20. Sharman, M.J., Cresswell, A.G., Riek, S.: Proprioceptive neuromuscular facilitation stretching: mechanisms and clinical implications. Sports Med. **36**(11), 929–939 (2006)
21. Furmanek, M.P., Schettino, L.F., Yarossi, M., Kirkman, S., Adamovich, S.V., Tunik, E.: Coordination of reach-to-grasp in physical and haptic-free virtual environments. J. Neuroeng. Rehabil. **16**, 78 (2019)
22. Harris, J.E., Eng, J.J.: Strength training improves upper-limb function in individuals with stroke a meta-analysis. Stroke. **41**(1), 136–140 (2010)
23. Blennerhassett, J., Dite, W.: Additional task-related practice improves mobility and upper limb function early after stroke: a randomised controlled trial. Aust. J. Physiotherapy. **50**, 219–224 (2004)
24. Dean, C.M., Richards, C.L., Malouin, F.: Task-related circuit training improves performance of locomotor tasks in chronic stroke: a randomized, controlled pilot trial. Arch. Phys. Med. Rehabil. **81**, 409–417 (2000)
25. Malouin, F., Richards, C.L., McFadyen, B., Doyon, J.: New perspectives of locomotor rehabilitation after stroke. Médecine/Sciences. **19**(10), 994–998 (2003)
26. Inman, D.P., Loge, K., Cram, A., Peterson, M.: Learning to drive a wheelchair in virtual reality. J. Spec. Educ. Technol. **26**(3), 21–34 (2011)
27. Wang, B., Shen, M., Wang, Y., He, Z., Chi, S., Yang, Z.: Effect of virtual reality on balance and gait ability in patients with Parkinson's disease: a systematic review and meta-analysis. Clin. Rehabil. **33**(7), 1130–1138 (2019)
28. Crosbie, J., Lennon, S., McNeill, M., McDonough, S.: Virtual reality in the rehabilitation of the upper limb after stroke: the user's perspective. CyberPsychol. Behav. **9**(2), 137–141 (2006)
29. Ustinova, K., Perkins, J., Szostakowski, L., Tamkei, L., Leonard, W.: Effect of viewing angle on arm reaching while standing in a virtual environment: potential for virtual rehabilitation. Acta Psychol. **133**(2), 180–190 (2010)
30. Ma, H.I., Hwang, W.J., Wang, C.Y., Fang, J.J., Leong, I.F., Wang, T.Y.: Trunk-arm coordination in reaching for moving targets in people with Parkinson's disease: comparison between virtual and physical reality. Hum. Mov. Sci. **31**, 1340–1352 (2012)
31. Alamri, A., Eid, M., Iglesias, R., Shirmohammadi, S., El Saddik, A.: Haptic virtual rehabilitation exercises for poststroke diagnosis. IEEE Trans. Instrum. Meas. **57**(9), 1876–1884 (2008)
32. Boian, R., Sharma, A., Han, C., Merians, A.S., Burdea, G.S., Adamovich, S.V., Poizner, H.: Virtual reality-based post-stroke hand rehabilitation. Stud. Health Technol. Inform. **85**, 64–70 (2002)
33. Jack, D., Boian, R., Merians, A.S., Tremaine, M., M., G.S. Burdea, S.V. Adamovich, H.: Poizner: virtual reality-enhanced stroke rehabilitation. IEEE Trans. Neural Syst. Rehabil. Eng. **9**(3), 308–318 (2001)
34. Adamovich, S.V., Merians, A.S., Boian, R., Lewis, J.A., Tremaine, M., Burdea, G.S., Poizner, H.: A virtual reality-based exercise system for hand rehabilitation post-stroke. Presence Teleop. Virt. Environ. **14**(2), 161–174 (2005)
35. Shakra, I., Orozco, M., El Saddik, A., Shirmohammadi, S., Lemaire, E.: VR-based hand rehabilitation using a haptic-based framework. In: Proceedings of the IEEE Conference on Instrumentation and Measurement Technology, Sorrento, pp. 1178–1181 (2006)
36. Alamri, A., Iglesias, R., Eid, M., El Saddik, A., Shirmohammadi, S., Lemaire, E.: Haptic exercises for measuring improvement of post-stroke rehabilitation patients. In: Proceedings of the IEEE International Workshop on Medical Measurement and Applications, Warsaw, pp. 1–6 (2007)
37. Lambercy, O., Dovat, L., Gassert, R., Burdet, E., Teo, C.L., Milner, T.: A haptic knob for rehabilitation of hand function. IEEE Trans. Neural Syst. Rehabil. Eng. **15**(3), 356–366 (2007)
38. Xu, Z., Yu, H., Yan, S.: Motor rehabilitation training after stroke using haptic handwriting and games. In: Proceedings of the 4th International Convention on Rehabilitation Engineering & Assistive Technology, Shanghai, p. 31 (2010)
39. Gupta, A., O'Malley, M.K.: Design of a haptic arm exoskeleton for training and rehabilitation. IEEE/ASME Trans. Mechatron. **11**(3), 280–289 (2006)
40. Lövquist, F., Dreifaldt, U.: The design of a haptic exercise for post-stroke arm rehabilitation. In: Proceedings of the 6th International Conference on Disability, Virtual Reality and Associate Technologies, Esbjerg, pp. 18–20 (2006)
41. Song, A., Wu, J., Qin, G., Huang, W.: A novel self-decoupled four degree-of-freedom wrist force/torque sensor. Measurement. **40**(9), 883–891 (2007)
42. Kayyali, R., Alamri, A., Eid, M., Iglesias, R., Shirmohammadi, S., El Saddik, A.: Occupational therapists' evaluation of haptic motor rehabilitation. In: Proceedings of the 29th Annual International Conference of the IEEE, Lyon, pp. 4763–4766 (2007)

43. Lo, H.S., Xie, S.Q.: Exoskeleton robots for upper-limb rehabilitation: state of the art and future prospects. Med. Eng. Phys. **34**(3), 261–268 (2012)
44. Loureiro, R., Amirabdollahian, F., Topping, M., Driessen, B., Harwin, W.: Upper limb robot mediated stroke therapy – GENTLE/s approach. Auton. Robot. **15**(1), 35–51 (2003)
45. Loureiro, R., Collin, C., Harwin, W.: Robot aided therapy: challenges ahead for upper limb stroke rehabilitation. In: Proceedings of the Fifth International Conference on Disability, Virtual Reality and Associated Technologies, Oxford, UK, pp. 33–39 (2004)
46. Burdea, G.C.: Virtual rehabilitation-benefits and challenges. Methods Inf. Med. **42**(5), 519–523 (2003)
47. Al-Issa, H., Regenbrecht, H., Hale, L.: Augmented reality applications in rehabilitation to improve physical outcomes. Phys. Ther. Rev. **17**(1), 16–28 (2012)
48. Jaffe, D.L.: Using augmented reality to improve walking in stroke survivors. In: Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication, Millbrae, pp. 79–83 (2003)
49. Espay, A.J., Baram, Y., Dwivedi, A.K., Shukla, R., Gartner, M., Gaines, L., Revilla, F.J.: At-home training with closed-loop augmented-reality cueing device for improving gait in patients with Parkinson disease. J. Rehabil. Res. Dev. **47**, 573–581 (2010)
50. Luo, X., Kline, T., Fischer, H.C, Stubblefield, K.A., Kenyon, R.V., Kamper, D.G.: Integration of augmented reality and assistive devices for post-stroke hand opening rehabilitation. In: Proceedings of the 27th Annual International Conference of Engineering in Medicine and Biology Society, Shanghai, pp. 6855–6858 (2005)
51. Lee, R.G., Tien, S.C., Chen, C.C., Chen, Y.Y.: Development of an augmented reality-oriented game system for stroke rehabilitation assessment. Biomed. Eng. Appl. Basis Commun. **24**(5), 435–445 (2012)
52. Wang, H.S., Hsu, C., Chiu, D., Tsai, S.N.: Using augmented reality gaming system to enhance hand rehabilitation. In: Proceedings of the 2nd International Conference on Education Technology and Computer, Shanghai, pp. 243–246 (2010)
53. Chinthammit, W., Merritt, T., Pedersen, S., Williams, A., Visentin, D., Rowe, R., Furness, T.: Ghostman: augmented reality application for telerehabilitation and remote instruction of a novel motor skill. Biomed. Res. Int. **2014**, 1–7 (2014)
54. Lin, J.K., Cheng, P.H., Su, Y., Wang, S.Y., Lin, H.W., Hou, H.C., Su, M.J.: Augmented reality serious game framework for rehabilitation with personal health records. In: Proceedings of the 13th IEEE International Conference on e-Health Networking Applications and Services, Columbia, pp. 197–200 (2011)
55. Regenbrecht, H., McGregor, G., Ott, C., Hoermann, S., Schubert, T., Hale, L., Hoermann, J., Dixon, B., Franz, E.: Out of reach? A novel AR interface approach for motor rehabilitation. In: Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, pp. 219–228 (2011)
56. Alamri, A., Cha, J., Eid, M., El Saddik, A.: Evaluating the post-stroke patients progress using an Augmented Reality Rehabilitation system. In: Proceedings of the IEEE International Workshop on Medical Measurements and Applications, Cetraro, pp. 89–94 (2009)
57. Alamri, A., Cha, J., El Saddik, A.: AR-REHAB: an augmented reality framework for poststroke-patient rehabilitation. IEEE Trans. Instrum. Meas. **59**, 2554–2563 (2010)
58. Lange, B., Chang, C.Y., Suma, E., Newman, B., Rizzo, A.S., Bolas, M.: Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. In: Proceedings of the Annual IEEE International Conference of Engineering in Medicine and Biology Society, Boston, pp. 1831–1834 (2011)
59. Ponto, K., Kimmel, R., Kohlmann, J., Bartholomew, A., Radwin, R.G.: Virtual exertions: a user interface combining visual information, kinesthetics and biofeedback for virtual object manipulation. In: Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI), Costa Mesa, pp. 85–88 (2012)
60. Klein, A., Assis, G.A.D.: A markerless Augmented Reality tracking for enhancing the user interaction during virtual rehabilitation. In: Proceedings of the XV Symposium on Virtual and Augmented Reality (SVR), Washington, pp. 117–124 (2013)
61. Watanabe, Y., Matsutani, A., Niikura, T., Komuro, T., Ishikawa, M.: High-speed estimation of multi-finger position and pose for input interface of the mobile devices. In: Proceedings of the IEEE 1st Global Conference on Consumer Electronics (GCCE), Tokyo, pp. 228–232 (2012)
62. Metcalf, C.D., Robinson, R., Malpass, A.J., Bogle, T.P., Dell, T.A., Harris, C., Demain, S.H.: Markerless motion capture and measurement of hand kinematics: validation and application to home-based upper limb rehabilitation. IEEE Trans. Biomed. Eng. **60**(8), 2184–2192 (2013)
63. Fugl-Meyer, A., Jääskö, L., Leyman, I., Olsson, S., Steglind, S.: The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. Scand. J. Rehabil. Med. **7**(1), 13–31 (1975)
64. Wolf, S.L., Catlin, P.A., Ellis, M., Archer, A.L., Morgan, B., Piacentino, A.: Assessing Wolf motor function test as outcome measure for research in patients after stroke. Stroke. **32**(7), 1635–1639 (2001)
65. Hsieh, Y.W., Wu, C.Y., Lin, K.C., Chang, Y.F., Chen, C.L., Liu, J.S.: Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. Stroke. **40**(4), 1386–1391 (2009)
66. McDonnell, M.: Action research arm test. Aust. J. Physiotherapy. **54**(3), 220 (2008)
67. Lyle, R.C.: A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int. J. Rehabil. Res. **4**(4), 483–492 (1981)
68. Yozbatiran, N., Der-Yeghiaian, L., Cramer, S.C.: A standardized approach to performing the action research arm test. Neurorehabil. Neural Repair. **22**(1), 78–90 (2008)
69. Bonato, P.: Advances in wearable technology and applications in physical medicine and rehabilitation. J. Neuroeng. Rehabil. **2**(1), 2 (2005)
70. Williams, N., Penrose, J., Caddy, C., Barnes, E., Hose, D., Harley, P.: A goniometric glove for clinical hand assessment construction, calibration and validation. J. Hand Surg. (Br. Eur. Vol.). **25**(2), 200–207 (2000)
71. Fahn, C.S., Sun, H.: Development of a fingertip glove equipped with magnetic tracking sensors. Sensors. **10**(2), 1119–1140 (2010)
72. Roberts, L., Singhal, G., Kaliki, R.: Slip detection and grip adjustment using optical tracking in prosthetic hands. In: Proceedings of the Annual IEEE International Conference of Engineering in Medicine and Biology Society, Boston, pp. 2929–2932 (2011)
73. Hermsdörfer, J., Hagl, E., Nowak, D., Marquardt, C.: Grip force control during object manipulation in cerebral stroke. Clin. Neurophysiol. **114**(5), 915–929 (2003)
74. Blennerhassett, J.M., Carey, L.M., Matyas, T.A.: Grip force regulation during pinch grip lifts under somatosensory guidance: comparison between people with stroke and healthy controls. Arch. Phys. Med. Rehabil. **87**, 418–429 (2006)
75. Schaefer, S.Y., DeJong, S.L., Cherry, K.M., Lang, C.E.: Grip type and task goal modify reach-to-grasp performance in post-stroke hemiparesis. Mot. Control. **16**(2), 245–264 (2012)
76. Keil, A., Elbert, T., Taub, E.: Relation of accelerometer and EMG recordings for the measurement of upper extremity movement. J. Psychophysiol. **13**(2), 77–82 (1999)

77. Manson, A., Brown, P., O'Sullivan, J., Asselman, P., Buckwell, D., Lees, A.: An ambulatory dyskinesia monitor. J. Neurol. Neurosurg. Psychiatry. **68**(2), 196–201 (2000)
78. Thielgen, T., Foerster, F., Fuchs, G., Hornig, A., Fahrenberg, J.: Tremor in Parkinson's disease: 24-hr monitoring with calibrated accelerometry. Electromyogr. Clin. Neurophysiol. **44**(3), 137–146 (2004)
79. Parnandi, A., Wade, E., Mataric, M.: Motor function assessment using wearable inertial sensors. In: Proceedings of the Annual IEEE International Conference of Engineering in Medicine and Biology Society, Buenos Aires, pp. 86–89 (2010)
80. Uswatte, G., Giuliani, C., Winstein, C., Zeringue, A., Hobbs, L., Wolf, S.L.: Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. Arch. Phys. Med. Rehabil. **87**(10), 1340–1345 (2006)
81. Choi, Y.: Ubi-REHAB: An android-based portable Augmented Reality stroke rehabilitation system using the eGlove for multiple participants. In: Proceedings of the International Conference on Virtual Rehabilitation (ICVR), Zurich, pp. 1–2 (2011)
82. Choi, Y., Gordon, J., Park, H., Schweighofer, N.: Feasibility of the adaptive and automatic presentation of tasks (ADAPT) system for rehabilitation of upper extremity function post-stroke. J. Neuroeng. Rehabil. **8**, 42 (2011)
83. Parsons, T.D., Iyer, A., Cosand, L., Courtney, C., Rizzo, A.A.: Neurocognitive and psychophysiological analysis of human performance within virtual reality environments. Stud. Health Technol. Inform. **142**, 247–252 (2009)
84. Fritz, S.L., Light, K.E., Patterson, T.S., Behrman, A.L., Davis, S.B.: Active finger extension predicts outcomes after constraint-induced movement therapy for individuals with hemiparesis after stroke. Stroke. **36**, 1172–1177 (2005)
85. Mathiowetz, V., Weber, K., Kashman, N., Volland, G.: Adult norms for the nine hole peg test of finger dexterity. Occup. Ther. J. Res. **5**, 24–38 (1985)
86. Hester, T., Hughes, R., Sherrill, D.M., Knorr, B., Akay, M., Stein, J., Bonato, P.: Using wearable sensors to measure motor abilities following stroke. In: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks, Massachusetts, pp. 4–8 (2006)
87. Zhao, M.Y., Ong, S.K., Nee, A.Y.C.: Augmented reality-assisted rehabilitation of activities of daily living. In: Proceedings of the 5th International Conference on Internet Technologies & Society, New Taipei City, pp. 89–93 (2014)
88. Zhang, D., Shen, Y., Ong, S.K., Nee, A.Y.C.: An affordable augmented reality based rehabilitation system for hand motions. In: Proceedings of the International Conference on Cyberworlds, Singapore, pp. 346–353 (2010)
89. Barnes, M.P.: Rehabilitation after traumatic brain injury. Br. Med. Bull. **55**, 927–943 (1999)
90. Kaber, D., Tupler, L.A., Clamann, M., Gil, G.H., Zhu, B., Swangnetr, M., Jeon, W., Zhang, Y., Qin, X., Ma, W., Lee, Y.S.: Evaluation of an augmented virtual reality and haptic control interface for psychomotor training. Assist. Technol. **26**, 51–60 (2014)
91. Zar, J.H.: Biostatistical Analysis. Prentice Hall, Upper Saddle River (1998)

Soh Khim Ong is Associate Professor at NUS. Her research interests are virtual and augmented reality applications in manufacturing and life cycle engineering. She is Fellow of CIRP. She has published 5 books and over 330 international refereed journal and conference papers. She received 2004 OYME Award from SME and 2009 Emerging Leaders Award in Academia (US Society for Women Engineers).



M.Y. Zhao received a BS degree in Biomedical and Information Engineering from Northeastern University, Shenyang, P.R.C., in 2010, specializing in lung segmentation in CT images. She received her PhD from the Mechanical Engineering Department, National University of Singapore, Singapore, specializing in augmented reality-assisted healthcare exercising system for upper-extremities.

Andrew Yeh Ching Nee received his Ph.D. and D.Eng. degrees from UMIST, Manchester. He is currently Emeritus Professor in the Department of Mechanical Engineering of NUS. He is Fellow of CIRP, SME, and Academy of Engineering Singapore. His research interests are in manufacturing engineering, augmented reality, and digital twin. He is editor-in-chief of IJAMT and Executive editor-in-chief of AIM.

Augmented Reality for Cognitive Impairments

31

Zhanat Makhataeva, Tolegen Akhmetov, and Huseyin Atakan Varol

Contents

31.1 **Introduction** 765

31.2 **Background Research** 766
31.2.1 Devices and Tools for AR 766
31.2.2 Computing Paradigms for AR 768

31.3 **Current Development** 769
31.3.1 Human Memory System 769
31.3.2 AR Applications in Research of Human Memory and Cognition 771
31.3.3 AR Applications in Cognitive Load Research 776
31.3.4 AR Applications for Restoring Perception 781
31.3.5 Issues and Challenges in the Application of AR for Cognitive Impairments 788

31.4 **Conclusions** 789

References 789

Abstract

Augmented reality (AR) is a rapidly developing technology that has introduced new approaches for the design of human technology interaction in various fields. However, application of AR in the area of assistive systems for cognitively impaired people is not deeply studied yet. In this chapter, we investigate the state of the art of AR in assistive systems for the cognitively impaired. Specifically, we will investigate the role of AR during the design, implementation, and performance assessment of perception and memory augmentation assistive systems. We start with a summary of various technologies utilized for the development of AR systems, including sensor and camera technology, data visualization methods, computing paradigms, and intelligent data processing algorithms. Then, we discuss the fundamental mechanisms behind human memory system and look at the examples of the first technology-based human memory and intelligence augmentation systems. We overview various methods for estimation of the human cognitive state and mental workload utilized during the evaluations of such assistive systems in human involved experimental studies. Our main objective in this work is to find out the current status, challenges, and future perspectives of AR in the research of human memory and perception augmentation systems for the people with cognitive impairments.

Keywords

Augmented reality · Cognitive impairments · Perception impairments · Assistive systems · Artificial intelligence · Computing paradigms · Cognitive load · Working memory

31.1 Introduction

According to the statistics of World Bank [1], approximately 10% of the world population have mental, neurological, and substance use disorders [2]. Twenty percent of these are children and adolescents. With the aging societies, the ratio of elderly people in the society also increases. In general, elderly have higher risk to become affected by memory, cognitive, and mental disorders. Such dysfunctions negatively impact the social relationships of the elderly and their close contacts with behavioral, psychological, and physiological consequences. Also, there are many affected by perception-related impairments, such as vision and hearing deficits.

Over the years, various technology-enhanced approaches to help people were developed. This leads to the concept of assistive technologies (AT) [3], which is an umbrella term describing solutions designed to address issues of people affected by a disorder or an injury. Within the context of cognitive impairments, Ong et al. [4] assumed that AT could allow people with cognitive disorders to live in their familiar

Z. Makhataeva · T. Akhmetov · H. A. Varol (✉)
Department of Robotics, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan City, Kazakhstan
e-mail: zhanat.mahataeva@nu.edu.kz; tolegen.akhmetov@nu.edu.kz; ahvarol@nu.edu.kz

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_31

environments and reduce the cost for long-term care institutions. The developments in computer science, especially in computer vision (CV), introduced new approaches within the research of AT. CV is the area of computer science that aims to replicate the human vision system by extracting rich information from digital images and videos. Leo et al. [5] discussed applications of AT enhanced with CV to address a number of areas such as cognitive restoration, personal mobility, sensory functions, and activities of daily living (ADLs). In addition, integration of CV into AT led to the emergence of assistive systems that are capable of self-localization, i.e., mapping and tracking the position of an assistive system or human with help of sensors. This, in turn, enabled the assistive devices to answer questions of its user such as "Where I am?" and "What I have seen?".

In recent years, AR has been employed frequently in the research of AT. AR is known as the environment where virtual objects are incorporated on top of the real environment in 3D space [6]. The history of AR goes back to the invention of virtual reality (VR) by Sutherland [7] in 1965. The origins of AR and VR can actually be further traced back to an invention of Charles Wheatstone. In 1838, Wheatstone created a stereoscope that allowed people to view a pair of separate images for each eye as a combined three-dimensional image [8]. While users are fully immersed in three-dimensional worlds in VR, augmented reality (AR) presents an integrated version of the virtual world with the real physical world in 3D space [6]. AR can incorporate several interface modalities such as voice control, audio/visual support, and gesture recognition. Also, AR introduces intuitive means of interaction with surrounding technology and environment [9]. In addition, AR can be used to compensate for the perception and memory impairments. It can also bolster cognitive abilities of healthy people via enhancing comprehension, analysis, and perception of the surrounding environment. In this chapter, we study recent applications where AR was integrated to assistive systems designed to help people with memory- and perception-related cognitive impairments. Based on the literature analysis, we summarize the general trends, challenges, design solutions, and communication methods for human-technology interaction in the research of AR-enhanced assistive systems. In addition, we consider the potential contributions of AR in the area of diagnosis of cognitive impairments, perception enhancement, and novel ways for setting the assistance. We believe this work will help developers of assistive systems in the field of cognitive impairments to: (1) understand mechanisms behind human memory and cognition, (2) familiarize with methods utilized for estimation of the cognitive status and cognitive load during the design of assistive systems, and (3) get a general picture of the current status, advantages, and issues of AR, as well as to acquire a vision for better integration of AR into the future research of assistive systems and memory/perception augmentation issues for the cognitively impaired.

The rest of the chapter is divided into two parts: background research and current development. In the background research (Sect. 31.2), we review devices and tools used in AR in Sect. 31.2.1 and computation paradigms for AR in Sect. 31.2.2. After introducing the technological landscape, we discuss the current developments in Sect. 31.3. The mechanisms underlying human memory and cognition are covered in Sect. 31.3.1. The discussion on how AR is incorporated into the research of human memory and cognition is presented in Sect. 31.3.2. This is followed by Sect. 31.3.3 which introduces cognitive load estimation techniques and summarizes AR-enhanced applications in the research of cognitive load and assistive systems. Section 31.3.4 highlights perception-related cognitive impairments as the playground for future AR research. It also summarizes AR-enhanced assistive systems addressing the perception-related issues of the cognitively impaired people. A summary of issues and challenges in the integration of AR into the research of cognitive impairments, specifically in memory and perception augmentation, appears in Sect. 31.3.5. Conclusion and predictions on future research prospects are drawn in Sect. 31.4.

31.2 Background Research

31.2.1 Devices and Tools for AR

There are different types of technology utilized in the research of AR, which are head-mounted displays (HMDs), mobile devices, and projector systems. Each group of the devices led to the development of wearable, mobile, and spatial types of AR, correspondingly. Two decades ago, first examples of AR-capable HMDs were cumbersome, had high latency, short battery life, limited computation power, small field of view, low-resolution displays, and had optics and calibration issues. However, present technological advances resulted in the invention of different types of sensors, cameras, and computational units. These significantly benefited research of AR. Specifically, miniaturization of the technology enabled to reduce the size and weight of AR-capable HMDs. In addition, advances in optics, environmental sensors, and image data processing algorithms led to the better tracking of the real environment and more accurate perception of AR features. However, modern wearables still have to overcome issues related to short battery life, small field of view, and see-through display resolution. Popular AR-capable HMDs utilized in the research of AR include Microsoft HoloLens, Magic Leap, Google Glass, Epson MOVERIO, and Vuzix Blade. According to the recent tech news, AR has been successfully integrated with software developed for real-time

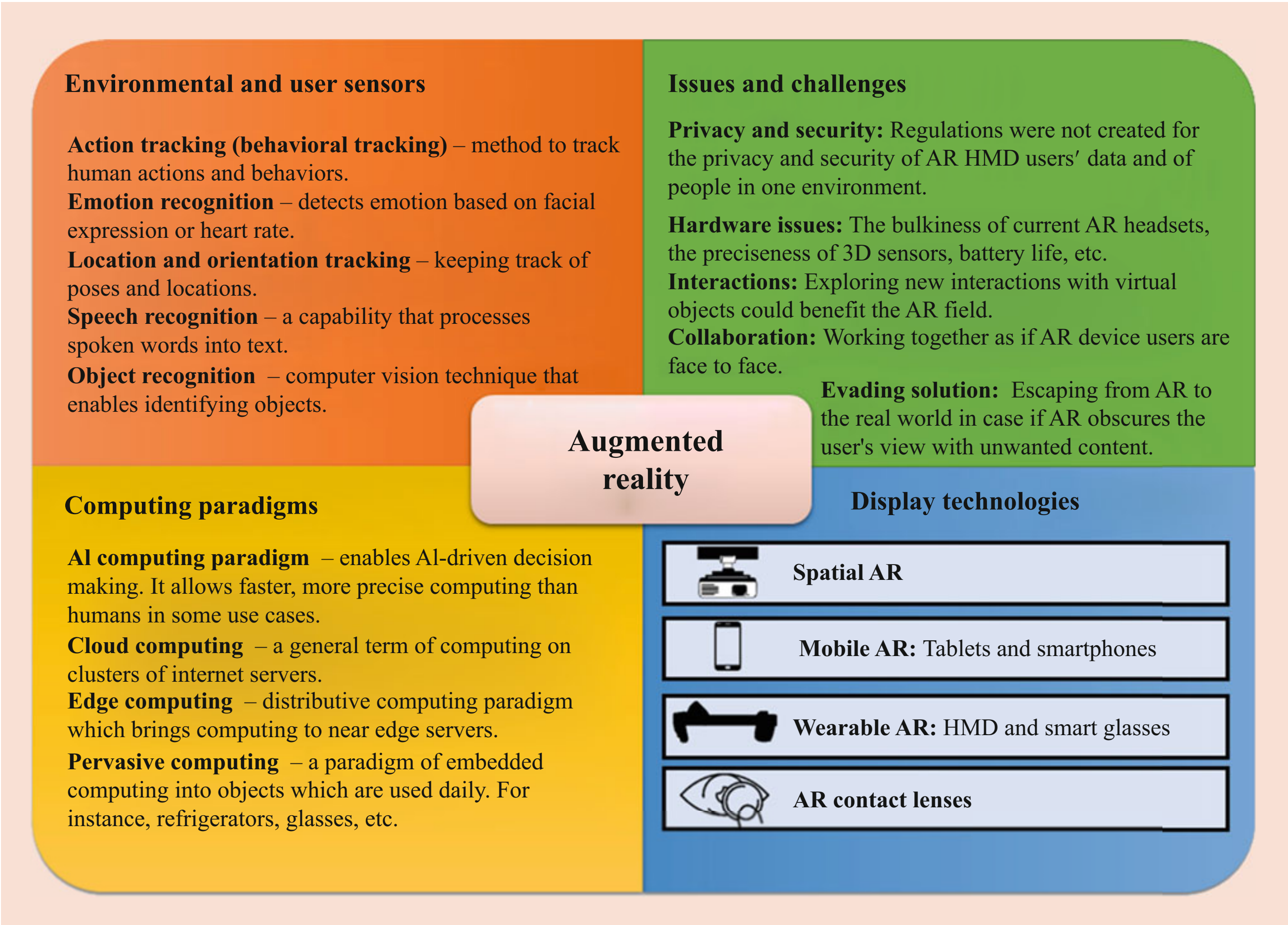


Fig. 31.1 Different sensors, computing paradigms, software, and display technologies for the state-of-the-art AR systems

tracking of drones [10]. Specifically, Microsoft HoloLens was utilized to visualize the spatial relationships between multiple flight paths and objects during the flight. Also, recently in automotive industry through AR environment service technicians collaborated with remote experts using HoloLens 2 and Microsoft Dynamics 365 application.

There are several requirements for the wearable devices in AR. First, for setting the AR environment wearables should be aware of its surroundings and current context of the user. This context sensitivity is achieved by integrating various sensors and camera technology into AR devices (see Fig. 31.1). Frequently used camera technology in AR are time of flight (ToF), depth, infrared, and stereo cameras. Popular camera-based technique utilized for placing virtual context into the real environment uses image markers (e.g., Vuforia [11]) or fiducial markers (e.g., ArUco [12, 13]). In order to track the movement of the device and define its orientation inertial measurement unit (IMU) sensors are used. As illustrated in Fig. 31.2, Microsoft HoloLens 2 is a noteworthy example to demonstrate the sensor density of new generation of AR goggles. Following sensors are utilized for sensing the environment and data collection: Visible Light Environment Tracking cameras, depth cameras, two versions of the IR-reflectivity stream, accelerometer, gyroscope, and magnetometer. The depth camera has two modes: the short-throw mode also called articulated hand tracking (AHAT) mode is responsible for hand tracking, while the long-throw mode is employed for spatial mapping. Environment cameras are utilized by the system for head tracking and map building. Depth of the surrounding environment is computed with the help of infrared sensors. In general, there are several ways to set the interaction between the human and the system in the AR environment. They are hand gesture control, voice control, and external controller-based communication. Going back to HoloLens 2 system, it supports gesture and voice control. However, the first generation of the HMD is capable to set the interaction with a help of external controller known as clicker. Clicker is connected to the main device via the Bluetooth network. There are also other examples of digital glasses in the market. For example, there is Smart Swim digital glasses from Vuzix designed for water sports, and Vuzix M-400, water- and dust-proof AR glasses designed for

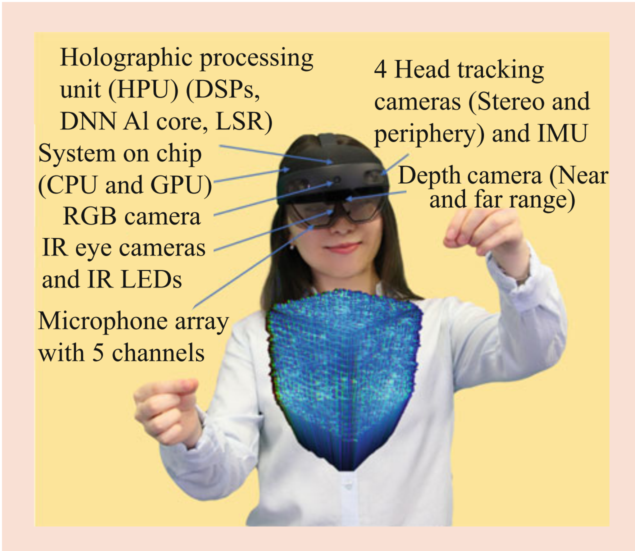


Fig. 31.2 Microsoft HoloLens 2 smart glass with annotated sensors and computational elements

difficult weather conditions. Glasses are equipped with GPS, Bluetooth, and Wi-Fi, enabling users to track performance and receive signals from other devices. Modern AR glasses such as Vuzix Blade and Hololens 2 contain an array of noise-canceling microphones that enable their use in noisy environments. Overall, it is worth to mention that issues of weight, size, thermal properties, battery life, as well as issues with optics and see-through display resolution of AR-capable digital glasses are still considered as the main sources of challenges and limitations in the research of wearable AR.

In addition to digital glasses, smartphones are also popular in AR. New generation of smartphones, tablets, and other mobile devices are equipped with smart features such as spatial mapping and object recognition. For example, in [14], smartphone was used for navigation assisted by AR-enhanced automatic object recognition framework. Also, many modern smartphones and tablets are equipped with ToF depth cameras and LiDAR sensors, which are utilized for the creation of 3D-immersive environment. 3D scanners can be paired with mobile devices in order to develop AR environment as well. Interestingly, mobile devices can be used to set up the WebAR environment, in which web browsers of mobile devices are utilized to acquire and display AR features [15]. General trends show up that mobile devices carry a significant potential for providing cost-efficient AR experience to people of twenty-first century.

In addition to wearable and mobile AR, there is also spatial AR environment, where projectors are utilized to superimpose or integrate virtual 3D objects onto a real physical environment. Within spatial AR, there is a concept of spatial user interface (SUI) [16], which lets users freely interact with both 3D virtual environment and traditional 2D workstation applications. In general, SAR has its unique constraints, which are the following: (1) SAR can be projected onto surfaces and cannot be projected in midair; (2) projector brightness limits the use in outdoor environments; and (3) it is vital for SAR to have geometric surfaces that must be known a priori. The reader is referred to the book [17] for a detailed treatment of SAR. Such limitation for outdoor use is also shared by the modern optical see-through (OST) HMDs [18]. Specifically, AR imagery of the HMDs loses the contrast between the physical environment in bright outdoor lighting.

Various software tools were introduced in recent years for simplifying the design and use of AR, including game development platform Unity, Vuforia engine, Mixed Reality Toolkit, and Mixed Reality capabilities in Robotic Operating System (ROS) (see Fig. 31.1). Game development engine Unity was enhanced with Mixed and AR Studio (MARS) to track objects and events using 3D objects as markers. Recently, Khronos group released OpenXR, a royalty-free standard for AR and VR devices and platforms. Most of the browsers enable WebXR Device Application Programming Interface (API) using OpenXR allowing consumers to experience 3D AR through their browsers. Also, major tech companies, such as Microsoft, Oculus, Samsung, Magic Leap, Apple, Unity Technologies, and Epic Games (developer of the Unreal Engine) updated their products for OpenXR compatibility. Overall, currently developers can design cross-platform applications or websites, and host online AR/VR environment making consumers experience AR/VR through their browsers or AR/VR headsets.

31.2.2 Computing Paradigms for AR

Widespread Internet connectivity and miniaturization of computing elements led to the emergence of different computing paradigms (see Fig. 31.1). The first one, artificial intelligence (AI) computing paradigm includes deterministic and nondeterministic algorithms which push technologies of speech and object recognition forward. The second paradigm, edge computing is a distributed processing paradigm that involves placing service provisioning, data, and intelligence closer to users and devices. Famous examples of edge computing are autonomous cars, smart street lights, automated industrial machines, smart homes, and mobile devices. In the third one, cloud computing, data is gathered and processed in the "cloud," actually a distant data center with substantial processing and storage capacity. The fourth paradigm, termed ubiquitous or pervasive computing, embeds computing to everyday objects instead of making computations in separate devices. In ubiquitous computing, computers may exist in the form of everyday objects such as a refrigerator or glasses.

AI computing paradigm welcomes new algorithms in recent years widening the possible areas of usage including AR. AI is the science of creating intelligent agents that can take actions based on the information received from the environment [19]. The current AI systems can match human ability and even surpass it in narrow tasks such as speech recognition, handwritten character recognition, and spam mail classification. Machine learning (ML), a branch of AI, analyzes data to find patterns which it then uses to make certain decisions. ML focuses on the questions of how to construct computer systems that improve by experience and what are the fundamental laws that govern these learning systems [20]. The state-of-the-art method for ML is deep learning (DL). It uses artificial neural networks with multiple layers to learn features obviating the need for feature engineering by experts [21]. ML algorithms have been utilized in AR for various purposes. In [22], by integration of ML and AR, the researchers designed a framework for assisting medical residents during surgery via enhancing their understanding of spatial relations between tools, implants, and anatomical objects. The learning algorithm used depth and X-ray images for object recognition, extracting their positions/orientations, and generating the augmented vision. Ichihashi et al. [23] proposed a visibility management algorithm for spatial AR, VisLP, that placed annotations and linkage lines based on the estimation of the visibility using ML. Advances in computing, especially the emergence of graphical processing units (GPUs), enabled the use of high-performance DL algorithms for object recognition. Accurate and fast object recognition adds a wide range of capabilities to AR. Sutanto et al. [24] presented a markerless AR method utilizing deep network for object detection. Hoppenstedt et al. [25] implemented an object recognition framework where voice commands from AR glasses were used for labeling objects.

For demanding computational tasks, AR leverages edge computing. Specifically, AR goggles are usually used as clients, while computationally intensive operations such as DL are performed on the server side. Microsoft developed the research mode extension for HoloLens to enable seamless communication with the server. Using this mode, devices can transmit data acquired by built-in sensors to a separate edge computing device. Recently, mobile laptops (“Backpacks”) enhanced with AR/VR features appeared in the market. Researchers used these specialized mobile laptops for DL-based object recognition to implement an adaptive projection AR system [26].

In line with the ubiquitous computing philosophy, the processing units inside the AR systems reduce latency by eliminating the data transmission between the edge and cloud computing devices. Microsoft HoloLens 1 and 2 provide a glimpse into the future of smart glasses with ubiquitous computing. These devices contain holographic processing units (HPUs), which make them self-contained holographic computers. However, it takes around 1 min for the HPU of Hololens 1 to run the state-of-the-art object recognition engine “YOLO” for a single frame [27]. Therefore, in heavy tasks, the limited computation power of the built-in processor in an AR system might decrease quality of service.

Cloud computing delivers on-demand computing services through the Internet. Microsoft Azure, Google Cloud, and Amazon Web Services are well-known service providers. These companies provide APIs for a wide spectrum of users from academia and industry. For example, the Custom Vision API of Azure was developed to train an object recognition engine for flower classification in [28]. In [29], authors presented a client-server AR system that directed demanding computational tasks to clusters. On the other hand, mobile phones were exploited for basic image processing tasks for reducing transmission costs. Overall, cloud computing integrated with AR can enhance the 3D immersive interactive services of users if the bandwidth and latency are not technical bottlenecks.

31.3 Current Development

31.3.1 Human Memory System

In this subsection, we will provide an overview of the mechanisms underlying human memory and cognition. A high-level illustration of human memory and the possible areas that can benefit from AR are shown in Fig. 31.3.

Human memory is a complex system for receiving and storing information comprised of auditory, visual, sound, smell, touch, and taste senses. The memory system is comprised of three components, which are sensory memory, working memory, and short-/long-term memory. Sensory memory is responsible for data acquisition from the surrounding environment and sending the data into the working memory. Working memory processes the sensed data and performs functions such as reasoning, learning, and understanding [30]. In other words, working memory acts as a register for incoming auditory, verbal, and visual signals. On the other hand, long-term memory forms and stores schemes that make humans perform tasks automatically, implying successful learning of the task. The processed information in the working memory goes further into the short-/long-term memory. The short-/long-term memory system is where the information is actually stored and can be retrieved upon the request of the human brain. It is assumed that memories can be stored for different periods ranging from fractions of a second (short-term memory) to the whole lifespan (long-term memory). Correspondingly, consumed storage space may also vary from a tiny buffer within the short-term memory up to a substantial capacity within the long-term memory. Long-term memory is comprised of conscious and unconscious

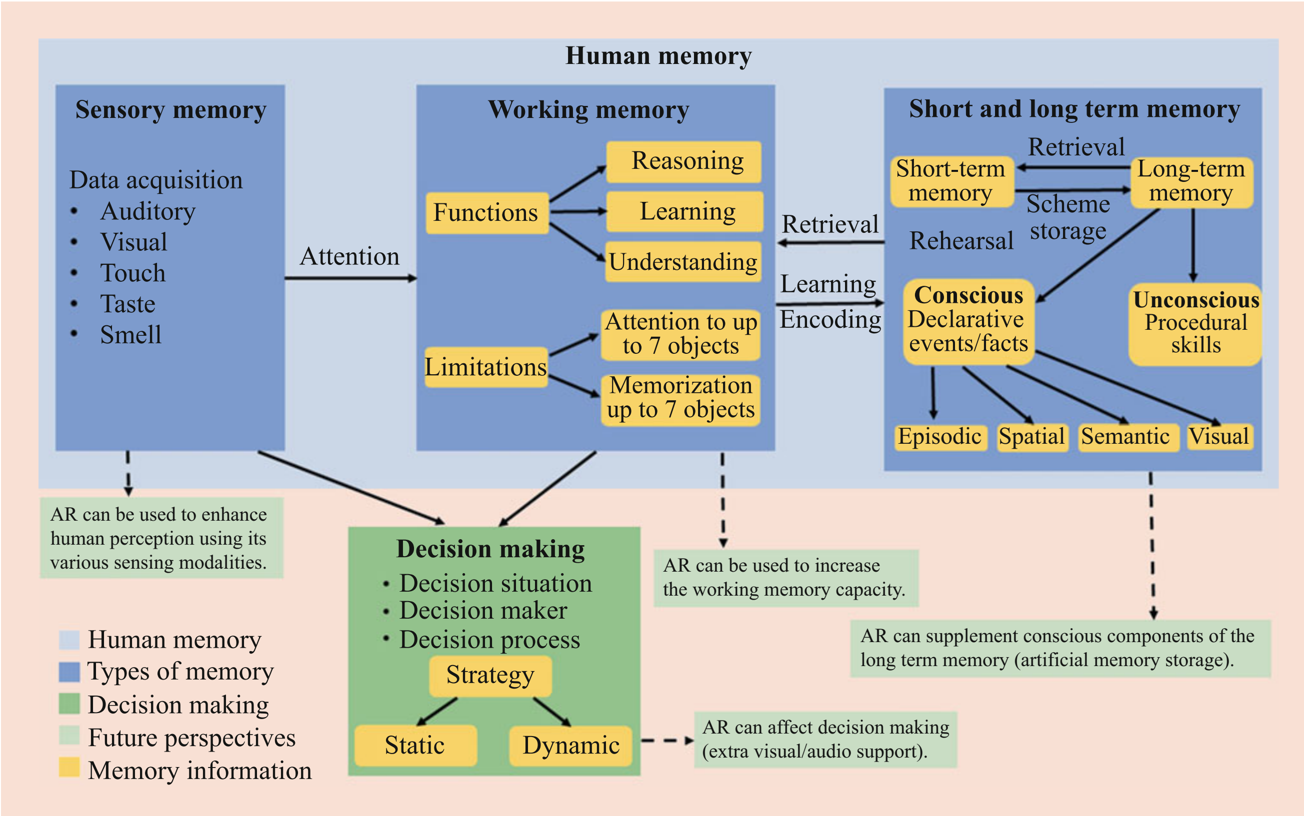


Fig. 31.3 AR-related research in assisting human memory and cognition

memory components. The conscious component is responsible for storing information of declarative events and facts. It can be further classified into episodic, spatial, semantic, and visual memories. While the unconscious component stores information on procedural skills.

In general, an individual's memory capacity varies from person to person. It is assumed that working memory has limited capacity, while long-term memory has unlimited capacity [31]. In the context of human memory analysis, ten potential sources were identified for the differences in human memory capacity [32]: rehearsal, grouping, chunking, retrieval strategies, item identification/ordering, capacity, susceptibility to interference, search rate, and the output buffer. These items were not yet studied in relation to an AR environment. Therefore, these constitute a wide range of open questions for future research in AR.

It is hypothesized that the human brain is constrained by the amount of information it can retain. For instance, an average human can pay attention to up to seven objects and memorize up to seven digits at once [33]. This is due to the natural differences in working memory. Further discussion on the limitations of the working memory can be found in the seminal paper of Cowan entitled "Magical Mystery Four" [34]. The author showed that due to the limitations of working memory, an average human subject can memorize up to four objects at once without experiencing cognitive overload. It was reported that properly designed instructions for a task can support learning and minimize working memory load [31].

Regarding the nature of individual differences within the working memory capacity (WMC), the case study was presented in [35]. The authors identified two determinants of WMC, which were related to the dynamic memory attention process happening in the primary memory component, and the memory searching process occurring in the secondary memory component of the working memory. According to the authors, the first difference arises due to one's ability to access the data stored in the primary memory. Whereas the second difference stems from one's ability to search information from data stored in the secondary memory.

WMC may affect not only one's ability to memorize but also another important cognitive process known as decision-making. Generally, decision-making has three constituent elements: decision situation, decision-maker, and decision process. An outcome of decision-making depends on the strategy selected by the decision-maker. Decision-making strategies can be grouped as static and dynamic. Static decision-making takes place when the environment changes and the human decision-making activities are independent of each other, or in other words, decisions made by human do not affect the environment and vice versa. Dynamic strategies depend on outcomes defined by the environment, and effects of previously made decisions [36]. Sensory memory and working memory both can have an impact on the decision-making of the human. Moreover, the potential of AR to affect dynamic decision-making is not well studied yet, thus can be explored by future researchers.

With regard to the sensory memory component of the general memory system, data acquisition depends on one's perception of audio, visual, and sensory information. Thus, via shaping the content and controlling the quality of provided information and the way it is delivered to the human, one can directly or indirectly shape the decision-making process itself. Since AR technology is capable to manipulate the data and the way it is delivered, it yields a new approach for more effective and seamless cognition of data and processes. Overall, present capabilities of AR technology can provide people with extra visual/audio support, and artificial memory storage capacity. Also, the study of AR-enhanced human memory augmentation might trigger new research on the exploration of the human decision-making process supported by AR. Specifically, case studies investigating the role of AR in dynamic decision-making can be a fruitful direction for future research.

31.3.2 AR Applications in Research of Human Memory and Cognition

In this subsection, we first look at the historical background of human intelligence augmentation concept. We also overview the first examples of memory augmentation systems in the literature. Then, we will focus on our primary research question, "What is the state-of-the-art of AR-enhanced assistive systems in the research of human memory and cognition?". We identified our keywords at the intersection of two key domains: "augmented reality" and "memory augmentation." In order to find corresponding literature, we applied the following advanced search strategy to the Scopus database: TITLE-ABS-KEY ("augmented reality" or "mixed reality") and TITLE-ABS-KEY ("memory assistant" or "memory augmentation," or "memory aid" or "memorization"). This search on 30 May 2021 returned 46 results. We first excluded review and survey papers. Then, by analyzing the papers, we identified the most relevant ones matching the scope of our work. We selected papers such that the discussed concepts and issues are not repeated. Overall, six papers are presented in Table 31.1, in which we describe unique features, methodology, hardware, and software of the developed wearable AR memory aid systems.

Historical Background

When introducing the futuristic concept "Man-Computer Symbiosis" in 1960, Licklider [37] envisioned that humans will easily cope with complex cognitive tasks thanks to joint decision-making with computers. Engelbart [38] from Stanford Research Institute's Augmentation Research Center focused on transforming human-technology interaction from a basic service relation to the level of collaborative work. Engelbart's team strived to enhance human intelligence via artificially increasing the cognitive capacity of the human mind. With time, these ideas matured into the concept of *Augmentationism* [39], which encompassed the philosophy that knowledge will not be stored in the human brain, but instead will stay within the supporting technology (e.g., artificial memory storage). Later, the concept of "Intimate Computing" [40] appeared in the literature as a novel model of computing paradigm in which computer accompanies its user all the time and can adjust its behavior according to the user's need, effectively transforming itself into a wearable personal assistant. Interestingly, while personal computers were supposed to provide users with access to devices' working context, intimate computing enhanced assistive systems were designed to provide access to the user's own working context.

First examples of human memory augmentation systems utilizing ideas of human episodic memory appeared in the literature two decades ago. Specifically, two broad groups of wearable digital assistants were developed around that time: personal digital assistants (PDAs) and wearable remembrance agents (RAs). Before the development of the PDAs and RAs, there were several works by researchers from the Rank Xerox Research Center, which discussed the role of technology in supporting personal memory such as video diary for memory recall [41], design of memory prosthesis [42], and activity-based information retrieval [43]. In 1994 researchers from the same team presented the "Forget-me-not" portable episodic memory aid system [40]. At the time of development, the application was in its early prototype stage, where a small portable device with a screen was used as its main platform. The application was able to collect the user data and represent it in the form of personal biography. The graphical user interface was utilized to convey the information to the user in the form of tiny symbols.

In 1997, researchers from the MIT Media Lab introduced a continuously running proactive memory aid named "Remembrance Agent" [44]. This system differed from the Forget-me-not in the way information presented to the user was selected. Specifically, RA did not create an autobiographical summary of actions during a period, but displayed notes that summarized events or activities only relevant to the current context of the user. When the context changed, the RA adapted the displayed notes accordingly. The built-in infrared receiver of RA could communicate with the indoor

Table 31.1 AR applications for human memory augmentation

Work	Application	Year	Methodology and unique features	Hardware and software	Target population
[47]	MOL for dementia patients	2015	MOL method of memorization, set of beacons associated with objects in AR	Android tablets, GPS, BLE (Bluetooth low energy) beacons	Dementia patients
[48]	Remotely controllable navigation system	2015	Constructive design method, light emitting diode indicators placed on the frame of eyeglasses	Eyeglasses frame, LED indicators, GPS tracker, camera, accelerometer, gyroscope, step detector sensors, and microcontroller	Alzheimer's dementia patients
[49]	External memory aid system ELEPHANT	2015	Digital flashcards of faces, places, objects, etc., dynamic reminders generated via machine learning tools based on detected information	Google Glass, "Glassware" software, GPS, Google Glass Mirror API, and Bluetooth	Patients with memory deficiencies
[50]	Memorization assistant NeverMind	2016	MOL, image placement on the locations on the defined route	iPhone app (user interface), Epson Move-rio BT-200 AR headset, and Unity 3D video game engine	General population
[51]	Cognitive assistant HoloLearn	2018	Helps to learn ADL patients with cognitive impairments	HoloLens 1, Unity, MRTK, and Visual Studio	ADL patients
[52]	External memory aid system EyeRemember	2018	Voice/video recording, picture taking, web searching, tracking of user emails, phone calls, text messages	Google Glass, smart watches, Bluetooth low energy (BLE) beacons	People with memory impairments

location beacons or markers for data collection and self-localization. In the same year, Wilson et al. [45] developed a portable paging system, "NeuroPage," supposed to help people with neurological impairments. The paging device accessed the patient's data from a communicating computer and sent reminders to the device's screen at the scheduled instances.

In order to retrieve information recorded by the wearable digital assistive devices (e.g., PDAs and RAs), the first thing to do was the process of indexing the information. Previously developed memory aids were designed mostly to make life easier for the computer, rather than for the user. For example, users were required to remember the filenames in order to browse the searching data. However, the developers of the PDAs and RAs exploited users' self-context as a valuable key for information indexing. Basically, these wearables were equipped with sensors that collected data of the users' surroundings and activity in order to create cues for recalling information later on. The idea of using physical context as a tool for information retrieval appeared in the literature long ago [46]. It resulted in various theories about episodic (autobiographical) memory. Within episodic memory, humans organize past memories into episodes, and can use characteristics of the episodes such as location, people involved, activities going on, and events that happened before or after in order to retrieve the searching information from the memory. In other words, these attributes are employed as cues for recalling the memory. Such phenomena of human memory made researchers contemplate the possibility of creating a prosthetic episodic memory device or memory prosthesis. Within this context, the digital devices were programmed to automatically structure recorded data in a way that human episodic memory naturally organizes the data. Thus, small cues associated with the event's context were supposed to be used to recall the forgotten details of the memory.

The symbols utilized in the graphical user interfaces of memory aid systems had to be easy to recognize and interpret by the users. Otherwise, they could contribute to the memory load. Another challenge that appeared was the difficulty of transforming raw data into human-recognizable episodes. For example, structuring the data describing activities recorded during the day supposed to be different from the data processing that described the activities that happened in the course of the whole year. Another issue was that the amount of text presented on the display of the memory aid systems could become messy and difficult for the user to follow. Moreover, sometimes, the displayed information could be irrelevant to the current context.

Overall, developers of memory aid systems aimed to utilize the capabilities of technology to enhance users' independence, employability, rehabilitation, and reduce their stress related to their memory problems. The health care system was projected to benefit from memory aids as well, due to the device's cost-effectiveness, reduction of caregiver effort, and in some cases medical drugs and treatment. However, developers were also cognizant of privacy and security issues that could become major concerns among users of memory aid systems. It was assumed that solutions for these issues will require a slate of developments both in technology and legislation [40].

AR-Enhanced Memory Aid Systems

Over the years, wearable AR appeared at the intersection of three research fields: AR, mobile/ubiquitous computing, and human ergonomics. An important role within this area of research is given to head-mounted AR displays. These see-through displays are responsible for overlaying the virtual content in the form of text, object, sound, image, video, or animation on top of the real scene. Within the context of assistive systems, AR has introduced a novel approach to address issues of people who need technological means to compensate their impairments. There are many cognitive impairments and diseases such as autism, dementia, Alzheimer's disease (AD), and various mild cognitive impairments. These impairments are followed by co-occurring deficits in language, speech, memory, social behavior, and motors skills of the patients, and can be overcome with the help of AI-enhanced AR-assistive systems. Several applications of AI for helping dementia patients were discussed by Sonntag [53] and included the following functionalities: speech dialogue and episodic memory retrieval, behavior tracking, and game-based rehabilitation. Overall, further development of technology can increase chances of cognitively impaired people for better rehabilitation, independence, happiness, and sense of self.

Cognitive disability (CD) is known as the term that covers a broad range of intellectual or cognitive deficits related with neurological disorders (e.g., autism), cognitive problems due to the brain injuries, and neurodegenerative diseases (e.g., dementia) [54]. There are several works in the literature that investigate AR-assisted applications designed to address issues of people with CD. One example is HoloLearn [51], which was developed to improve the autonomy of cognitively impaired people in ADLs. Application introduced to the user an AR environment, where several activities composed of dragging and dropping an object were designed. It was implemented on AR goggles, Microsoft HoloLens first generation, and presented the "Lay the table" and "Garbage collection" activities, during which holograms were overlaid on the physical environment and taught cognitively impaired patients perform routine tasks. In the "Lay the table" task, patients learned to lay real glasses, plates, cutlery, and bottles on the table. While in "Garbage collection" activity patients learned to collect different types of waste into a box. The level of difficulty, number, and type of virtual objects, as well as behavior of virtual assistant in AR were able to be adjusted, according to the needs of the patient. Such AR-enhanced application aims to develop independence of CD patients thanks to the presented training activities.

Another wearable AR assistant system for CD patients, specifically people with Alzheimer's dementia, was developed by Aruanno et al. [55]. Application was designed to train short-term and spatial memories, and thus slow down the mental decline of such patients. The work presented two activities for training short-term memory and one activity for training long-term memory. In the first activity, the subjects searched for a reference object hidden in eight boxes arranged in two rows displayed by AR goggles. During the second activity, patients were asked to find two similar objects in those eight boxes. The third task was similar to the second one but the boxes were placed in different locations in the room, and the subjects had to move around the room, memorize the location of the objects, and match them. The application was tested by elderly subjects with no cognitive impairments, and based on performed study authors assumed that such AR application can help to train memory and delay memory loss of AD patients. However, this argument should be supported with further tests with cognitively impaired subjects. During the implementation of such wearable assistant training systems in AR, designers faced several technical problems mostly while setting the two-way interaction and communication between the CD patients and wearable AR devices. Another challenge that developers faced was in need for personalization, since most CD cases are supposed to be unique and characterized by enormously different cognitive and motor skills, as well as different therapeutic and educational needs.

Wearable AR has the potential to improve the quality of life of both people with dementia and their close ones. Category of assistive technologies designed to help elderly during navigational tasks are in demand among senior citizens and can improve their quality of life. Within this context, wearable application developed for the elderly with mild and moderate dementia is presented in [48]. This application helps patients with memory loss to navigate while performing usual everyday tasks. The navigational signals are sent remotely by the caregiver. Such systems transform the one-to-one care-taking process into a one-to-many format, resulting in more efficient use of health care resources. In this application, video/image data of the user's sight, as well as data from the GPS, step counter, and gyroscope were sent to the caregiver. In order to set voice communication between the user and the remote unit, microphone and earphones were utilized. Even in remote navigational applications for elderly, a user-friendly interface and smart image processing and localization tools are required, making AI applicable in the area of AR-enhanced assistive technologies as well. This concept suffered from some of the technical shortcomings of the prototype, i.e., the navigation cues distracted the users. The authors noted the necessity of improved user interface and lightweight hardware components for the introduction of this technology to the market.

Another AR application developed for supporting AD patients, especially the one with apraxia and action disorganization syndrome is known as Therapy Lens [56]. The framework was created in Unity with HoloToolKit for Microsoft HoloLens. Tea making task was chosen by the authors for demonstration. The task has multiple subtasks that should be performed in an organized order that can be problematic for AD patients. Ten participants with mild cognitive impairments participated in the experiment. Control (natural tea making) and experimental (AR-supported tea making) conditions were assessed in the study. The duration of the trials, the successful task completion, and Mini Mental State Examination (MMSE) scores were used for quantitative evaluation, while qualitative evaluation was conducted in the form of semi-structured interviews. The time to complete a task successfully was higher with the Therapy Lens application compared to the control group. However, the acceptability of the application was high, with the exception that HoloLens was bulky and uncomfortable for the majority of participants. One limitation of the application was that the instructions did not have the necessary level of detail for some participants. For instance, placing the kettle to its base had to be cued to one of the participants for task completion. This highlights the necessity for the personalization of the application according to the needs of each patient. Participants with comorbidities such as partial loss of vision and hearing problems benefited from the multidimensional cues (audio, subtitle, and holograms) integrated into the application.

Many recent works that implemented memorization aid systems in wearable AR investigated features of human memory system and popular mnemonic techniques. Within human memory system, the short-term and spatial memories are utilized to memorize things via creating associations with the specific locations in a familiar environment. Associations are related with associative memory of the human that support learning and remember relations between items that are not directly related with each other. Method of loci (MOL) is known in the literature as the classic mnemonic technique that helps to memorize things by creating mental associations between to-be remembered items and the familiar locations in 3D space [57]. According to the definition, MOL has several components and requirements (see Fig. 31.4). First, a familiar place is selected and the specific travel route in the place is specified. Then, different locations (loci) along the travel route are defined and each location is assigned with an item to be remembered. After the walking activity along the travel route is completed, the subject can mentally visit the place, imagine walking along the route, and visualize the locations. At this stage, the subject should be able to retrieve from the memory items associated with the visualized locations (loci).

AR-enhanced memorization aid system known as Never-Mind is presented in [50]. Developed system utilizes spatial memory of the user during memorization and uses pairing technique between the spacial navigation and memory for easier memorization. More enhanced version of such memorization aid system is HoloMoL [58]; it uses both spatial and associative memory techniques. Both memory aid systems utilized visual textures (e.g., image or marker) in order to place corresponding information in the AR environment. Authors investigated in case studies if AR technologies can be efficiently integrated with general object memorization techniques and come up with novel AR-supported memorization assistant. Also, the work investigated if a HoloMoL-like application can be utilized to address social issues through the game playing (real-world locations similar to one in the MOL are utilized in the game design). The major limitation of the HoloMoL system preventing it from effective integration into the game environment was that information associated with the different locations were preregistered and depend on specific markers. Thus the information delivery lacked the dynamic features making HoloMoL unable to deliver information in correspondence with the current status of the game and the context. One more limitation was that one marker was able to deliver only one message, while different players in the game ideally should receive different information when looking at the same marker.

There are also works in the literature that implemented MOL technique in wearable AR to improve memory skills of people with mild cognitive impairments. One of such works presents a wearable AR assistant "EyeRemember" designed to help people with early dementia for memorizing relatives' names by utilizing associative learning, i.e., creating coupling between an object and the person to be remembered by a story [52]. Specifically, this application requires the storytelling from the patient about the person when the object is recognized. It is assumed that storytelling strategy is helpful for training the memory of people with dementia, enhance mutual recognition, and trigger positive emotions. The system facilitated recall of the names and relevant information about close contacts of the subjects. The developers utilized hand/finger gestures in combination with head movements, eye winking, and voice for providing input to the AR system. They also presented audio complemented visual output displayed on the HMD. Clinical tests of this memory aid system revealed a lack of knowledge about the HMD technology (e.g., Goggle Glass) among clinicians and caregivers. Therefore, in order to test EyeRemember among a wide range of patients with brain injury, the case studies should be first tested in a controlled environment within hospitals and rehabilitation centers.

Before technological advances, memory aid systems were mainly passive, requiring the user to check the memory assistant from time to time. Over the years, active strategies were implemented, such that memory aid systems become capable to send reminders to the users according to the detected context. One such AR-enhanced active memory aid system utilized AI to dynamically retrieve photographic memories with respect to the current context of the user's sight [49]. Recorded memories were automatically annotated with the information of location, date, time, image type (e.g., face,

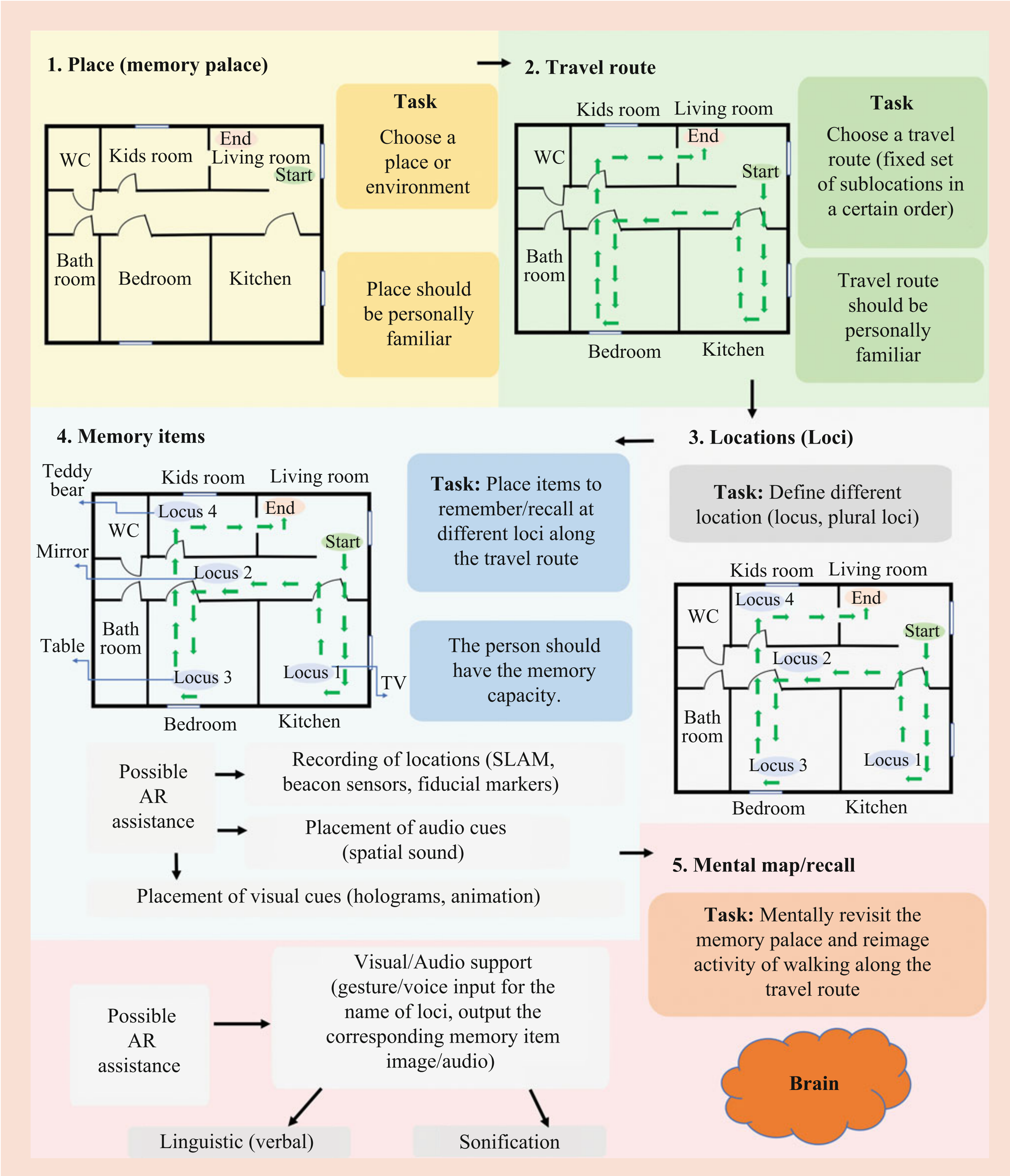


Fig. 31.4 Illustration of method of loci consisting of five steps

place, or object), and type of activity. AI algorithms were utilized to search the matching information from the database prompting the dynamic reminders. Reminders were in the form of digital flashcards, while each flashcard was characterized by the title, date, time, and a picture corresponding to a subject or a memory. For interaction with the assistant, users used hand gestures, specifically they were able to go over the flashcards and review the information offered by the system about some particular subject. Also, the system had voice recognition module that enabled voice-to-text conversion, when getting user inputs to assign titles for the flashcards or memories.

Several memory aid systems used mnemonic strategies that aim to encode information into the long-term memory. Such encoding involves chunking of the to-be remembered information into a picture format, resulting in decreased cognitive load and accelerated decision-making. This also leads to higher task performance. Multiple AR studies mentioned potential of AR to complement human cognition, thanks to integrated information access, error reduction, enhanced motivation, and concurrent training. Most of the assistive systems that implemented MOL strategy in AR were designed for indoor scenes and utilized small sensors (beacons) attached to the objects, which were detected by AR goggles via Bluetooth communication. Use of such sensors put constraints on the general use of cognitive assistants and make them case dependent. Some of the systems used marker-based object recognition which also had shortcomings in robustness and personalization. Based on the post-experimental studies of various AR applications, majority of works indicated that users prefer to communicate with the AR environment using voice control. Gestures required additional training, were not intuitive, and overwhelmed the users. Moreover, when users had to choose between voice and gesture control, they preferred voice control. Marker-based communication schemes in AR also had some limitations. Specifically, application became marker dependent, making it harder to personalize for the user.

It is worth mentioning that most of the recent AR-enhanced assistive applications in the domain of memory and cognition are still in the prototype stage. As a consequence, further studies and experimental investigations should be conducted both in AR and cognitive science. The spatial perception ability of wearable AR systems enables self-localization in the real environment, and can create the spatial map of the registered environment. This explains the popularity of MOL technique in AR-enhanced memorization aid systems since both utilize spatial features of the environment. Most works utilize popular AR headsets such as HoloLens and Goggle Glass in their applications, while the hologram initialization is achieved using marker-based techniques. This significantly impedes capabilities of AR, and results in static and predesigned AR environments. In more realistic AR-enhanced assistive applications, the AR environment should be dynamic and change its state according to the user's current context. Ideally, cognitive assistants must be able to acquire knowledge from the surrounding world and present that knowledge to the user via communication channels. In order to form the knowledge from the surrounding, any cognitive assistant should have functionalities such as object recognition, person identification, self-localization, and object tracking within the dynamic scene. Even though these technical issues still present a challenge, their solutions are introduced rapidly.

31.3.3 AR Applications in Cognitive Load Research

In this subsection, we will introduce the concept of cognitive load and review popular methods utilized for the evaluation of cognitive load. Then, we will focus on our primary research question, which is "What is the state-of-the-art of AR in research of cognitive load?". We identified our keywords at the intersection of two main areas: "augmented reality" and "cognitive load." In order to find corresponding literature, the following advanced search strategy was applied in Scopus: TITLE-ABS-KEY ("mixed reality" or "augmented reality") and TITLE-ABS-KEY ("cognitive load"). The search on 30 May 2021 returned 209 results. First, we excluded papers with the words "education," "learning," and "assembly" and ended up with 114 papers. From these, we selected original research papers and excluded review/survey papers. Overall, we selected five papers for the discussion and analysis presented in Table 31.3. Specifically, we selected papers such that the general concepts and study designs are not repeated, as we wanted to present the general trends in the research of AR and cognitive load.

Cognitive Load

Cognitive load is a multidimensional metric that measures a human's mental state during some activity or performance [59]. Cognitive load is associated with the internal activity of the brain during the information processing, and cannot be measured directly. However, perceived mental effort can be measured via subjective post-experiment questionnaires which are considered in the literature as an indicator of the cognitive load [60]. Cognitive load is related with quantities such as mental load, mental effort, and performance. The science dealing with cognitive load is known in the literature as the cognitive load theory (CLT). The main idea behind the science of cognitive load is that excessive load caused by complex learning processes can be removed, if properly designed instructions are provided [61]. According to CLT, there are three types of cognitive load: (1) intrinsic cognitive load, (2) extraneous cognitive load, and (3) germane cognitive load. We will describe these loads using an example (see Fig. 31.5). During studying a new subject, a student's cognitive load is distributed as follows: (1) intrinsic cognitive load represents load due to the inherent nature of the study material itself, (2) extraneous cognitive load depends on instructional design and caused by the way the material is presented (unnecessary cognitive load), and (3) germane cognitive load defines the effort necessary to form the scheme

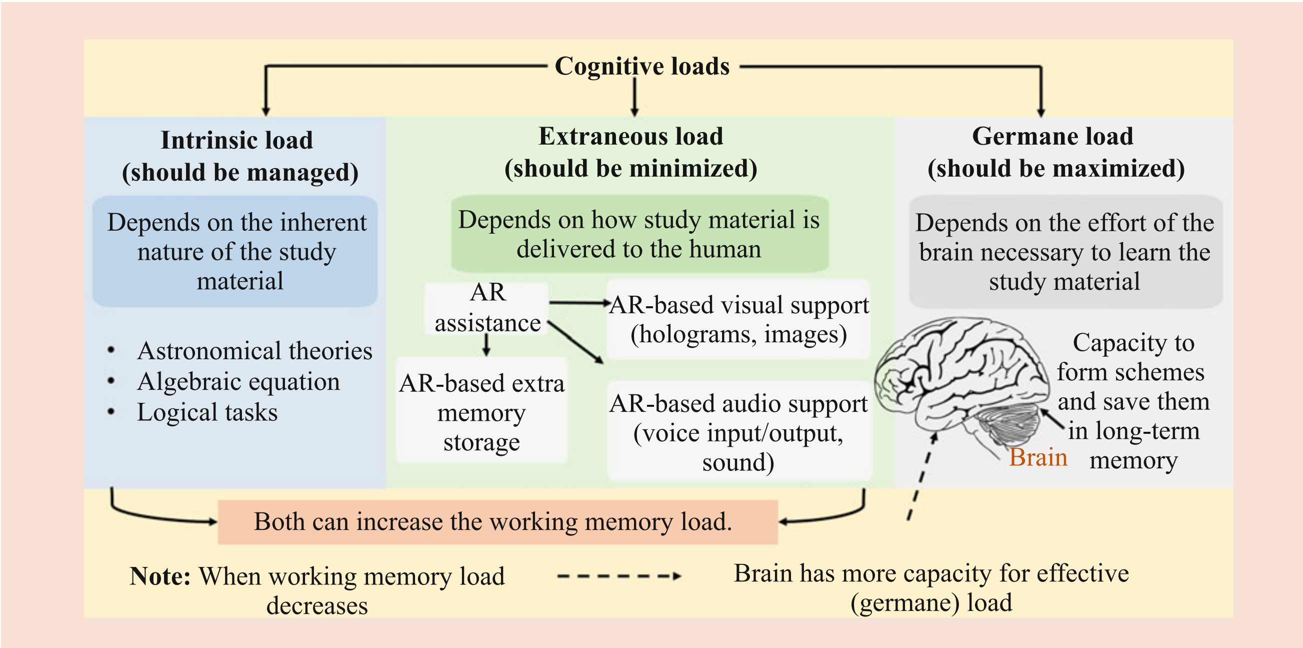


Fig. 31.5 Cognitive loads (study example)

(good cognitive load) [59]. The main objective of effective instructional design is to decrease the extraneous cognitive load and increase the germane cognitive load. Intrinsic and extraneous cognitive loads can increase working memory load [62]. All of these requires techniques for cognitive workload measurement, and still considered as the major challenge in the literature of cognitive load.

Wierwille et al. [63] divided mental workload measurement techniques into three groups: (1) subjective techniques, (2) task- and performance-based techniques, and (3) physiological techniques. The cognitive load measurement tests and their descriptions are summarized in Table 31.2.

Subjective techniques use a human's feedback on the experienced mental workload after the task is performed. Popular subjective techniques are: (1) Modified Cooper Harper (MCH) Scale [64], (2) Subjective Workload Assessment Technique [65], and (3) NASA Task Load Index (NASA TLX) [66]. For example, NASA TLX measures task-related mental/physical/temporal demands, frustration, effort, and performance to estimate experienced cognitive load.

Task- and performance-based measures compare the subject's performance first on the primary task alone and then on the primary task with a secondary task introduced [63]. Simple activities, such as detection of visual and auditory cues, are common secondary tasks. Basically, the goal of the secondary task is to distract the subject. Reaction time, accuracy, and error rate are the major performance estimation variables of these measures [67].

Physiological cognitive load estimation techniques use biological responses of humans to task demands. These responses include heart rate, brain activity (e.g., task-evoked brain potentials), and eye reactions (e.g., pupil size, blink rate, and saccadic motions) [67].

Psychometric tests are also employed to measure cognitive load. These tests depend neither on the material nor the instructions but consider intrinsic characteristics of one's mental state and cognition [68]. Therefore, psychometric tests can provide independent measures of cognitive capacity. For example, a paper-and-pencil test named Mental Rotations [69] can examine a subject's general spatial cognitive abilities. There are other psychometric tests in the literature, such as Rey Auditory Verbal Learning Test (RAVLT), Trail Making Test, Grooved Pegboard Test, Digit Span Test, Stroop Test, Letter–Number Replacement Test, and Short Cognitive Performance Test (SKT) [70].

Over the history of assistive systems, a number of studies utilized subjective methods during the evaluation of the cognitive load. However, there were also works that studied psychophysiological measurements during the assessment of the human's cognitive state. Within this context, in the early 2000s, several studies on human-machine interaction designed for military settings appeared in the literature [77, 78]. Specifically, Augmented Cognition program [77] was launched to empirically study to what extent the existing psychophysiological measures are capable of registering changes in human cognitive activity during real-time tasks. The program utilized 20 psychophysiological measures of cognitive states during execution of control tasks designed to evaluate perceptual, motor, spacial, auditory, verbal, memory, and decision-making cognitive activities of the user. In order

Table 31.2 Cognitive load measurements techniques

Method	Test name	Test description
Subjective	NASA Task Load Index (NASA TLX)	Questionnaire designed to evaluate a task or system via rating the resulting cognitive workload. Following items (scales) are investigated: mental/physical/temporal demands, frustration, effort, and performance [66]
	Modified Cooper Harper (MCH) Scale	Subjective measure developed to assess an aircraft's design and performance. Following items are investigated: aircraft controllability, workload experienced by the pilot, and attainable performance goals [71]
	Subjective Workload Assessment Technique (SWAT)	Card sorting pretask procedure is performed by the subject (cards differ in mental load) and then task (event) scoring procedure takes place. Following items are investigated: time load, mental effort load, and psychological stress load [72]
Task and performance based	Dual-Task Performance Test	Highly utilized in experimental psychology and multimedia studies. Subjects are introduced with primary and secondary (distracting) tasks during which the following items are investigated: reaction times, accuracy, and error rate
Physiological	Physiological measures of cognitive load	Following sensory formations are investigated: electroencephalography (EEG), heart rate, brain activity (e.g., task-evoked brain potentials), and eye reactions (e.g., pupillary size, blink rate, and saccadic motions)
Psychometric	Paper-and-pencil Mental Rotations Test	Chronometric study of mental imagery or spatial visualization using stimuli (2D images of 3D objects). Time limits are introduced and correct rotations are counted [69]
	Rey Auditory Verbal Learning Test	Neuropsychological assessment tool evaluating verbal-memory capacity of the subject. Audio recordings of a list of nouns are presented to the subject and the subject is requested to list memorized items after some time [73]
	Trail Making Test	Neuropsychological test designed to assess visual attention, task switching capability, and executive control. Subjects are asked to draw a line connecting consecutive numbers from 1 to 25. Time of completion and accuracy are shown in [74]
	Digit Span Test	Designed to assess the subject's immediate verbal recall (short-term memory), intentional capacity, and working memory. During the test, strings of digits (string length increases from 3 to 9) are read to the subjects. Then, the subjects are asked to list the digits in the correct order, forward or backward [75]
	Stroop Color and Word Test	Neuropsychological test utilized to assess one's ability to inhibit cognitive interference. During the test, subjects are asked to name the colors of words presented in black ink in one case and different ink in another case [76]

Table 31.3 AR applications in research of cognitive load measurement

	Application	Methodology and unique features	Hardware and software	Cognitive load measurement
[84]	AR-enhanced cognitive assistant	Mnemonic pictorial method to encode data in long-term memory via chunking the data into pictures	Optical see-through (OST) HMD (Juxtopia LLC), helmet, micron tracker, visualization toolkit (VTK)	NASA TLX subjective method
[85]	AR glasses in automotive industry	Age-related effects on cognitive load of AR glasses	QR code, Sony SmartEyeglass SED-E1	Hamilton depression and anxiety scales, EEG, and NASA TLX
[83]	AR-assisted physics laboratory courses	Spatial and temporal contiguity, split-attention effect	HoloLens 1, Unity, and Vuforia	Conceptual knowledge test (pre- and posttest), adapted version of the cognitive load scale (CLS) by [81]
[86], [87]	Shared gaze interaction in collaborative AR	Eye tracking, target identification, and virtual partner	HoloLens 1, Unity, PC, Pupil Labs eye tracker, MSI backpack computer	Seven-point Likert scale, questions from NASA TLX, System Usability Scale (SUS) questionnaire, and short interview
[79]	Warfighter problems with embodied assistant	Audio-only and embodied assistants	HoloLens 1, Unity, Vuforia markers, Adobe Fuse4, Vocalware, FaceGen, and Mixamo	NASA TLX, Temple Presence Inventory (TPI) questionnaire, and short interview

to manipulate the task load, primary task and secondary verbal-memory task were combined. In the primary task, different aspects of cognitive activity such as perception, motor activity, memory, attention, and decision-making were studied during a 15-min computer program in which a random number of aircraft (tracking objects) moved on the display. In the secondary task, the participants were required to listen to the audio messages about the current status of three systems in a ship, and answer multiple-choice questions when they were prompted by the captain via an audio message. By varying the numbers of aircraft (tracking objects) in a wave, their class (e.g., an aircraft classified as friend can be ignored, when classified as enemy needs to be processed properly, and undetermined ones need more processing from the operator's side), and the presence/absence of the verbal-memory task, experimenters were able to manipulate the task load which was considered an equivalent of cognitive activity and workload. The desktop software for the experimental studies were designed such that scenario events, user response times, and errors were recorded in real time. The response times were synchronized with the events registered by the external sensory devices such as brain activity, cortical blood flow, users' hand pressure, and changes in body posture in response to changes in task demands. The major challenge faced by the researchers was interpretation of the raw sensory data as meaningful cognitive state measurements.

AR in Cognitive Load Research

General trend in the literature of AR shows that many studies are designed to compare the cognitive load with and without AR support. Subjective methods were utilized to measure changes in attention, distraction, and cognitive load. While objective measures were employed to study the cognitive impact, i.e., the cognitive load caused by the AR technology on the users. Most frequently used cognitive load measurement techniques are the average of the area under the EEG graph (objective) and NASA-TLX scores (subjective). NASA-TLX scale was designed to assess the user's mental, physical, and temporal demand, performance, effort, and frustration spanning a range of values between very low and very high. Neurophysiological measurements such as EEG are utilized to describe different cognitive states of the user while performing different tasks. Specifically, EEG is a quantitative method that measures the electrical activity of neurons in the brain recorded from the scalp. EEG system can be portable and wireless. Also, it is cheaper in comparison to other systems such as functional magnetic resonance imaging (fMRI). Relative easiness of data collection and affordability makes EEG a popular objective method for the cognitive state measurement.

In the study presented by Demelo et al. [79], cognitive load of participants was measured, while they were solving problems in military settings with the help of a virtual embodied assistant in AR, voice assistant, and no assistant. The virtual embodied assistant with audio and video modes was implemented using AR goggles (Microsoft HoloLens). Thirty-four participants were instructed that they survived a plane crash in the desert which is 70 miles away from the nearest known habitation. They were tasked to prioritize between 15 survival items (jackknife, water, pistol, etc.). In this task, the participants were paired with an embodied virtual assistant, voice assistant, and no assistant. They followed the suggestions of the virtual assistants and received feedback from them. Several hypotheses were tested in this experiment: (H1a) Does the embodied assistant outperform the voice-only assistant?; (H1b) Do both assistants increase performance compared to no assistant?; (H2) Does the embodied assistant lead to lower cognitive load than the voice assistant?; and (H3) Does social presence and richness are perceived higher with the embodied assistant than the voice assistant?. The performance of participants with assistants surpassed no assistance (confirming H1b); nevertheless, higher performance with the embodied assistant compared to the audio assistant was not observed (not supporting H1a). The cognitive load of participants was evaluated with NASA TLX. The cognitive load during problem-solving with the embodied assistant was less than the audio assistant (confirming H2). The social richness with an embodied assistant was perceived higher than voice assistant (confirming H3). This study reduced cognitive burden using natural conversation for human-machine interaction. One limitation of the study was the simple emotional expressions (subtle and strong smiling) of the embodied assistant.

In the literature, there are lots of works that overview application of AR in assembly engineering. For instance, effects of AR glasses on the cognitive loads of the employees in automotive industry was discussed in [80]. Statistical analysis of the data in this paper showed that when performing assembly tasks with AR support the cognitive load of the employees decreased according to both objective and subjective tests. During the data collection assembly engineers were performing diffusion task that involves cognitive activities such as decision-making, memory processes, and visual stimuli. NASA-TLX answers indicated that completing the task with AR glasses help required less effort and temporal demand from the participants.

There are also studies in the literature that investigate whether psychometric instruments can differentiate types of cognitive load such as intrinsic, extraneous, and germane. For example, researchers in [81] investigated effect of instructional design on different types of cognitive load in the example of learning the statistical concept, Bayes' theorem. Four learning scenarios were investigated: problem-problem, problem-example, example-problem, and example-example, and the results indicated that the two learning scenarios which are example-problem and example-example resulted in reduced germane cognitive load in comparison with the other two learning scenarios. Such studies support an assumption that germane and intrinsic cognitive loads can be differentiated from the extraneous cognitive load with the help of psychometric instruments, e.g., subjective questionnaires, principal factor analysis, and analysis of variance.

Several AR applications aiming reductions of the cognitive load try to deal with split-attention effect, and this way

31

improve the task performance and mitigate the extraneous cognitive load. Split-attention effect takes place when the learning components are split up within space and time, making learners spend additional mental activity for integrating these components, and construct a coherent mental model [82]. Such mental activity is not desirable since it consumes resources necessary for the actual learning process. AR allows to combine real and virtual learning components, utilize features of spatial and temporal perception in order to avoid split-attention effect, and thus foster learning and reduce resulted extraneous cognitive processing. Within these contexts, there are several works that explore effects of AR-based instructional scenarios on the learning outcomes and extraneous cognitive load.

In [83], the efficacy of AR-assisted framework for university physics laboratory courses was studied. The experiment was held as part of a university course for bachelor students. For comparing classical and AR-assisted workflows of physics experiments, the students were divided into control and intervention groups. The control group conducted experiments in a traditional way while the AR-integrated method was used by the intervention group. The workflow consisted of the preparation, pretest, test, posttest, and analysis steps. Performance and cognitive load were evaluated as dependent variables during the experiment. The cognitive load scale (CLS) questionnaire by Leppink et al. [81] was used for assessing cognitive load. The questionnaire was administered in the form of an HTML-based survey on tablets. Two hypotheses were tested: “Does AR lead to higher learning gain?” and “Does AR reduce the extraneous cognitive load?”. After comparing the students’ pre- and posttest scores, a higher learning gain was not observed for the AR-integrated method. However, the AR framework reduced extraneous cognitive load supporting the second hypothesis. The researchers also found that the simultaneous visualization of multiple data does not affect intrinsic cognitive load. This AR-assisted teaching method has the advantage of reduced extraneous processing in physics experiments. There are also a few shortcomings of this study: complete randomization was not achieved, the sample group was unbalanced in the pretest, and high pre- and posttest scores diminished the ability to measure differences in conceptual knowledge. Overall, the work studied if AR-enhanced visualization system added to the traditional learning scenario benefits the students’ conceptual knowledge and reduces the irrelevant cognitive load. The work demonstrated that AR-based visualization of the live data integrated to the traditional laboratory learning environment resulted in reduced extraneous cognitive load. The idea behind this visualization system is related to the term of multiple external representations (MERs) which stand for different forms in which a scientific concept is expressed, demonstrated, depicted, or communicated. In addition to words, graphs, algebraic expressions, pictures, diagrams, and tables, AR introduces animations, videos, and holograms into that list of MERs. Capabilities of AR might also fit into the area of multimedia instructions that are composed of visual and auditory information, and aim to reduce the extraneous cognitive load resulting from the instructional design or the way instructions are delivered to the learner.

A primary challenge in the utilization of wearable AR in daily life is the way information is presented and how inputs are registered from users such that cognitive load is mitigated. Based on these necessities, an adaptive AR system was introduced capable of changing its content based on the recordings of the user’s physiological and cognitive state. Application of such an adaptive AR visualization system in medical surgery was considered in [84]. In the medical surgery domain, the operator might be cognitively overloaded during image-guided neurosurgical operations. During this process, surgeon repeatedly looks to the computer monitor and back to the patients causing ergonomic issues. These might increase the operation time, fatigue, and potential errors. Thus several works in literature address such problems with the help of AR-enhanced digital glasses integrated with precise tracking systems. The system was utilized for studying the operator’s cognition and memory during a surgery supported by AR-based navigational instructions. A head-mounted system was utilized for head tracking, registration, and display. The major drawback of the system was the low update rate of 20 Hz, which restricted the head movements of the operator and also caused visualization errors.

Wearable AR technology can understand user’s point of interest and direction of attention utilizing the gaze, which is conceptualized as a directional ray from the camera system of AR headset to the 3D environment. There are studies which assess the potential of gaze cues to benefit users’ interpersonal spatial communication, improve task performance, and reduce experienced cognitive load. It was determined that four types of errors (accuracy, precision, latency, and dropout) affect the users’ performance in gaze-based assistive systems [86].

In [86], the researchers measured the cognitive load and other performance degrading factors for shared gaze interactions. Pupil Labs eye trackers and HoloLens AR glasses were utilized along with complementary computing devices such as desktop and backpack computers. Subjects were tasked to work with the virtual human partner, a police officer. The task aimed to identify a virtual human who can cause threats in a crowd of virtual people. The virtual partner helped subjects to identify a potential threat with gaze information. Gaze information was transmitted in the form of a line and cursor which connected the human head position to the target object. Subjects had to say a floating number on top of the threat object as an answer. The accuracy, precision, latency, and frame dropout errors were measured across multiple trials. Tasks were performed in three target distances ranging from 3 to 7.2 m. The subjective perception of subjects was measured using a seven-point Likert scale. Questions from NASA TLX and System Usability Scale (SUS) questionnaires and questions about gaze behavior realism were included in the questionnaire to assess the cognitive load, performance of the framework, and user preferences. This work provided valuable information on the error sources in gaze interactions with state-of-the-art AR devices. The authors reported that the limited field of view of HoloLens increased response time and error rate at close and medium distances.

Within the paradigm of Industry 4.0, it is assumed that AR can help employees to increase productivity and quality of their work. In addition, with the help of current capabilities of AI people can process large amounts of data in superhuman speeds and perceive the environment in wavebands, e.g., infrared, extending beyond human ability. However, the premise of AI can only be fully leveraged if successful collaboration with the human users is established. This collaboration and interaction between human and AI is a challenge. Thus, various works studied the workload during the collaborative task execution. In this context, AR has the potential to support and further enhance the collaboration with AI with the help of virtual agents (objects, animations, spatial sounds, or even virtual embodied assistants) placed within the real environment.

Within the context of diagnosing cognitive impairments, AI can be combined with psychophysiological measurement tools for the automatic diagnosis of the wide range of disorders related with dementia and memory decline. In order to ease the data collection process and analysis, the capabilities of AR can be utilized. There are already works in the literature that study AI in the diagnosis of dementia. For example, in [88], biomarker-based prognostic AI models were constructed for detecting people in the presymptomatic stages of dementia. For increasing the generalizability and robustness of the models, data from clinics in Europe and the USA were collected. Clinical evaluation involved Mini-Mental State Exam (MMSE), Clinical Dementia Rating (CDR), Wechsler Memory Scale, and Memory Box Score. A hundred subjects performed Altoida iADL tests at home. Altoida iADL test was a mobile test app that assigned AR interactive tasks to participants. It consisted of different levels of complexity and AR tasks were inspired from instrumental activities of daily living (Altoida iADL tasks) related to spatial memory, prospective memory, executive function, and psychomotor processing speed. The app collected Neuro Motor Index (NMI) outcome of subjects during the task sequence. NMI was defined as screen touch frequency, hands' micromovements, correct replies, response time, walking cycle and pace, navigation route, and others. Digital biomarkers (NMI) obtained by the test app were used as a baseline dataset for ML analytics. Biomarkers including cerebrospinal fluid, brain magnetic resonance imaging (MRI), and apolipoprotein E genotype were collected from participants to form a baseline when Altoida iADL tasks were performed. Digital biomarker-based prognostic models allowed to discriminate people who were at risk to develop dementia within 3 years from healthy people with the precision of 86%. Good generalizability and robustness were achieved in predicting the progression of dementia.

Even though AR technology has several promising advantages, it has certain limitations as well. In addition to the computational, field of view, calibration, and battery limitations, AR might obscure environmental cues and this can distract, disrupt, and cognitively overload the user. Such issues are termed in the literature as the attention tunneling, i.e., AR diverts user's attention from the important cues of the physical environment. To overcome such limitation, AR-enhanced assistive systems should be designed such that technology would be aware of the user's current cognitive state and adapt AR system accordingly. In addition, in order to improve user's experience and performance, AR-enhanced systems should constantly monitor user's psychophysiological signals, and personalize AR experience.

31.3.4 AR Applications for Restoring Perception

In this subsection, we will cover perception-related cognitive impairments. Then, we will focus on our primary research question, "What is the state-of-the-art of AR in the research of perception diseases?". We identified our keywords at the intersection of the following areas: "augmented reality" and "perception diseases." In order to find the corresponding literature, the following advanced search strategy was applied in the Scopus database: TITLE-ABS-KEY ("augmented reality" or "mixed reality") and TITLE-ABS-KEY ("visual agnosia" or "topographic agnosia," or "achromatopsia" or "alexia" or "agraphia" or "aphasia" or "auditory agnosia" or "prosopagnosia" or "color blindness" or "hearing impairment" or "vision impairment"). The search performed on 30 May 2021 returned 209 results. From the results, we prefiltered original research papers and excluded review/survey papers. Overall, ten works were included for the analysis presented in Table 31.5, in which the application for the impairment, methodology, unique features, and utilized hardware/software systems are presented. We selected papers such that the concepts, techniques, and general features are not repeated. This way, we were able to create an overall picture of AR in the research of perception-related cognitive impairments.

Perception-Related Cognitive Impairments

In cognitive science and psychology, perception is known as the process that makes a human aware of the surrounding world and understand the information about it via acquired

Table 31.4 Different types of perception impairments

Perception	Impairment	Description
Vision	Achromatopsia	Color perception deficit or color blindness
	Prosopagnosia	Face perception deficit or face blindness
	Environmental agnosia	Selective loss in the ability to recognize familiar places
	Prosopamnesia	Inability to learn the identity of new faces
	Topographic agnosia	Inability to orient in 3D space to locations, objects, and landmarks
	Metamorphopsias	Deficits characterized by size or object distortions
	Macropsia/Micropsia	Objects seem larger/smaller than their actual size
	Associative object agnosia	Inability to assign meaning to an object while perception of the object remains intact
	Color and object anomia	Deficits dealing with the loss of color or object naming
	Optic aphasia	Inability to name an object via sight when the ability to name an object when touched or described is preserved
	Pure alexia	Inability to read words while the recognition of letters and writing is preserved
	Surface dyslexia	Condition when the link between word forms and semantics is lost. Reading is performed based on phonology
	Alexia associated with speech/language disorders	Alexia with aphasia or alexia with agraphia
Hearing	Amusia	Inability to recognize music
	Auditory agnosia	Inability to interpret/recognize sounds while the pure tone hearing ability is preserved
	Perceptual form of auditory agnosia	Inability to interpret detected sounds
	Developmental auditory agnosia	Congenital form of amusia characterized by tone/tune deafness
	Nonverbal agnosia or environmental sound agnosia	Selective difficulty in recognizing nonverbal environmental sounds
	Verbal agnosia	Inability to recognize/detect spoken language while the abilities for speech production, reading, and spelling are normal
	Phonagnosia	Inability to recognize and discriminate voices
	Central auditory disorders related with central speech/language disorders	For example, hearing impairments associated with aphasia and agraphia

sensory information [89]. The sensory information is acquired by a human through hearing, seeing, smelling, touching, and tasting via sensory organs that can recognize/identify different types of stimuli. Cognitive impairments associated with deficits in visual and auditory perception are summarized in Table 31.4.

A taxonomy of visual perception disorders is presented in [90]. According to researchers, there is a class of vision perception disorders known as deficits, when the patients suffer from selective loss of one specific visual function. These include achromatopsia (color perception deficit or color blindness), prosopagnosia (face perception deficit or face blindness), and environmental agnosia (selective loss in the ability to recognize familiar places). Prosopagnosia is often accompanied by prosopamnesia, i.e., the patients are unable to learn the identity of new faces [91]. There is also topographic agnosia (patients cannot orient themselves to locations, objects, and landmarks) [92]. With regard to color blindness, there is a condition known as color vision deficiency (CVD) characterized by disability to distinguish different color hues.

Visual deficits characterized by size or object distortions are grouped as metamorphopsias. For example, an object seeming larger/smaller than the actual size is termed macropsia/micropsia. Associative object agnosia stands for the loss of the ability to assign meaning to an object while the perception of the object remains intact. There are also specific deficits dealing with the loss in color or object naming known as color and object anomia, correspondingly. The case when the ability to name an object via touch or description is preserved but the ability to name it via the sight is lost is known as the optic aphasia. Pure alexia stands for the deficit in reading words while the recognition of letters and writing is preserved. Surface dyslexia is described as the condition when the link between word forms and semantics are lost. Such patients read based on phonology and make mistakes while reading the irregular words. There are also types of alexia associated with speech and language disorders such as aphasia and agraphia. Aphasia patients lose the ability to comprehend and formulate the language. In extreme cases, the patients might have impaired speaking, reading, and writing. Agraphia patients have deficits in writing or spelling while patients not necessarily have problems in reading and speaking.

Neumann and Stephens define types of hearing disorders in [93]. According to the authors, there is a class of hearing impairments known as central auditory disorders related to demonstrable brain lesions. This group contains the following hearing impairments: amusia (inability to recognize music), auditory agnosia (inability to interpret/recognize sounds while the pure tone hearing ability is preserved), perceptual form of auditory agnosia (inability to interpret detected sounds), developmental auditory agnosia (congenital form of amusia characterized by tone/tune deafness), nonverbal agnosia or environmental sound agnosia (selective difficulty in recognizing nonverbal environmental sounds), and verbal agnosia (inability to recognize spoken language while the abilities for speech production, reading, and spelling are normal). There is also a special type of auditory agnosia known as phonagnosia when patients have difficulty with voice recognition and discrimination [94].

AR-Enhanced Applications in Research of Perception Impairments

In this part of the work, we will investigate AR approaches for addressing perception-related cognitive impairments, such as vision and hearing. In general, visual modality of AR is popular in the literature of assistive systems designed to compensate perception-related impairments. In addition, AR has also other capabilities such as auditory, haptic, olfactory, and gustatory information overlay. Most wearable AR-enhanced assistive systems were designed for the blind, and were expected to support a wide range of functions in visual cognition including navigation, obstacle avoidance, route finding, object recognition, reading of signage, understanding of the surrounding real-world context or scene, and operations with spatial memories (e.g., formation and recall). There are works in the literature that present AR-enhanced assistive systems designed to compensate the deficiency to recognize and differentiate colors. Also, there are several works that cover auditory AR-based navigation for the visually impaired. AR might introduce new ways during the design of assistive systems utilizing techniques of sensory substitution in order to address needs of cognitively impaired people. Moreover, its integration with AI might result in even more advanced and effective assistive systems. The role of AI within the research of sensory substitution of the visually impaired was discussed in the literature three decades ago. One of the pioneers of the sensory substitution method, Collins [95] stated that AI is a powerful tool that can process large amounts of visual data, compensate missing functions of the eye, and perform most of the preprocessing operations with the data usually performed by the brain. However, at that times, these goals were mostly out of reach. Fortunately, advances in computer vision, AI, computing, and miniaturization of technology transformed these goals into a fast-approaching reality. Overall, the invention of AR and wearable sensors created new ways to expand human senses and interaction with the world. At the same time, this caused developers of the AR-based sensory augmentation systems to deal with the challenge of establishing the interaction between sensory augmentation system and the user.

In the literature, there are two major approaches to restore vision for the blind. First is the prosthetic approach, in which raw image data is sent directly to the brain of the user with the help of external encoding devices. These devices send electrical signals to the operating cells within the retina of the eye or neurons of the visual cortex part. The major limitation of such systems is the lack of bandwidth. The alternative approach is to restore vision at the cognitive level. Within that approach, external sensors collect the video/image data, extract the necessary information, and convey data to the user via not impaired sensory modalities, in most cases through hearing or touch. One example of such a user interface implemented with AR is presented in [96]. Interestingly, an AR environment was designed such that each object within the 3D space in real world was assigned with a voice, and was capable to communicate with the user upon request. In general, there are several ways how developers of wearable AR systems can establish communication between user and device in an AR environment, e.g., hand gestures, voice control, and gaze control. Also, markers can be used for interaction. For example, each marker can be assigned with a specific action, and that action is triggered whenever the marker is recognized via the camera of the wearable AR [97]. A popular implementation of sensory substitution for the blind on wearable AR is CARA cognitive augmented reality assistant [96]. The system conveyed messages to the user through natural language voice commands instead of the traditional scene sonification. Three tasks were performed by blind and blindfolded subjects. In the first task, subjects were tasked to localize objects randomly placed in azimuth 1 m away. Subject pressed the clicker to hear a sound spatialized on target. Afterward, the correct angle confirmation audio was played. The second task consisted of scan and recall phases. The subjects had to remember five objects placed 2 m away spaced 30° apart in azimuth from each other. The scan phase lasted for 60 s and subjects had to recall the direction of each object when asked in random order. In the recall phase, subjects were asked to turn to the direction of a specific object. In the third task, a virtual chair was placed 2 m away at a random angle. The subjects could press the clicker to hear the call emanating from the chair. The task was completed successfully when the subject entered 0.5 m vicinity of the virtual chair. Subjects performed several trials by going to the object and back to their places. The system was compared to the traditional techniques used by blind people. Specifically, the subjects were also asked to find a real chair using a cane. The chair was placed in the same space. Subjects had to touch the chair to complete the trial. With the help of CARA, subjects found the chair faster than using the classical walking aid. In the fourth task, a 36 m guide path was designed by the experimenters from the first-floor lobby to the second-floor office of a building. It included nine waypoints in the prescanned environment. A virtual guide was calling "Follow me" every 2 s. When the subject arrived within 1.2 m to the target, the task was completed with the voice confirmation "You have arrived."

Another example of a sensory substitution system is the vOICe [98]; the system was designed to render video to soundscapes and vice versa. The most recent version of the vOICe application can run on a variety of platforms (AR smart glasses such as Vuzix M4000, M400, Blade M300XL, M300, and M100, VISION-800, Google glass, Android smartphones, and Windows desktop computer with USB camera glasses). vOICe was already used by blind people for different tasks. For instance, the visually impaired Indian photographer Pranav Lal used vOICe technology to take photos. The creators of the CARA system compared its performance with the vOICe system in a set of tasks. CARA was running on the Microsoft HoloLens while the vOICe setup consisted of an HTC Vive VR headset and a desktop computer. CARA outperformed vOICe in this benchmark suite. The main advantage of CARA was its capability to work standalone without an edge device for computation. This enabled it to do real-time automatic wayfinding using navigation tools such as NavMesh "Baking" and "UnityEngine.AI." One more subtle benefit of CARA was the seamless integration of voice controls in addition to the clicker. Voice control included keyword recognition and dictation modules. Keyword recognition detected known voice commands, whereas words said by the subjects were converted to text by the dictation module. The voice control was implemented using HoloToolkit scripts. The downside of the framework was the loss of spatial tracking in narrow areas with white walls since they have fewer visual features.

The natural functionality of sight makes people understand "What is where?" by looking to the environment. Cognitive assistant systems such as CARA or vOICe compensate for this seeing process of the eye and the visual system. Such a cognitive system might be modified to convey information that is not seen by the sight, thus not only replacing vision for the visually impaired but also enhancing the overall perception of the scene by the sighted user. Also, it is worth mentioning that rather than solving the technical problems, developers of CARA focused more on the user interface. Thus, several user studies were conducted in order to find the most effective and useful user interface for the blind. Within the context of spatial memory, authors investigated if the object voices can help to construct the mental image of the scene for the blind. Experimental studies confirmed that active exploration of object voices stimulate mental map formation, scene understanding, and furthermore support recall and orientation within the real-world environment.

Within the context of assistive systems for the blind, CamIO system helps blind and visually impaired people by providing audio labels to the places which the user touches with a stylus in real time [99]. First, the 3D locations of interest (hotspots) were labeled with corresponding audio labels using a customized stylus. The wooden stylus had barcode patterns on four sides to facilitate pose estimation. OpenCV library was utilized for detecting barcodes whereas the object and stylus poses were estimated with Perspective-n-Point algorithm. The efficacy of the CamIO app was assessed by six blind participants in tasks with objects such as the 3D model of a plant cell, a tactile street map, and a microwave oven. The objects were also attached with barcodes for estimating their poses. Specifically, the study investigated the localization of audio annotations with a stylus and smartphone for blind people which requires cheaper hardware than hand tracking. In the task with the plant cell, CamIO running on iPhone 8 was attached to a table and the participants had to point to hotspots and record them. In the tactile street map task, the participants were asked to add audio labels to street intersections when experimenters directed them verbally. All participants were able to record hotspots correctly. In the last task of labeling the buttons on the microwave oven, more than half of the participants performed the task successfully. In semi-structured interviews, the participants were asked for their suggested improvements, what they liked and disliked, and whether their concept of ideal audio cueing changed after using CamIO. Most of the participants expressed that they would prefer to use their own fingers for labeling instead of the stylus.

Virtually Enhanced Senses (VES) [100] is another MR application for assisting the blind and visually impaired. The system consists of one server device (a smartphone or a laptop) and client devices (smartphone and earphone attached to a headset and another smartphone for hand tracking). A virtual replica of a physical environment was built and the participants were provided with specific sound stimuli by the system during the experiment. The system provides wall tracking, acoustic compass, and geometry-based virtual acoustic space (GbVAS) cues. Wall tracking helps participants to walk parallel to a wall by warning the users when they move away from the wall. Acoustic compass is a virtual sound from a static place that allows participants to perceive obstacles and projected sounds from other objects along the way. GbVAS algorithm has three modes: Projected Virtual Acoustic Space (PVAS), Aural Virtual Acoustic Space (AVAS), and Flashlight. PVAS gives acoustic feedback that allows placing objects and their types in the users' mental map. AVAS creates a virtual sound source for each object which collides with the virtual aura of the user. Flashlight mode gives a synthetic speech response with the name of the object when the user taps the screen. Sixteen normal-sighted subjects equally divided into two groups participated in the experiments. The first and second groups used only PVAS and PVAS with AVAS, respectively. Before the experiment, participants were trained with the system for 6 min. The experiment consisted of three tests. Participants had to explore an unknown environment in the first two tests and find the route to the exit in the third one. In the first test, the PVAS group performed worse than the second group (PVAS with AVAS). This was attributed to the underestimation of the field of view by PVAS alone. Nevertheless, the performance of the second group decreased in the second test. Participants perceived AVAS as heading cues and were lured toward virtual walls. In the third task, more than a third of the participants in both groups reached the exit successfully. The researchers pointed out that the advantage of the GbVAS was providing perceptual and higher cognitive factors. The lack of visually impaired participants might be noted as a weakness of this study.

Several works in the area of assistive systems focus on face recognition algorithms instead of the common user studies. Marquez et al. [101] presented a face recognition algorithm capable of identifying faces with different expressions. The software was integrated into a prototype assistive system, composed of a headband with a camera, computation unit, battery, and the speakers that expressed the name of the person registered by the camera. AR-enhanced applications with built-in face recognition also can be considered as external memory aids. For example, two core techniques, dynamic face recognition and adaptive information retrieval, were utilized for a social interaction assistant [102]. For the information retrieval process, the data were first recorded and labeled using the current context of the user and scene recognition technique, which is highly utilized in prosthetic memory aid systems. The major limitation of this assistant system is that scene recognition is performed among the pretrained scenes within a specific building. Thus, the system might fail to register context within an unconstrained environment. This again indicates the necessity of utilizing AI in assistive systems designed for augmenting human senses and assisting people with perception deficits.

Within the field of assistive computer vision, there are works that utilize wearable AR glasses to help CVD patients. For example, Dai et al. [103] patented an assistive system that displays instructions on the AR headset that assists drivers with color vision deficiency. The system shows information about what vehicular action should be taken in response to the detected color signals on the road. This AR system detects the location of the registered color signals, specifically the brake signal from other vehicles, using GPS and depth cameras. The core technical feature of such AR assistants is automatic pixel-level color classification algorithms [104]. A well-known computer vision technique utilized in assistive systems for the colorblind or people with color vision deficiency (CVD) is pixel mapping from a camera to a display system. A work implementing this pixel mapping algorithm on the images acquired by wearable AR camera system was presented in [105] and called the ChromaGlasses. These computational glasses as vision aids utilized Moverio BT-300 OST HMD glasses. Three studies were performed: feasibility using user perspective camera and Chromaglasses, and evaluation of usability and memory load. The state-of-the-art OST HMDs such as Moverio BT-300 and HoloLens have their own off-axis cameras. These cameras cannot show the view from the user's perspective. The authors integrated a third-party PointGrey BlackFly camera to OST HMD for showing the image from the user's perspective. The cameras were placed near the eyes using beamsplitters virtually. The cameras were calibrated to reduce the error between users' eyes and displays. In the first study, tests with Ishihara plates were performed by participants using user-perspective cameras instead of Chromaglasses. The limitation of these cameras was the occlusion of the direct view to the test plates. Different correction methods were used: RGBShift, LMSShift, Edges, RGBShiftAdjusted, and LMSShiftAdjusted. Nineteen participants with red-green color blindness participated in the experiment. Two hypotheses were tested: Would these correction methods improve participant's ability (H1) and would participants feel more confident when recognizing the correct content (H2)? Both hypotheses were confirmed in the experiments. Particularly, the participants felt more confident with default color shifts (RGBShift and LMSShift). In the second study, participants donned Chromaglasses directly. Nineteen color-blind participants performed Ishihara plate tests. Most of the participants were red-green blinds including one total color-blind participant. Nearly all participants performed the second task with custom color shifts (RGBShiftAdjusted and LMSShiftAdjusted) successfully with a high confidence rate. The third task aimed to test the usability and cognitive load of Chromaglasses compared to off-axis Google Glass–based approach. NASA TLX was used for measuring cognitive load and SUS for assessing usability. The participants using Chromaglasess exerted less cognitive load, and the usability of both devices was found to be similar. The strength of this work is the evaluation of the usability and cognitive load of novel color correction methods using user-perspective cameras. In [106], an image-level color classification framework was developed for assisting CVD patients. Parametric fuzzy sets were utilized for classifying pixel to color levels. Google Glass application was designed to improve the color classification of users by highlighting certain colors. The experiment was conducted to assess the efficacy of the system. Two groups of participants performed in the experiment. The control group consisted of four students without CVD, whereas four students with CVD were in the test group. Ishihara tests were performed by participants. The test group was assisted with the Google Glass application while the control group was not. The average response time of CVD individuals was 3.6 s compared to 1.39 of non-CVD individuals. All users in the test group responded correctly with the assistance of smart glasses. The study presented the algorithm and setup which could improve the color-blinded people. The usability study and workload assessment of ChromoGlasses indicated lowered cognitive load of CVD patients even though the system was bulky and need further improvements for everyday usage.

An AI-based warning system was developed using AR for people with peripheral vision loss [110]. Peripheral vision loss (PVL) is the absence of side vision, while central vision remains fine. A severe case of PVL is also known as tunnel vision. AI was used for head motion detection and object detection. AR was leveraged for projecting the target hazard in the central vision region within the screen of Moverio smart glasses. Moreover, AR can seamlessly supplement human cognition and provide context-aware assistance via multimodal perceptual cues (e.g., animation, graphics, text, video, voice, and tactile feedback using digital gloves). In the area of context- and location-aware digital assistants, Wiener et al. [111] presented the "wayfinding" concept for human pathfinding and spatial problem-solving. Also, Jacobson [112] discussed issues of cognitive mapping and spatial learning during the design of an auditory "hypermap system" for visually impaired people. There is a trilateral nature in interactions between environment, cognitive agent, and resulting human cartographic mental map [113]. In this context, Li et al. [114] developed a dynamic Bayesian network–based framework for indoor navigation. This system interfaced with the user as an audio assistant generating verbal navigational instructions. In [108], a virtual speaking head for the Slovak language was developed for students with hearing impairments. Deaf people usually use lip reading for understanding people with normal hearing. Speech visemes are the visual representation of speech sounds. The system leverages visemes to teach new words to students with hearing disabilities. Unity and Lipsynctool development engines were utilized for implementing speech phonemes and visemes, virtual human head, text-to-speech (TTS), speech-to-text (STT), and text components of the application. The feasibility and usability of the system were not tested with human subjects.

Augmenting human senses might improve perception and make it appear as an extension of natural vision. The main challenge of such perception augmentation systems is that control of the augmented vision should not result in more effort than the control of the natural vision. Before the integration of such systems in everyday life, researchers still need to solve issues of how to design corresponding interfaces and controls, ensuring that mental effort and cognitive loads are mitigated. For example, developers of the Virtual Ubiquitous Microscope [109] studied two types of AR-enhanced visualization (full-screen view and windowed view which is a small display area on the screen), and three types of controls (e.g., hand gestures, physical controller or clicker, and voice commands). These design elements were evaluated during experiments, and the following performance measures were recorded: task completion time, training time, and error rate. In order to measure the resulting cognitive load NASA-TLX questionnaire was used, and short semi-structured interviews were conducted in order to find the personal preference of the participants. The study indicated that system control via clicker has the fastest rate in zooming and results in lower cognitive load, as well as users prefer the windowed view for the visualization (Table 31.5).

AR combined with multimodal sensors and AI introduce the concept of "digital objects" that could acoustically augment the objects of interest. In addition, current technology can support attention-driven signal enhancement, noise suppression, and context awareness in assistive systems for people with hearing loss. An AR-enhanced hearing aid system was presented in [115], where authors discuss the potential of AR to improve user's speech understanding in noisy environments. This is accomplished thanks to present capabilities of sensory technology and AI which was used for automatic understanding of listener's intent, speaker identification, noise reduction, and signal enhancement. According to authors, AR-enhanced hearing aids should have properties such as environment understanding and adaptability based on the current context and performance evaluation. This requires the device to be aware of: (1) the user's location and what the user does at that location; (2) device's own position, orientation, and velocity; and (3) position of the sound sources and characteristics of the noise in the occupying space. Also, through AR glasses utilizing text-to-speech AI algorithms, information can be delivered to the user in noisy environments via sight in the form of captions presented on the wearables' screen. Berger et al. [116] proposed speech recognition system for deaf people. The system was connected to Google Cloud Speech-to-Text and ML Kit Face Detection services. The architecture comprised the server, mediator (Android phone), and client (Google Glass). The solution projected the speech in the form of text to the user via Google Glass utilizing neural networks which run on a mediator device. Neural networks evaluate the ambient sound for speech recognition. Visual lip detection was performed for recognizing the speaker. In addition, AR can present new ways to study problems of people with hearing deficits in a more comfortable way.

During the implementation of assistive systems capable of complementing functions of sight such as face recognition and color identification developers utilized wearable AR glasses for visualization and mobile phones or other

Table 31.5 AR applications for restoring perception

Work	Impairment	Application	Methodology and unique features	Hardware and software
[102]	Prosopagnosia (face blindness)	Socioglass (AR enhances social interactions assistant with build-in face recognition)	Information about person's identity and corresponding data are provided through phone's screen, external memory	Goggle Glass paired with a mobile phone, digital database of faces and corresponding data, OpenCV library
[96]	Blindness	Cognitive AR assistant (CARA) for the blind	Spatial sound-based navigation in AR, obstacle avoidance, spatial memorization, image-to-sound audio support	HoloLens 1 AR goggles, Unity 5.6.1f1, HoloToolkit-Unity- v1.5.5.0
[105]	Color Vision Deficiency (CVD)	Chromaglasses (AR-enhanced system compensating color blindness)	Pixel-precise mapping from camera into the user's view, detection of critical colors and shifting them on a per pixel base in AR display	AR goggles Epson Move-rio BT-300, half-silvered mirrors, PointGrey Black-fly cameras, Qt, OpenGL, GLSL shaders, OpenCV
[107]	Prosopagnosia (Face blindness)	FaceReminder (AR-enhanced framework for learning faces and assigning names)	Deep network-based learning of faces, showing the names of the person on the see-through display	VUZIX Blade Smart Glass, smart phone (utilized for computation), MTCNN facial detection technique
[101]	Prosopagnosia (face blindness)	AR assistant for learning, storing, and retrieval of identities of people met during the day	Face identification under different facial expressions, face recognition–based hybrid model of alpha-beta associative memories	Plastic head-mounted structure integrated with a camera, mini PC Jetson TK1, face recognition algorithm
[108]	Hearing impairments	Virtual Slovak speech speaking head (assistant) in AR	Smart education and smart e-learning	Unity 3D, Lipsynctool, converter text to speech, speech to text, generator of computer speech
[103]	Achromatopsia (color blindness)	AR-enhanced vehicular assistant for color-blind drivers that help to understand/interpret color signals on the road (e.g., traffic lights)	Colors recognition of light signals and assigning specific vehicular actions corresponding to the light signals presented as the word instructions on the AR display	AR system composed of onboard vehicle computer, electronic control unit, AR glove, GPS, wireless messaging protocol, and AR goggles
[99]	Blindness	CamIO (smartphone application for audio labeling of 3D spaces and objects)	Handheld stylus is used to point into the 3D space (hotspot) and record audio label correspondingly	Smartphone (iPhone), paper barcode, styluses, Google Cardboard, OpenCV, Perspective-n-Point algorithm
[100]	Blindness	Virtually Enhanced Senses (VES)	Testing of multisensory human-machine interfaces for blind, uses various types of virtual acoustic spaces	Server tablet or laptop, 2 smartphones, earphone (client), Unity, Resonance Studio, Vuforia, ARCore
[109]	Virtual Ubiquitous Microscope (VUM)	Incorporation of microscopy to AR. Study of user control techniques (e.g., gestures, physical controller, and voice commands) and display techniques	HoloLens 1, Unity, HoloLens clicker, Vuforia-based object recognition	Task completion time (TCT), training time, error rate, task work load, NASA TLX questionnaire, and short semi-structured interview

external devices for computation. However, most applications developed to assist blind in navigation and environment understanding utilized mobile phones as camera and server computation unit to perform calculations. In order to understand the environment and user's position, the majority of applications were using data from GPS, external sensors, and image markers. For example, the assistant system for the blind in [99] used image markers assigned to 3D objects in order to assign audio labels to objects and using these labels afterward for environment understanding. In order to register markers, a mobile phone was carried by the user, and interaction with the objects was set with the help of two types of sticks (styluses) designed to be recognized by the same mobile phone camera. User studies indicated that people prefer not to carry the mobile phone and not to use different types of sticks (styluses) in order to interact with the AR-enhanced environment, instead preferring hand gestures and touch. It is assumed that people with hearing deficits experience more cognitive load within environments like restaurants where complex noise is present. Increased mental effort for compensating the hearing loss consumes the limited resources of the working memory resulting in memory problems as well. The assistive systems for the colorblind need user-centered calibration for computing the eye-display relationships. The physiological differences of the eye and head geometry for the different users constituted the major problem for this approach. Such systems utilized the color detection and processing algorithms that shift the colors that the user had difficulty recognizing toward colors that are distinguishable for the user. In general, assistive systems that acquire data from the camera are expected to deal with challenges caused by the changes in lighting conditions and the environment. Assistive systems designed to help people with hearing impairments also need to deal with environmental noise.

To sum up, there are comparatively not many works in the literature utilizing AR for cognitive impairments due to perception, especially in the domain of hearing impairments. Thanks to the sensing and computational capabilities of AR devices, research in this area might generate significant results to restore the abilities of a substantial number of patients.

31.3.5 Issues and Challenges in the Application of AR for Cognitive Impairments

In this subsection, we will summarize issues and challenges in the research of AR-enhanced assistive systems, cognitive impairments, and memory aid systems.

Various strategies such as spatial memory, memory games (exercises), mnemonics, and external memory aids were applied to compensate for memory disorders. However, people using external memory aids face substantial difficulties since they need to remember how to use these memory assistants. Common aspirations of user interface designers include ease of learning and use, as well as interface reliability and trustworthiness. These have a special significance during the design of human memory support systems. For example, during the design of Forget-me-not, researchers had to ensure that it is easier to remember how to use the system, rather than remembering the forgotten fact. The issues related to the human-technology interaction should be solved as well. Specifically, the AR device for the specific application should be selected carefully. Various studies on AR-enhanced systems indicated that spatial AR might help to increase performance and reduce cognitive load, while the limited field of view and low resolution of AR HMDs might result in the additional extraneous (undesirable) cognitive load.

In addition, during the implementation of the memory aid systems, developers need to come up with smart solutions for processing the sensed data into the episodes (the so-called assigning the labels or cues to the acquired data). This is a possible area where AI algorithms can be applied, including users' localization, automatic object/person/place/activity recognition from the image data, as well as voice recognition and natural language processing from the speech data, and general sound recognition from the audio data. Specifically, AI can implement automatic face recognition in order to save a person's name during a conversation. GPS technology can be used to automatically localize users within the AR scene, and indoor sensors can be utilized to collect the user information, where GPS is not available. This way, memory aid systems might use the full potential of physical context in order to label and match data accordingly. Overall, in most wearable AR applications for cognitively impaired, object recognition was an important technical issue.

Also, AR can be utilized on the early diagnosis of cognitive diseases, such as dementia and early memory decline. One of the challenges in the research area of cognitive impairments is the difficulty of diagnosing these disorders and related issues with the evaluation methods. Many quantitative and qualitative tests should be conducted before an individual can be finally diagnosed with one of these cognitive impairments. In addition, currently, most studies of AR-enhanced systems utilize subjective measures to assess the cognitive load and user experience. Tests utilized for assessing the cognitive state and cognitive workload load entail tedious manual work for the experimenter considering that the tests should be performed several times for large groups of people. AR goggles can be utilized to simplify and automate this process. For instance, subjects can receive visuospatial inputs through AR display, followed by instructions to perform the tasks utilized in various tests. Then, the AR device can detect the task completion and register the corresponding times. We believe that the integration of AR into the clinical study of cognitive load can make data collection, evaluation, and analysis simpler and more intuitive.

Currently, mobile applications process data received from smartphone sensors, cloud, and GPS to come up with suggestions for the user. However, we prophesize that AR-enhanced systems can go beyond just assisting the impaired, but even restore and supplement their abilities using artificial memory, vision, and other senses. However, it is worth mentioning that some of the technical limitations of AR (e.g., the narrow field of vision of head-mounted displays and limited computation power) still need to be overcome for the widespread use of AR in assistive technologies for cognitive impairments. Also, we should keep in mind that AR technologies might give rise to security and privacy issues due to the systems' capability of memory augmentation, enhanced perception, and decision-making in everyday life.

Another problem for AR is malware, i.e., the applications might be exploited by malicious software easily, potentially aggravating privacy and security issues. Malicious applications can confuse AR users displaying false information about the physical world [117]. For example, in AR systems, applications might frequently shuffle between the frames and increase the user cognitive load. In AR applications deployed on devices such as AR contact lenses, the user might have difficulty in switching from the AR environment to the real one. Within the automobile industry, there are cases when AR technology is integrated into the windshields.

For instance, Jaguar Land Rovers project virtual holographic cues to drivers on the windshields. In this case, malicious applications might negatively affect the safety of the drivers that rely on the AR content while driving. In general, before AR-enhanced solutions become widespread in everyday life, the issues of malicious applications should be addressed first. Since AR depends on computations on the cloud and browsers, the developers should pay attention to computer browser attacks such as phishing and clickjacking. Popular defensive approaches used in mobile apps can be employed for AR as well. Specifically, AR-enhanced system might request access permission from the user to process certain user-dependent data or use microphone, camera, and installed software. Also, before the introduction of AR applications to the public, they should go through app store review processes. Also, AR-enhanced applications utilize external sensors for data collection in order to monitor the psychophysiological state of the users in real time. This also might cause privacy and security issues .

31.4 Conclusions

In this chapter, we provided an overview of current trends in AR technology and computing paradigms to introduce the capabilities of AR-enhanced assistive systems to the research community. Also, we briefly covered the basic mechanisms underlying human memory system, mental workload, and perception. We discussed in detail state of the art of AR in research of memory augmentation systems, cognitive load, and perception decline–related cognitive impairments.

We expect a multitude of assistive systems and cognitive tools to be developed for diagnosing and restoring cognitive impairments using AR supported by AI and novel computing paradigms. For instance, a single AR device can provide task-related visuospatial data to the human user, detect if task is completed, then record the task variables (e.g., performance times), and synchronize data recorded from the other external sensing devices as well.

Even though AR has a vast application area, we focused our attention on cognitive impairments which cause deficits in memory and perception, including declines in recognition of objects, people, places, as well as verbal and nonverbal audio signals. We were surprised by the scarcity of research works that address audio-related cognitive disorders using AR. However, AR-enhanced assistive systems were extensively developed for addressing people in need for extra memory, as well as visual, navigational, and training support. Restoration of cognitive abilities due to advances in wearable devices, sensor technology, computing paradigms, and AI frameworks (e.g., object recognition, face identification, and natural language processing) remains an attractive open research problem.

Thanks to its rapid development, AR technology can act as an invaluable tool for measuring cognitive state and experienced workload, as well as for compensating memory and perception declines of cognitively impaired patients. Specifically, perception, data recording, and analysis features of AR might simplify the work of the researchers involved in the development of new approaches in assistive systems. AR capabilities can also be used to enhance perception and memory of healthy people as well. However, limitations of AR technology such as limited field of vision, low computational power, short battery life, and potential to obscure environmental cues need to be taken into account while developing AR-enhanced assistive applications.

References

1. World Bank Group. https://www.worldbank.org/en/topic/mental-health. Accessed 22 Jan 2021
2. Patel, V., Chisholm, D., Parikh, R., Charlson, F.J., Degenhardt, L., Dua, T., Ferrari, A.J., Hyman, S., Laxminarayan, R., Levin, C.: Addressing the burden of mental, neurological, and substance use disorders: key messages from disease control priorities. Lancet. **387**(10028), 1672–1685 (2016)
3. Lancioni, G.E., Sigafoos, J., O'Reilly, M.F., Singh, N.N.: Assistive Technology: Interventions for Individuals with Severe/Profound and Multiple Disabilities. Springer Science & Business Media (2012)
4. Ong, S.K., Shen, Y., Zhang, J., Nee, A.Y.C.: Augmented reality in assistive technology and rehabilitation engineering. In: Furht, B. (ed.) Handbook of Augmented Reality, pp. 603–630. Springer, New York (2011)
5. Leo, M., Medioni, G., Trivedi, M., Kanade, T., Farinella, G.M.: Computer vision for assistive technologies. Comput. Vis. Image Underst. **154**, 1–15 (2017)
6. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. Environ. **6**(4), 355–385 (1997)
7. Sutherland, I.E.: The ultimate display. In: Proceedings of the International Federation of Information Processing Congress (1F1P), pp. 506–508. Macmillan and Co, London (1965)
8. Wade, N.J.: Charles Wheatstone (1802–1875). SAGE, London (2002)
9. Makhataeva, Z., Varol, H.A.: Augmented reality for robotics: a review. Robotics. **9**(2), 21 (2020)
10. NASA. https://www.nasa.gov/feature/ames/augmented-reality-air-traffic-management. Accessed 22 Jan 2021
11. Xiao, C., Lifeng, Z.: Implementation of mobile augmented reality based on Vuforia and Rawajali. In: Proceedings of the 1FFF International Conference on Software Engineering and Service Science, pp. 912–915. IEEE, Beijing (2014)
12. Hübner, P., Weinmann, M., Wursthorn, S.: Marker-based localization of the Microsoft Hololens in building models. Int. Arch. Photogramm. Remote Sens. Spatial Informat. Sci. **42**(1), 195 (2018)
13. Rekimoto, J.: Matrix: a realtime object identification and registration method for augmented reality. In: Proceedings of the IEEE Asia Pacific Computer Human Interaction, pp. 63–68. IEEE, Shonan Village Center (1998)
14. Lin, C.-H., Chung, Y., Chou, B.-Y., Chen, H.-Y., Tsai, C.-Y.: A novel campus navigation app with augmented reality and deep learning. In: Proceedings of the IEEE International Conference on

Applied System Invention (ICASI), pp. 1075–1077. IEEE, Chiba (2018)
15. Ren, P., Qiao, X., Huang, Y., Liu, L., Dustdar, S., Chen, J.: Edge-assisted distributed DNN collaborative computing approach for mobile Web augmented reality in 5G networks. IEEE Netw. **34**(2), 254–261 (2020). https://doi.org/10.1109/MNET.011.1900305
16. Marner, M.R., Smith, R.T., Walsh, J.A., Thomas, B.H.: Spatial user interfaces for large-scale projector-based augmented reality. IEEE Comput. Graph. Appl. **34**(6), 74–82 (2014)
17. Bimber, O., Raskar, R.: Spatial Augmented Reality: Merging Real and Virtual Worlds. CRC Press (2005)
18. Erickson, A., Kim, K., Bruder, G., Welch, G.F.: Exploring the limitations of environment lighting on optical see-through head-mounted displays. In: Proceedings of the Symposium on Spatial User Interaction. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3385959.3418445
19. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 4th edn. Pearson (2020)
20. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science. **349**(6245), 255–260 (2015). https://doi.org/10.1126/science.aaa8415
21. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature. **521**(7553), 436–444 (2015)
22. Pauly, O., Diotte, B., Fallavollita, P., Weidert, S., Euler, E., Navab, N.: Machine learning-based augmented reality for improved surgical scene understanding. Comput. Med. Imaging Graph. **41**, 55–60 (2015)
23. Ichihashi, K., Fujinami, K.: Estimating visibility of annotations for view management in spatial augmented reality based on machine-learning techniques. Sensors. **19**(4), 939 (2019)
24. Sutanto, R.E., Pribadi, L., Lee, S.: 3D integral imaging based augmented reality with deep learning implemented by faster R-CNN. In: Proceedings of the International Conference on Mobile and Wireless Technology, pp. 241–247. Springer, Singapore (2017)
25. Hoppenstedt, B., Kammerer, K., Reichert, M., Spiliopoulou, M., Pryss, R.: Convolutional neural networks for image recognition in mixed reality using voice command labeling. In: Proceedings of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics, pp. 63–70. Springer, Santa Maria al Bagno (2019)
26. Huynh, B., Orlosky, J., Hollerer, T.: Semantic labeling and object registration for augmented reality language learning. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 986–987. IEEE, Osaka (2019)
27. Lebeck, K., Cuervo, E., Philipose, M.: Collaborative Acceleration for Mixed Reality MSR-TR-2018-20. Microsoft (2018)
28. Ali, G., Le, H.-Q., Kim, J., Hwang, S.-W., Hwang, J.-I.: Design of seamless multi-modal interaction framework for intelligent virtual agents in wearable mixed reality environment. In: Proceedings of the International Conference on Computer Animation and Social Agents, pp. 47–52. Association for Computing Machinery (ACM) (2019)
29. Huang, B.-R., Lin, C.H., Lee, C.-H.: Mobile augmented reality based on cloud computing. In: Proceedings of the Anti-counterfeiting, Security, and Identification, pp. 1–5. IEEE, Taipei (2012)
30. Baddeley, A.D.: Human Memory: Theory and Practice. Psychology Press (1997)
31. Sweller, J., Van Merrienboer, J.J., Paas, F.G.: Cognitive architecture and instructional design. Educ. Psychol. Rev. **10**(3), 251–296 (1998)
32. Dempster, F.N.: Memory span: sources of individual and developmental differences. Psychol. Bull. **89**(1), 63 (1981)
33. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. **63**(2), 81 (1956)
34. Cowan, N.: The magical mystery four: how is working memory capacity limited, and why? Curr. Dir. Psychol. Sci. **19**(1), 51–57 (2010)
35. Unsworth, N., Engle, R.W.: The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. Psychol Rev. **114**(1), 104 (2007)
36. Wang, Y., Ruhe, G.: The cognitive process of decision making. Int. J. Cognit. Informat. Nat. Intell. **1**(2), 73–85 (2007)
37. Licklider, J.C.: Man-computer symbiosis. IRE Trans. Human Factors Electron. **1**, 4–11 (1960)
38. Engelbart, D.C., English, W.K.: A research center for augmenting human intellect. In: Proceedings of the Fall Joint Computer Conference, pp. 395–410. Association for Computing Machinery (ACM), New York (1968)
39. Skagestad, P.: Thinking with machines: intelligence augmentation, evolutionary epistemology, and semiotic. J. Soc. Evol. Syst. **16**(2), 157–180 (1993)
40. Lamming, M., Flynn, M.: Forget-me-not: intimate computing in support of human memory. In: Proceedings of the International Symposium on Next Generation Human Interface, vol. 4. Institute for Personalized Information Environment, Meguro Gajoen (1994)
41. Eldridge, M., Lamming, M., Flynn, M.: Does a video diary help recall? In: People and Computers, p. 257. Cambridge University Press (1992)
42. Lamming, M., Brown, P., Carter, K., Eldridge, M., Flynn, M., Louie, G., Robinson, P., Sellen, A.: The design of a human memory prosthesis. Comput. J. **37**(3), 153–163 (1994)
43. Lamming, M.G., Newman, W.M.: Activity-based information retrieval: technology in support of personal memory. Proc. TETP Congr. **14**, 68–81 (1992)
44. Rhodes, B.J.: The wearable remembrance agent: a system for augmented memory. Pers. Technol. **1**(4), 218–224 (1997)
45. Wilson, B.A., Evans, J.J., Emslie, H., Malinek, V.: Evaluation of NeuroPage: a new memory aid. J. Neurol. Neurosurg. Psychiatry. **63**(1), 113–115 (1997)
46. Tulving, E.: Elements of Episodic Memory. Clarendon Press (1983)
47. Morel, A., Bormans, K., Rombouts, K.: Memory palaces to improve quality of life in dementia. In: Proceedings of the Conference on Raising Awareness for the Societal and Environmental Role of Engineering and Training Engineers for Participatory Design (Engineering4Society), pp. 80–84. IEEE, Leuven (2015)
48. Firouzian, A., Asghar, Z., Tervonen, J., Pulli, P., Yamamoto, G.: Conceptual design and implementation of indicator-based Smart Glasses: a navigational device for remote assistance of senior citizens suffering from memory loss. In: Proceedings of the International Symposium on Medical Information and Communication Technology (TSMTCT), pp. 153–156. IEEE (2015). https://doi.org/10.1109/ISMICT.2015.7107518
49. Way, T., Bemiller, A., Mysari, R., Reimers, C.: Using Google Glass and machine learning to assist people with memory deficiencies. In: Proceeding of the International Conference on Artificial Intelligence (ICAI), p. 571. CSREA Press, Las Vegas (2015)
50. Rosello, O., Exposito, M., Maes, P.: NeverMind: using augmented reality for memorization. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 215–216. ACM (2016)
51. Garzotto, F., Torelli, E., Vona, F., Aruanno, B.: HoloLearn: learning through mixed reality for people with cognitive disability. In: Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), pp. 189–190. IEEE, Taichung (2018). https://doi.org/10.1109/AIVR.2018.00042

52. Wallace, T., Morris, J.T.: Development and testing of EyeRemember: a memory aid app for wearables for people with brain injury. In: Miesenberger, K., Kouroupetroglou, G. (eds.) Computers Helping People with Special Needs, pp. 493–500. Springer International Publishing, Cham (2018)
53. Sonntag, D.: Kognit: intelligent cognitive enhancement technology by cognitive models and mixed reality for dementia patients. In: Proceedings of the AAAI Fall Symposium, pp. 47–53 (2015)
54. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 5th edn. American Psychiatric Association (2013)
55. Aruanno, B., Garzotto, F., Rodriguez, M.C.: HoloLens-based mixed reality experiences for subjects with Alzheimer's disease. In: Proceedings of the Biannual Conference on Italian SIGCHI Chapter. Association for Computing Machinery, New York (2017). https://doi.org/10.1145/3125571.3125589
56. Rohrbach, N., Guide, P., Armstrong, A.R., Hartig, L., Abdelrazeq, A., Schroder, S., Neuse, J., Grimmer, T., Diehl-Schmid, J., Hermsdorfer, J.: An augmented reality approach for ADL support in Alzheimer's disease: a crossover trial. J. Neuroeng. Rehabil. **16**(1), 1–11 (2019)
57. Lea, G.: Chronometric analysis of the method of loci. J. Exp. Psychol. Human Percept. Perform. **1**(2), 95 (1975)
58. Yamada, Y., Irie, K., Gushima, K., Ishizawa, F., Sada, M.A., Nakajima, T.: HoloMoL: human memory augmentation with mixed-reality technologies. In: Proceedings of the International Academic Mindtrek Conference, pp. 235–238. Association for Computing Machinery (ACM) (2017)
59. Paas, F.G., Van Merriënboer, J.J.: Instructional control of cognitive load in the training of complex cognitive tasks. Educ. Psychol. Rev. **6**(4), 351–371 (1994)
60. Paas, F.G., Van Merriënboer, J.J., Adam, J.J.: Measurement of cognitive load in instructional research. Percept. Mot. Skills. **79**(1), 419–430 (1994)
61. Van Merrienboer, J.J., Sweller, J.: Cognitive load theory and complex learning: recent developments and future directions. Educ. Psychol. Rev. **17**(2), 147–177 (2005)
62. Engle, R.W.: Role of working-memory capacity in cognitive control. Curr. Anthropol. **51**(S1), S17–S26 (2010)
63. Wierwille, W.W., Eggemeier, F.T.: Recommendations for mental workload measurement in a test and evaluation environment. Hum. Factors. **35**(2), 263–281 (1993)
64. Wierwille, W.W., Casali, J.G.: A validated rating scale for global mental workload measurement applications. In: Proceedings of the Human Factors Society Annual Meeting, vol. 27, pp. 129–133. SAGE, Los Angeles (1983)
65. Reid, G.B., Eggemeier, F.T., Shingledecker, C.A.: Subjective Workload Assessment Technique. Air Force Flight Test Center Edwards AFB CA (1982)
66. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, Advances in Psychology, vol. 52, pp. 139–183. North-Holland (1988)
67. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.: Cognitive load measurement as a means to advance cognitive load theory. Educ. Psychol. **38**(1), 63–71 (2003)
68. Hammond, S.: Using psychometric tests. Res. Methods Psychol. **3**, 182–209 (2006)
69. Vandenberg, S.G., Kuse, A.R.: Mental rotations, a group test of three-dimensional spatial visualization. Percept. Mot. Skills. **47**(2), 599–604 (1978)
70. Hanning, C.: Postoperative cognitive dysfunction. Br. J. Anaesth. **95**(1), 82–87 (2005)
71. Cummings, M., Myers, K., Scott, S.D.: Modified Cooper Harper evaluation tool for unmanned vehicle displays. In: Proceedings of the UVS Canada: Conference on Unmanned Vehicle Systems Canada. Unmanned Systems Canada, Montebello (2006)
72. Luximon, A., Goonetilleke, R.S.: Simplified subjective workload assessment technique. Ergonomics. **44**(3), 229–243 (2001)
73. Bean, J.: Rey auditory verbal learning test, Rey AVLT. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds.) Encyclopedia of Clinical Neuropsychology, pp. 2174–2175. Springer, New York (2011)
74. Arbuthnott, K., Frank, J.: Trail making test: part B as a measure of executive control: validation using a set-switching paradigm. J. Clin. Exp. Neuropsychol. **22**(4), 518–528 (2000). https://www.tandfonline.com/doi/abs/10.1076/1380-3395(200008)22:4;1-0;FT518
75. Ostrosky-Sol's, F., Lozano, A.: Digit span: effect of education and culture. Int. J. Psychol. **41**(5), 333–341 (2006). https://doi.org/10.1080/00207590500345724
76. Scarpina, F., Tagini, S.: The Stroop color and word test. Front. Psychol. **8**, 557 (2017). https://doi.org/10.3389/fpsyg.2017.00557
77. John, M.S., Kobus, D.A., Morrison, J.G., Schmorrow, D.: Overview of the DARPA augmented cognition technical integration experiment. Int. J. Human Comput. Interact. **17**(2), 131–149 (2004)
78. Schmorrow, D.D., Kruse, A.A.: DARPA's Augmented Cognition Program: tomorrow's human computer interaction from vision to reality: building cognitively aware computational systems. In: Proceedings of the IEEE Conference on Human Factors and Power Plants, pp. 1–7. IEEE, Scottsdale (2002). https://doi.org/10.1109/HFPP.2002.1042859
79. de Melo, C.M., Kim, K., Norouzi, N., Bruder, G., Welch, G.: Reducing cognitive load and improving warfighter problem solving with intelligent virtual assistants. Front. Psychol. **11**, 3170 (2020). https://doi.org/10.3389/fpsyg.2020.554706
80. Atici-Ulusu, H., Ikiz, Y.D., Taskapilioglu, O., Gunduz, T.: Effects of augmented reality glasses on the cognitive load of assembly operators in the automotive industry. Int. J. Comput. Integr. Manuf. **34**(5), 487–499 (2021)
81. Leppink, J., Paas, F., van Gog, T., van der Vleuten, C.P., van Merriënboer, J.J.: Effects of pairs of problems and examples on task performance and different types of cognitive load. Learn. Instr. **30**, 32–42 (2014). https://doi.org/10.1016/j.learninstruc.2013.12.001
82. Mayer, R.E., Moreno, R.: A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. J. Educ. Psychol. **90**(2), 312 (1998)
83. Thees, M., Kapp, S., Strzys, M.P., Beil, F., Lukowicz, P., Kuhn, J.: Effects of augmented reality on learning and cognitive load in university physics laboratory courses. Comput. Hum. Behav. **108**, 106316 (2020). https://doi.org/10.1016/j.chb.2020.106316
84. Doswell, J.T., Skinner, A.: Augmenting human cognition with adaptive augmented reality. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) Proceedings of the Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems, pp. 104–113. Springer, Cham (2014)
85. Ikiz, Y.D., Atici-Ulusu, H., Taskapilioglu, O., Gunduz, T.: Usage of augmented reality glasses in automotive industry: age-related effects on cognitive load. Int. J. Recent Technol. Eng. **8**(3), 1–6 (2019)
86. Erickson, A., Zi, N., Kim, K., Schubert, R., Jules, J., LaViola, J.J., Bruder, G., Welch, G.F.: Sharing gaze rays for visual target identification tasks in collaborative augmented reality. J. Multimodal User Interf. **14**(4), 353–371 (2020)
87. Norouzi, N., Erickson, A., Kim, K., Schubert, R., LaViola, J., Bruder, G., Welch, G.: Effects of shared gaze parameters on visual target identification task performance in augmented reality. In: Proceedings of the ACM Conference on Spatial User Interaction, pp. 1–11. Association for Computing Machinery, New Orleans (2019). https://doi.org/10.1145/3357251.3357587

88. Buegler, M., Harms, R.L., Balasa, M., Meier, I.B., Exarchos, T., Rai, L., Boyle, R., Tort, A., Kozori, M., Lazarou, E.: Digital biomarker-based individualized prognosis for people at risk of dementia. Alzheimers Dement. **12**(1), e12073 (2020)
89. Qiong, O.U.: A brief introduction to perception. Stud. Lit. Lang. **15**(4), 18–28 (2017). https://doi.org/10.3968/1005
90. Ffytche, D.H., Blom, J.D., Catani, M.: Disorders of visual perception. J. Neurol. Neurosurg. Psychiatry. **81**(11), 1280–1287 (2010). https://doi.org/10.1136/jnnp.2008.171348
91. Tippett, L.J., Miller, L.A., Farah, M.J.: Prosopamnesia: a selective impairment in face learning. Cogn. Neuropsychol. **17**(1–3), 241–255 (2000)
92. Farah, M.J.: Visual Agnosia. MIT Press (2004)
93. Neumann, K., Stephens, D.: Definitions of types of hearing impairment: a discussion paper. Folia Phoniatr. Logop. **63**(1), 43–48 (2011). https://doi.org/10.1159/000316412
94. Van Lancker, D.R., Cummings, J.L., Kreiman, J., Dobkin, B.H.: Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex. **24**(2), 195–209 (1988)
95. Collins, C.C.: On mobility aids for the blind. In: Electronic Spatial Sensing for the Blind, pp. 35–64. Springer (1985)
96. Liu, Y., Stiles, N.R., Meister, M.: Augmented reality powers a cognitive assistant for the blind. eLife. **7**, e37841 (2018). https://doi.org/10.7554/eLife.37841
97. Hirzer, M.: Marker detection for augmented reality applications. Comput. Graph. Vis. TUG **25** (2008). https://www.tugraz.at/fileadmin/user_upload/Institute/ICG/Documents/lrs/pubs/hirzer_tr_2008.pdf
98. Seeing with Sound. https://www.seeingwithsound.com/. Accessed 22 Jan 2021
99. Coughlan, J.M., Shen, H., Biggs, B.: Towards accessible audio labeling of 3D objects. J. Technol. Persons Disabilit. **8**, 210 (2020)
100. Real, S., Araujo, A.: VES: a mixed-reality system to assist multisensory spatial perception and cognition for blind and visually impaired people. Appl. Sci. **10**(2) (2020). https://doi.org/10.3390/app10020523
101. Márquez-Olivera, M., Juárez-Gracia, A.-G., Hernández-Herrera, V., Argüelles-Cruz, A.-J., López-Yáñez, I.: System for face recognition under different facial expressions using a new associative hybrid model Amαβ-KNN for people with visual impairment or prosopagnosia. Sensors. **19**(3), 578 (2019)
102. Xu, Q., Chia, S.C., Mandal, B., Li, L., Lim, J.-H., Mukawa, M.A., Tan, C.: SocioGlass: social interaction assistance with face recognition on Google Glass. Sci. Phone Apps Mobile Dev. **2**(1), 7 (2016)
103. Dai, S., Arechiga, N., Lin, C.-W., Shiraishi, S.: Augmented reality vehicular assistance for color blindness, US Patent (7 January 2020)
104. Wang, X.-Y., Wang, T., Bu, J.: Color image segmentation using pixel wise support vector machine classification. Pattern Recogn. **44**(4), 777–787 (2011)
105. Langlotz, T., Sutton, J., Zollmann, S., Itoh, Y., Regenbrecht, H.: ChromaGlasses: computational glasses for compensating colour blindness. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–12. Association for Computing Machinery, New York (2018)
106. Fuller, T.L., Sadovnik, A.: Image level color classification for colorblind assistance. In: Proceedings of the JEFF International Conference on Image Processing (ICIP), pp. 1985–1989. IEEE, Easton (2017). https://doi.org/10.1109/ICIP.2017.8296629
107. McKelvey, C., Dreyer, R., Zhu, D., Wang, W., Quarles, J.: Energy-oriented designs of an augmented-reality application on a VUZIX Blade smart glass. In: Proceedings of the IFFF International Green and Sustainable Computing Conference (IGSC), pp. 1–8. IEEE, Alexandria (2019)
108. Pajorová, E., Hluchý, L.: Augmented reality as a higher education form for students with delimited ability. In: Uskov, V.L., Howlett, R.J., Jain, L.C. (eds.) Smart Education and e-Learning 2019, pp. 461–469. Springer, Singapore (2019)
109. Kiss, F., Woundefinedniak, P.W., Biener, V., Knierim, P., Schmidt, A.: VUM: understanding requirements for a virtual ubiquitous microscope. In: Proceedings of the International Conference on Mobile and Ubiquitous Multimedia, pp. 259–266. Association for Computing Machinery, Essen (2020)
110. Younis, O., Al-Nuaimy, W., Alomari, M.H., Rowe, F.: A hazard detection and tracking system for people with peripheral vision loss using smart glasses and augmented reality. Int. J. Adv. Comput. Sci. Appl. **10**(2), 1 (2019). https://doi.org/10.14569/IJACSA.2019.0100201
111. Wiener, J.M., Büchner, S.J., Holscher, C.: Taxonomy of human wayfinding tasks: a knowledge-based approach. Spatial Cognit. Comput. **9**(2), 152–165 (2009)
112. Jacobson, R.D.: Cognitive mapping without sight: four preliminary studies of spatial learning. J. Environ. Psychol. **18**(3), 289–305 (1998)
113. Freksa, C., Klippel, A., Winter, S.: A cognitive perspective on spatial context. In: Proceedings of the Dagstuhl Seminar. Schloss Dagstuhl-Leibniz-Zentrum fur Informatik (2007)
114. Li, L., Xu, Q., Chandrasekhar, V., Lim, J.-H., Tan, C., Mukawa, M.A.: A wearable virtual usher for vision-based cognitive indoor navigation. IEEE Trans. Cybernet. **47**(4), 841–854 (2016)
115. Mehra, R., Brimijoin, O., Robinson, P., Lunner, T.: Potential of augmented reality platforms to improve individual hearing aids and to support more ecologically valid research. Ear Hear. **41**, 140S–146S (2020)
116. Berger, A., Kostak, M., Maly, F.: Mobile AR solution for deaf people. In: Awan, I., Younas, M., Ünal, P., Aleksy, M. (eds.) Mobile Web and Intelligent Information Systems, pp. 243–254. Springer, Cham (2019)
117. Roesner, F., Kohno, T., Molnar, D.: Security and privacy for augmented reality systems. Commun. ACM. **57**(4), 88–96 (2014). https://doi.org/10.1145/2580723.2580730

Zhanat Makhataeva received the B.S. and M.S. degrees both in robotics from Nazarbayev University (Kazakhstan) in 2017 and 2019, respectively. Currently, she continues her doctoral studies at the Institute of Smart Systems and Artificial Intelligence (ISSAI) of Nazarbayev University. Her research interests include augmented reality and human-robot interaction.

Tolegen Akhmetov received the B.S. degree in Information Systems from Suleyman Demirel University (Kazakhstan) in 2012, and the M.S. degree in Information Systems from Yessenov University, Aktau, Kazakhstan, in 2016. Currently, he pursues his PhD at the Institute of Smart Systems and Artificial Intelligence of Nazarbayev University. His main research interests are task and motion planning for augmented reality.

Huseyin Atakan Varol received his PhD from Vanderbilt University (the USA) in 2009. He joined Nazarbayev University (Kazakhstan) as a robotics faculty member in 2011. Since 2019, he serves as the founding director of the Institute of Smart Systems and Artificial Intelligence at Nazarbayev University. His research interests include machine learning, soft robotics, augmented reality, sensor fusion, and tensegrity systems.

Part VII

Convergence with Emerging Technologies

32 The Augmented Reality Internet of Things: Opportunities of Embodied Interactions in Transreality

Kangsoo Kim, Nahal Norouzi, Dongsik Jo, Gerd Bruder, and Gregory F. Welch

Contents

32.1 **Introduction** . . . 798

32.2 **Background** . . . 799
32.2.1 Augmented, Virtual, and Mixed Reality . . . 799
32.2.2 Artificial Intelligence and Ubiquitous Computing . . . 800

32.3 **The Augmented Reality Internet of Things in Transreality** . . . 802
32.3.1 Convergence of Augmented Reality with Artificial Intelligence and Ubiquitous Computing . . . 802
32.3.2 Transreality and The Augmented Reality Internet of Things . . . 807

32.4 **AR-IoT Framework and Interaction Design** . . . 808
32.4.1 Object-Centric AR-IoT Data Management . . . 810
32.4.2 Scalable AR-IoT Recognition and Tracking . . . 812
32.4.3 Context-Based AR-IoT Interactions and Content Interoperability . . . 814
32.4.4 AR-IoT Framework Evaluation . . . 815

32.5 **Opportunities of Embodied Interactions in AR-IoT Environments** . . . 817
32.5.1 Embodied Interactions in AR-IoT Environments . . . 817
32.5.2 Embodied AR-IoT Agent Prototypes . . . 818
32.5.3 Embodied AR Agent Insights . . . 819
32.5.4 Potential Use Cases . . . 820

32.6 **Conclusions** . . . 821

References . . . 822

Abstract

Human society is encountering a new wave of advancements related to smart connected technologies with the convergence of different traditionally separate fields, which can be characterized by a fusion of technologies that merge and tightly integrate the physical, digital, and biological spheres. In this new paradigm of convergence, all the physical and digital things will become more and more intelligent and connected to each other through the Internet, and the boundary between them will blur and become seamless. In particular, augmented/mixed reality (AR/MR), which combines virtual content with the real environment, is experiencing an unprecedented golden era along with dramatic technological achievements and increasing public interest. Together with advanced artificial intelligence (AI) and ubiquitous computing empowered by the Internet of Things/Everything (IoT/IoE) systems, AR can be our ultimate interface to interact with both digital (virtual) and physical (real) worlds while pervasively mediating and enriching our lives.

In this chapter, we describe the concept of *transreality* that symbiotically connects the physical and the virtual worlds, incorporating the aforementioned advanced technologies, and illustrate how such transreality environments can transform our activities in it, providing intelligent and intuitive interaction with the environment while exploring prior research literature in this domain. We also present the potential of virtually embodied interactions – e.g., employing virtual avatars and agents – in highly connected transreality spaces for enhancing human abilities and perception. Recent ongoing research focusing on the effects of embodied interaction are described and discussed in different aspects, such as perceptual, cognitive,

K. Kim (✉)
Department of Electrical and Software Engineering, University of Calgary, Calgary, AB, Canada
e-mail: kangsoo.kim@ucalgary.ca

N. Norouzi · G. Bruder
University of Central Florida, Orlando, FL, USA
e-mail: nahal.norouzi@ucf.edu; bruder@ucf.edu

D. Jo
School of IT Convergence, University of Ulsan, Ulsan, Republic of Korea
e-mail: dongsikjo@ulsan.ac.kr

G. F. Welch
University of Central Florida, Orlando, FL, USA
e-mail: welch@ucf.edu

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_32

and social contexts. The chapter will end with discussions of potential research directions in the future and implications related to the user experience in transreality.

Keywords

Augmented reality · Internet of Things · AR-IoT · Transreality · Embodied interaction · Virtual avatars · Virtual agents

32.1 Introduction

In a definition of augmented reality (AR) given by Ron Azuma in 1997, he introduced AR as a variant of virtual reality (VR), allowing "the user to see the real world, with virtual objects superimposed upon or composited with the real world" [1]. An AR experience is commonly achieved through stereoscopic video/optical see-through head-mounted displays (HMDs) or mobile phones utilizing an inside-out or outside-in tracking method to appropriately register virtual objects to the physical world. Although the history of AR goes back to the 1960s [2], the technological advances in recent years have resulted in a dramatic increase in AR research and public interest and adoption [3].

While AR is the product of interdisciplinary research, human society is encountering a new wave of advancements through the convergence of further traditionally separate fields, called the *Fourth Industrial Revolution*, which can be characterized by a fusion of technologies that merge and tightly integrate the physical, digital, and biological spheres [4]. In this new paradigm of convergence, all the physical and digital things will become more and more intelligent and connected to each other through the Internet, and the boundary between them will blur and become seamless.

While both academia and the AR industry are experiencing dramatic technological advances, advanced artificial intelligence (AI) and ubiquitous computing empowered by the Internet of Things (IoT) have been actively merged with the AR technology, and AR researchers and practitioners have found such a convergence as the key for realizing more intelligent and interactive environments [5–8]. IoT generally involves a network of computing devices embedded in the physical environment as a form of ordinary daily objects for intelligently sensing, collecting, communicating, and even interacting with the users and the objects themselves. This can provide the basis for *smart* environments through collective big data analyses and context-based services, e.g., real-time analytics and automation [9–12]. AR users in such environments can perceive and understand the world more effectively and efficiently by extending their sensing abilities and intelligence through the intelligent IoT devices and achieve better user experience and performance in given contexts/tasks, e.g., solving problems or searching information [13, 14].

From a traditional standpoint, it might seem that AR and IoT have different objectives with seemingly unrelated concepts, but they can, in fact, be complementary to each other [15, 16]. In principle, a spatially registered and visually augmented content in AR offers a direct and semi-tangible interface and is, thus, easy to comprehend and highly useful, particularly for everyday and/or anywhere usage [15]. Such AR interfaces enable the users to visualize and interact with IoT-enabled smart objects and their associated data in more convenient and intuitive ways. The AR client, such as a mobile or head-mounted glasses-like device, is capable of instantly connecting to an IoT device. Through this connection, users can access/receive context-relevant object/environment-specific data, control information and associated AR datasets for the given targeted service, understand the state (or how to operate) with current datasets from the IoT product, and interact with the physical object using direct control by natural interaction [15]. The interface can be visualized and operated in real time via augmented virtual content that is connected to IoT objects in the real world [17]. Conversely, for AR, IoT as an infrastructure for pervasive "anywhere" services offers an efficient way to make AR "scalable" to the same degree by handling the necessary data management (e.g., tracking data and content) in a distributed and object-centric fashion [5]. Thus, any IoT device can be accessed locally in a seamless manner, and the scalable interface allows for location-based geographical and AR services using AR clients [15]. Additionally, context-aware AR services are made possible by tapping into the refined environment information offered by the IoT infrastructure [5].

To further promote the idea, that is, the synergistic convergence of AR and IoT, this chapter describes the concept of *transreality* that symbiotically connects the physical and the virtual worlds, incorporating elements achieved through the convergence of AR, AI, and UbiComp. Further, this synergy illustrates how such transreality environments equipped with AR-IoT devices can transform our activities, providing highly intelligent and intuitive interactions with the environment, such as AR in situ interfaces adopting natural communication metaphors, and embodied interaction of AR agents and avatars facilitating a bidirectional relationship between entities in the real and the virtual worlds. The overarching goals and contributions of this chapter include the following:

- Introduce the concept of AR-IoT-based transreality empowered by the convergence of AR, AI, and UbiComp (Sect. 32.3),
- Provide a comprehensive knowledge base through a survey of the literature, which covers previous findings and

recent trends of the convergence research empowering AR-IoT-based transreality interactions and environments (Sects. 32.2 and 32.3),
- Propose an AR-IoT framework, identify key features/benefits, and present supporting evidence on its efficacy (Sect. 32.4),
- Describe the transreality realm of interactions in the context of in situ virtual interfaces and embodied agents and avatars (Sect. 32.5),
- Present exemplary use cases of AR-IoT while providing insights for potential future research directions related to pervasive and physically interactive AR (Sect. 32.5).

Through the remainder of this chapter, we first provide a fundamental overview of AR and IoT, covering broader technological concepts, e.g., mixed reality (MR), AI, and ubiquitous computing, which are traditionally considered separate technology thrusts but offer synergistic benefits.

Then, we introduce a concept of *transreality* – where humans, virtual, and physical objects are interacting with each other dynamically and intuitively through AR and smart objects, such as IoT-embedded devices [18]. To underline the importance of and trends toward this transreality, we survey the current state of AR research and services together with IoT and AI technologies while considering their scalability and integration into a unified framework as a control/interaction interface. We discuss the issue of scalability, which relates to the number of objects, object recognition, and data management, as well as tracking techniques that can support AR objects without significant latency, which is a crucial factor for the AR system performance.

Given this concept of transreality, we present a high-level AR-IoT framework for a smart and interactive environment while addressing three aspects of data management, object-guided tracking, and the interface design, which we believe are key components for the development of an efficient AR-IoT infrastructure [19].

We then explore how virtually embodied interactions, such as facilitated by virtual avatars or agents in AR, could be used in AR-IoT-embedded transreality environments and how such interactions could influence the users' perception and behavior in social contexts while manipulating physical or virtual objects around them. In addition, we illustrate use case scenarios that could incorporate and benefit from the AR-IoT framework and transreality environments while discussing how AR can offer an intuitive and natural method for communication with IoT objects, compared to other methods, such as using a graphical user interface (GUI) with no visual, contextual, or spatial registration.

Finally, we provide concluding remarks and a summary of this chapter while presenting insights and potential future research directions. Note that this chapter is written largely based on the authors' prior research and other relevant research in this domain, in particular two previous publications that the authors led on and wrote [19, 20].

32.2 Background

In this section, we provide a fundamental overview of traditionally independent technology thrusts, including AR/VR/MR, AI, ubiquitous computing, and the IoT paradigm, before we introduce the convergence of these technologies and the concept of transreality.

32.2.1 Augmented, Virtual, and Mixed Reality

In the fields of computer science and human-computer interaction (HCI), VR refers to a technology-mediated simulation that can create real or artificial experiences, where people can interact with a simulated computer-generated virtual environment. While VR-related concepts and technologies have been anticipated and discussed in science fiction books since well before the 1950s, the term "virtual reality" was coined and popularized by Jaron Lanier in the 1980s at a time when the technology was becoming available to researchers and end users [21, 22]. Anticipating future advances in this field, up to the point where human users may not be able to distinguish the virtual from the real, various terms such as simulated reality, hyperreality, and synthetic reality have been introduced and defined [23]. In such forms of reality, users may not be aware of whether the outside world is simulated or not through high-fidelity simulated multisensory feedback that is indistinguishable from natural sensations and perception [24].

Unlike VR, which strives toward complete immersion of users in a virtual environment, in AR, users can experience virtual content blended with real objects and environments as if the content exists as part of the real world. AR has existed in many forms since Sutherland presented the first prototype optical see-through HMD and discussed the "Ultimate Display" in the 1960s [2, 25]. Although AR/VR-related concepts were evolving with respect to different perspectives, the broader definition of MR is traced back to a seminal paper written by Milgram and Kishino in 1994 [26], which is the first academic paper to use the term "mixed reality" in the context of computer interfaces. They defined MR as "a particular subset of VR related technologies that involve the merging of real and virtual worlds somewhere along the 'virtuality continuum' which connects completely real environments to completely virtual ones." As shown in Fig. 32.1, the "reality–virtuality continuum" spans from completely real to completely virtual environments and includes AR and augmented virtuality (AV). AR denotes experiences that superimpose virtual objects on the real environment, while AV can be

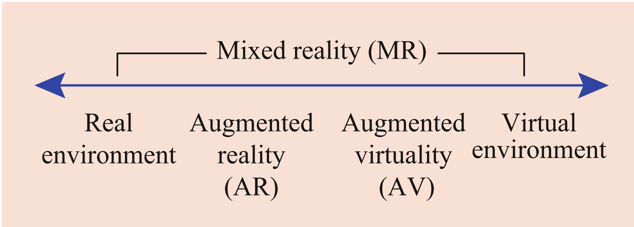


Fig. 32.1 Reality-virtuality continuum by Milgram and Kishino [26]

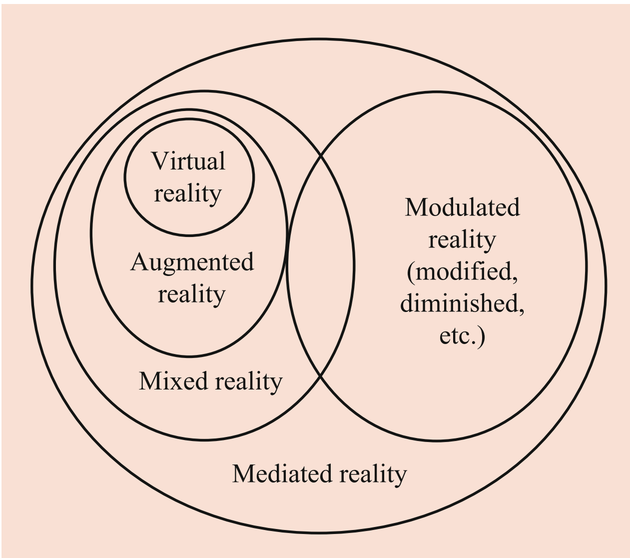


Fig. 32.2 The concept of mediated reality by Mann [27]

thought of as the other way around when superimposing real objects on the virtual environment.

Later, Mann [27] placed MR in a more generalized concept of "mediated reality" including VR and other forms of "modulated" reality like *diminished* reality, which tries to remove objects visually from the perceptible real world (Fig. 32.2). Mann et al. [28] recently illustrated a variety of concepts related to "reality" including "eXtended Reality (XR)" that tries to include the extreme ends of the virtuality continuum, i.e., reality and virtuality, which Milgram and Koshino did not include under the concept of MR. They presented a broader term, "Multimediated Reality (All R)," which is multidimensional, multisensory, multimodal, and multidisciplinary, by building a sophisticated taxonomy that covers different continuums related to virtuality and reality in a perspective of interactive multimedia.

Given the context of AR and IoT in this chapter, our focus is on AR/MR rather than immersive VR experiences. The most popular definition of AR is from Azuma in 1997 [1], where he defined AR with three characteristics: (i) it combines real and virtual objects in a real environment; (ii) it runs interactively in real time; and (iii) it registers (aligns) real and virtual object with each other (in 3D). In a 2017 paper, Azuma predicted a brighter future for AR compared to VR in commercial markets because of the potential to improve the user's understanding of and interaction with the real world, while it could replace all other display form factors, such as smartphones and tablets, via wearable devices like head-worn AR displays [29]. Recently, the term MR has become synonymous with a highly interactive version or expansion of AR [30]. In that sense, AR/MR involves not only the interaction between users and computers – e.g., virtual entities – but also the surrounding environment in the context of interaction.

Recent developments in AR/MR technology provided powerful but affordable computing devices for compelling AR/MR experiences with an unprecedented public interest in AR/MR [3, 31]. For example, a multiuser mobile AR game, Pokémon Go, was widely adopted around the world by 25 million active users in the United States and 40 million worldwide [32], and AR/VR startups raised over $3 billion investment in 2017 [33], while major IT companies, such as Apple, Meta (Facebook), Google, and Microsoft, have been investing and developing their own AR/MR platforms and technologies.

32.2.2 Artificial Intelligence and Ubiquitous Computing

As a separate technological thrust, AI denotes a field concerned with intelligence demonstrated by computing machines mimicking the natural intelligence of humans or animals, which has a long research history in computer science and cognitive science from the mid-1950s [34]. The term "artificial intelligence" was first coined by John McCarthy in 1956 [34], who is considered as one of the founders and leaders of AI, together with Marvin Minsky, Allen Newell, Arthur Samuel, and Herbert Simon. Recently, the field of AI has gained public attention and experienced significant technological achievements, thanks to novel computing devices, the increasing amounts of available data, and advanced data processing techniques, such as deep learning [35]. AI has become an essential part of the IT industry, and consumers are gaining more and more access to AI when performing tasks in their ordinary lives [36].

Ubiquitous computing (UbiComp), and the related areas of pervasive computing or calm technology [37], is a related technological concept that describes the notion that all computing occurs anytime and anywhere so that users do not distinctly realize that it is happening. Mark Weiser, who is widely considered to be the father of UbiComp, claimed that "the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it." in his pioneering article, *The Computer for the 21st Century* in 1991 [38].

For example, a door could intelligently recognize people who have authority to enter the room and open automatically for them, and a refrigerator could identify missing food items in it and issue delivery orders to fill it up.

While AI and UbiComp relate to each other complementarily, the technologies tend to have fundamentally different intentions. For example, UbiComp envisions computer technology as an invisible tool that is available throughout the physical environment while disappearing from the user's consciousness, but AI is more keen to turn computer technology into social/intelligent agents [39]. Nowadays, these novel concepts of technologies, which the research pioneers introduced and proposed decades ago, have come true and become absorbed in our daily lives more and more through various technological advances and realizations. Commercialized intelligent virtual agents, IoT, and the concept of smart connected environments, which we describe in the rest of this section, are some of the examples/concepts that can benefit from the convergence with AR technology.

Intelligent Virtual Agents

In AI literature, computational *intelligent agents* are defined as any entity or device that can autonomously perceive its environment or context through observations using sensors and take actions to maximize its chance of successfully achieving its goals incorporating any physical/virtual actuators [40].

Not surprisingly, intelligent agents have been an important component for UbiComp systems in different forms, such as for system-level decision-making and natural user interfaces like chatbots and virtual assistants [41]. A chatbot, with which the users have conversations through text or text-to-speech, is a software application popularly used in online chat services. Although the concept of chatbot traces back to the 1950s when Alan Turing proposed the Turing test as a criterion of intelligence, the term "ChatterBot" was originally coined by Michael Mauldin in 1994 [42], and since then, various chatbot systems have been proposed and examined in terms of the user's perception and the benefits of the systems [43, 44].

The concept of intelligent virtual agents (IVAs) has become prevalent to the public thanks to the recent commercialized digital assistant systems, such as Amazon Alexa, Apple Siri, Google Assistant, and Microsoft Cortana [45]. The products are embedded in various types of computing devices ranging from desktops to smartphones, even including cars and home appliances, and use sophisticated microphones and signal processing to capture human speech in various places, e.g., one or more rooms of one's house. In this way, users can interact with IVAs in *natural* ways, e.g., verbal conversations, and the IVAs can perform a variety of tasks, such as playing music, answering basic questions, checking on sports scores or weather, and more via Internet-based services, while supporting human users [46]. Products like the Amazon Echo Show further add a screen for displaying basic content such as photos, videos, and weather forecasts.

Many IVA researchers focus on how to develop more effective agent systems and understanding of human perception of and behaviors with the agent systems, compared to the conventional interaction methods, e.g., improving satisfaction [47] and task performance [48]. Recently, regarding security and safety issues, research on reliability or trust in IVA systems have gained attention from both researchers and the public and become an important research topic [49].

Internet of Things

The term Internet of Things (IoT) was first introduced by Kevin Ashton in 1999, and 10 years later, he reiterated his vision of IoT through the addition of various sensors to computers and *things* to collect information and understand their environment without the need for human input [50]. Over the years, increasing interest in the IoT paradigm resulted in a multitude of industrial and research efforts, and many researchers aimed at capturing its different definitions, applications, implementations, and challenges [51–54]. For instance, in a highly cited review of IoT, Atzori et al. [51] survey the various definitions and the different perspectives adopted in defining this paradigm, such as "*things oriented*" or "*Internet oriented*" points of view. From the "things-oriented" perspective, although the conception of IoT from Auto-ID Labs [55] was initially only considering radiofrequency identification (RFID) tags as *things*, the concept of things soon evolved to include a more broader definition, such as the vision of the International Telecommunication Union that said, "from anytime, any place connectivity for anyone, we will now have connectivity for anything" [56]. Other descriptions, such as the one given by the European Commission, emphasized the intelligent nature of this connectivity by saying, "Things having identities and virtual personalities operating in smart spaces using intelligent interfaces to connect and communicate within social, environmental, and user contexts" [57]. On the other side, the "Internet-oriented" point of view, such as the Internet Protocol for Smart Objects (IPSO) alliance [58], focuses mostly on the networking means for connection of smart things. More recently, Rayes and Salam proposed their definition of IoT, also including the role of people as Internet users, "IoT is the network of things, with device identification, embedded intelligence, and sensing and acting capabilities, connecting people and things over the Internet" [59].

Overall, with the increased ubiquity of IoT devices, advanced AI and IVA technologies in pervasive environments can enable not only more sophisticated interpretations of and semantic understanding in the given context but also appropriate, secure, and useful actions performed by pervasive computing modules, such as IoT devices. Deep learning solutions for better data extraction and enhanced security are just a few examples [60, 61]. Reportedly, the total number of connected IoT units will reach 83 billion by 2024, rising from 35 billion in 2020: a growth of 130% over the next four years [62]. Manufacturers in consumer industry, e.g., Amazon and other companies, networked their IVA devices to IoT and related smart home appliances and found an important application field for IVAs, resulting in a novel research thrust and mutually beneficial overlap between various research fields. Many of the research topics pursued in the IoT field, such as privacy [63, 64], the relationship between edge and cloud processing [65], and network traffic optimization [66], will need to be reevaluated when IoT is deployed in the context of AR and IVA. Furthermore, some IoT applications, such as smart healthcare [67–69], can benefit from the addition of AR and IVA techniques.

Smart Connected Environments

The potential of the aforementioned AI and UbiComp, particularly IVA and IoT technologies, and their products further extend to home automation and more general interactions in smart connected environments with the increasing present smart devices [70, 71]. The term *smart* implies the acquisition of context information and the automated reaction to the context [39, 72, 73]. The concepts of smart objects and environments trace back to the late 1960s [73].

Smart environments are a space or a world where various sensors and computers are integrated with everyday objects and connected to each other through networks as an extension of UbiComp [74]. Smart objects normally have a physical representation in the real world [75, 76], while virtual entities could be designed to be smart in the digital world, mimicking the characteristics of physical objects, such as appearance and behavior [72]. Related research mostly deals with the smart objects' interactions that take place as reactions to a context, which can be defined by other objects, the environments, and the users, for example, sensing human behaviors or activities [77], effective customized interfaces [78], security [79], and efficient communication techniques among the devices and the framework [80] in the smart environment.

The term "smart" in technology and the connected environments became popular together with the emergence of IoT because they are fundamentally complementary in research and for practical realizations [76, 81]. However, while UbiComp and IoT focus more on the tools that integrate in the physical world disappearing from consciousness through adaptation to the contexts, smart connected environments extend this understanding to intelligent and social interaction covering the concept of AI and agency. In other words, the IoT and agents establish the framework for smart connected environments with smart acting objects through connectivity [39].

When it comes to smart connected environments, smart homes are an important direction and driving force for research [70, 77, 79, 80, 82]. However, the scope of smart environments is growing beyond the scale of home environments to include urban areas and cities [83]. For example, the paradigm of Industrial 4.0 entails the trend toward automation and data exchange in manufacturing technologies and processes [4] and factories incorporating smart machines that are augmented with wireless connectivity, sensors, and actuators, hence the term smart factories [84]. Information about buildings can be continuously monitored through various sensors embedded in smart buildings, such as building operation and maintenance requirements [85]. Smart cities equipped with IoT devices and connected networks can help us find sustainable solutions to the issues that we encounter while the population is growing, e.g., effective and efficient city management and public community services [86]. In such smart environments where all smart objects are embedded and we become oblivious to their existence, AR and virtual avatars/agents can be an effective interface to interact with the user's local or even remotely connected environments.

32.3 The Augmented Reality Internet of Things in Transreality

In the previous section, we covered different technological concepts, which traditionally have developed separate research thrusts – broadly AR, AI, and UbiComp. Here, we survey previous research focusing on the convergence of the technologies in depth and introduce the concept of *transreality*, where the synergistic benefits of AR and IoT technologies, empowered by this convergence research, are emphasized in the blended virtual and real world.

The National Science Foundation emphasized the importance of convergence of different fields of research, as no discipline alone can resolve the complex challenges facing the world today [87]. They also differentiated between "convergence research" and various types of multidisciplinary research, as convergence of research fields is captured from the earliest to the last stages of the research process facilitated by teams formed upon intellectual diversity. Most importantly, "convergence research" not only works toward solving the complex problems through integration of knowledge and technology but contributes to the creation of novel research methodologies and avenues [88], which in our case is focused on the nexus of AR, AI, and UbiComp, and more specifically the convergence of AR and IoT.

32.3.1 Convergence of Augmented Reality with Artificial Intelligence and Ubiquitous Computing

Over the years, continuous research in the three fields of AR, AI, and UbiComp has led to individual growth in knowledge and advances in technology within each of the three fields. However, in line with the visions of "convergence research," a trend of converging knowledge and technology is growing among the three fields, facilitating the development of novel research areas and questions. In this section, we present some of the previous research focused on the convergence between the three fields of AR, UbiComp, and AI.

Artificial Intelligence and Ubiquitous Computing

As addressed earlier in Sect. 32.2.2, AI and UbiComp are relatively close to each other and complementary in terms of the research goals and directions. The intersection of AI, UbiComp, and communication was first captured in the term *ambient intelligence* by Eli Zelkha and Simon Birrell in 1998 [73, 76]. The main feature of ambient intelligence was to be human-centric, described by Ducatel et al. [89] as, "it is aware of the specific characteristics of human presence and personalities, takes care of needs and is capable of responding intelligently to spoken or gestured indications of desire, and even can engage in intelligent dialogue." This facilitation of the user is achieved through observing and understanding the environment in a non-intrusive, transparent, and ethically acceptable way [90]. Ramos et al. [91] emphasized the importance of AI in the ambient intelligence vision, dividing the ambient intelligence system into two parts: the operational layer and the intelligent layer. While the operational layer covers the technologies that control and manage the data and actions, such as hardware sensors and actuators, communications, computer graphics, and ubiquitous computing, the intelligent layer includes a multitude of core AI, such as natural language processing, computer vision, incompleteness, and uncertainty, to enable use in a wide range of environments and deal with unprecedented events that might arise.

Over the years, various research efforts were conducted to realize the ambient intelligence vision and identify the important challenges of the field to facilitate various applications. One such challenge is context awareness of the system, which requires a full understanding of the environment and the user's behavior and activities [92, 93]. Plotz et al. [92] pointed out the limited generalizability of activity recognition approaches necessary for context-awareness which rely on domain knowledge. They presented a domain-independent feature-learning framework, which outperformed classical heuristic approaches tested on four different datasets. Doctor et al. [94] proposed a type-2 fuzzy intelligent agent that would learn from the user through long-term interaction in order to perform on their behalf. In a 5-day experiment, they tested the capabilities of their agent using an intelligent dormitory called "iDorm" equipped with various sensors and actuators for sensing and affecting different aspects of the environment, such as light and temperature. They found that their approach was capable of dealing with environmental and personal uncertainty in a non-intrusive manner. Patterson et al. [95] designed a cognition assistant system for patients, mainly with Alzheimer's disease, for spatial and daily tasks. Their presented system was based on the detection of location, state, and activity. For instance, patients could be reminded of the required sequence of steps to complete a daily task based on their location and their motion.

Recently, UbiComp technology and environments are becoming more advanced and complex, presenting new challenges that require more robust solutions. Georgievski and Aiello pointed out the importance of AI and planning methods as a solution to continuously evolving UbiComp environments [96]. They surveyed 53 papers focused on AI planning within the UbiComp field and develop a framework classifying the literature and capturing the main elements in three dimensions of environments, planning, and interpretation. Based on their findings, they presented opportunities for future research, such as focusing on user preference and the requirement of more detailed spatial information for specific scenarios, e.g., recognizing human posture during emergency scenarios.

UbiComp's *invisible* nature and their increasing appearance in our daily lives have brought about new challenges in terms of trust in the system and users' sense of security [97]. To address this concern, D'Angelo et al. [97] introduced a trust-based architecture, where entities within the network make trust decisions for other entities using soft computing and data mining techniques. Although their model was capable of learning malicious tactics, they emphasized the importance of more research in this area, considering possible attacks on the network.

In recent years, fueled by public interest, a large body of research at the intersection of AI and UbiComp focused on IVAs, which are becoming more capable in understanding and communicating with humans verbally and nonverbally. Developing one of the earlier voice-based home assistant prototypes, Soda et al. [98] noted that voice-only interactions can be less informative and, borrowing from IVA literature, integrated different embodied virtual agents in their system. Their findings indicate that the presence of the embodied agents increases users' willingness to interact with the system and enhance their experience by giving feedback about the users' command. With the increasing popularity of intelligent virtual assistants, such as Amazon Alexa, Google Assistant, Apple Siri, and Microsoft Cortana, many researchers conducted comparative studies between these consumer products. For instance, López et al. [99] compared the usability of the four intelligent assistants mentioned above in terms of the quality of services they provide. They pointed out opportunities for more studies to understand the usability of such devices in areas such as counseling or working in tandem with robots. To mitigate factors such as social isolation that can negatively influence the lives of older adults, Reis et al. [100, 101] proposed and tested a model that would facilitate older adults by maintaining their social connections using consumer intelligent assistants. Druga et al. [102] investigated the impact of virtual/robotic intelligent agents, such as Amazon's Alexa and Anki's Cozmo on children's engagement and perceptions of the agent. Various ideas arose from their study, such as the importance of agent's mobility, having facial expressions, and mirroring the child's interaction style. Knote et al. [103] conducted a systematic literature review to capture the theories and principles used in designing smart personal assistants and their application areas. Their findings resulted in five main principles: context awareness, self-evolution, multimodality, anthropomorphism, and platform integration and extensibility. Austerjost et al. [104] pointed out the opportunity of extending the use of smart assistants to spaces other than homes, such as laboratories. They designed their voice-based agent, which was capable of interfacing with laboratory equipment and facilitating tasks, such as modifying the settings of a device or reading out values.

In recent years, researchers have utilized simulators to better design and test the features of smart home environments, where AI and UbiComp technologies are incorporated and merged; however, certain limitations still exist, such as unavailability of varied and realistic data from sensors. Lee and Cho et al. [105] and Helal et al. [106] designed simulators equipped with virtual sensors acting as real sensors, IVAs in the environment behaving as human occupants, and a configurable space with a 3D GUI.

IVAs that are designed to interface with IoT devices face many of the security concerns that exist in the basic concepts of UbiComp technologies, such as continuous and invisible monitoring of the user for the purpose of facilitating their needs. Chung and Igora et al. [107] analyzed the potential security vulnerabilities of such IVAs as a means to identify areas in need of security provision. Separately, information from many IVAs can potentially be used for crime investigation. Chung and Park et al. [108] developed a tool for digital forensics called *CIFT* aimed at understanding the use of Amazon Alexa in this application domain. To account for potential issues that may arise with the dominance of one type of IVA over the others, such as privacy and interoperability, Campagna et al. [109] developed an open IVA called *Almond* equipped with features and structures to address these issues.

Overall, past research in the intersection of AI and UbiComp covered a wide range of topics, from the development of ambient intelligence prototypes to the development of AI algorithms to better provide for the wide range of scenarios involving ubiquitous computers in areas such as action recognition, system validation, trust, and IVAs. The further adoption of AR interfaces in ambient intelligence spaces will bring about novel interactions and pose new questions for both the industry and academia.

Augmented Reality and Artificial Intelligence

AR, or more broadly MR, environments are capable of creating a multimodal experience for users with visual and auditory stimuli as the most common sources of information [20]. The inclusion of AI in AR experiences can foster improved spatial scalability and context awareness of AR applications.

One of the highly researched topics in this convergence area has been the development of *embodied* IVAs. To understand the differences in the realization of these embodied IVAs in AR, Holz et al. [110] suggested the terminology of mixed reality agents (MiRA) and presented a taxonomy based on the three dimensions of agency, corporal presence, and interactive capacity. One of the earliest examples of AR agents was first presented at 1993 SIGGRAPH and was called *The ALIVE system* [111]. The ALIVE system's semi-intelligent agent was comprised of various features, such as internal needs and activity hierarchy, and responded to the user detected by a vision-based approach. Another early example of AR agents was Anabuki et al.'s Welbo [112], which was embodied as a robot and helped users design an AR living room. In their work, they emphasized the value of such agents' abilities in interacting with both virtual and physical entities. Also, to the best of our knowledge, the first academic paper that introduced an interactive virtual human in AR is Balcisoy and Thalmann's application in real-time interactive drama with real and virtual actors in 1997 [113].

Although an AR agent's capability to interact with physical entities could enhance a user's experience, it is a technologically difficult task as it requires very precise tracking and semantic understanding of the physical environment [114]. In this area, Barakonyi and Psik et al. [115] presented an animated agent framework that facilitated interactions in AR through sentient computing by forming an understanding and being influenced by its physical surrounding. They discussed specific aspects that are applied to IVAs interacting in AR, such as the ability to both respond and influence the physical environment. For instance, in a multiplayer game called the *MonkeyBridge* developed by Barakonyi and Weilguny et al. [116], embodied agents are able to observe the changes in the physical and virtual pieces in the game and accordingly make decisions on how to reach the game's target. Similarly, to facilitate IVA's interaction within physical environments that are dynamically varying, Checklov et al. [117] developed an approach using the simultaneous localization and mapping framework, allowing the agent to detect planar surfaces on novel objects. More recently, Azad et al. [118] described an MR playground, tested in VR, that would benefit from procedural content generation, enriching the game and taking advantage of the real objects and surfaces in the user's environment, such as a *Mario* style game where the furniture acts as a track for the player's avatar.

One of the areas where IVAs, and more generally AI, have facilitated AR applications is education. Hantono et al. [119] reviewed the literature on the use of augmented reality agents in the education domain from 2005 to 2015. In their work, they identified some of the main challenges of the field, such as personalized customization of the information for each user, and designing mechanisms to reduce task load through the provision of appropriate instructions. In one of the first examples of integration of IVAs for educational use, Wagner et al. [120] designed an art history game and compared the player's experience through different devices. Their findings indicated that the AR experience with a virtual character was rated as most enjoyable compared to other game content conditions: text-only, text and audio, 2D image, and non-AR 3D character. Outside of the context of embodied AR agents, Holstein et al. [121] designed a teacher awareness tool using AR glasses called *Lumilo*. Lumilo is paired with an intelligent tutoring system, through which the teacher can view information about the state of each student and the whole class, such as common errors made when solving problems.

Entertainment is another application area where the convergence of AI and AR has been explored and researched. Cavazza et al. [122] created an interactive storytelling prototype, where the virtual character's role and the storyline evolve using hierarchical task networks, which is affected by inputs from the user. Dow et al. [123] also presented their interactive drama prototype called *Facade*, where players interacted with IVAs in the story through either AR or two desktop-based approaches. Their findings indicated that as expected, the AR interface increased users' sense of presence; however, this effect did not necessarily lead to higher user engagement, which was potentially caused by the type of interactive scenario and a lack of distance for players to easily engage in the game.

Continuous advances in the quality and usability of AR devices and tracking and AI algorithms show promise for the inclusion of intelligent AR interfaces and agents for facilitating users in a more robust and ubiquitous manner.

Augmented Reality and Ubiquitous Computing

The adoption of UbiComp with pervasive physical components opened new research ideas in many AR research areas, such as AR applications, user interfaces, and tracking. For example, the convergence of AR and UbiComp fostered novel approaches for creating multimodal AR user experiences with their environment, such as the ability to interact with physical and virtual objects through situated AR interfaces. In particular, the entertainment domain has been a very popular topic within the AR and UbiComp convergence field by bringing entertainment to users' physical space and taking advantage of elements within that space. Cheok et al. [124] designed an AR game called *Touch-Space*, and included elements of ubiquitous, tangible, and social computing in their game. For instance, a player's location in the physical space could trigger the embedded information about potential mines and treasures affecting the user's score. Montola [125] conducted a survey of pervasive AR games explaining how such games are distributed through time, space, and social structures. He later describes four game categories based on how the environment is utilized in the game, such as *local games* that are developed for the context of a certain location like tourist games. Vogiazou et al. [126] implemented an AR tag game called *citiTag*, where players' locations were transmitted over the network, allowing them to tag players from other teams or get tagged themselves. Their experimental findings reflected players' engagement in the AR game as they borrowed elements from real-world experiences, such as bending the rules in creative ways or wanting to spend more time within the game. Interestingly, players got involved in testing the limits of technology, for instance, by trying to understand the boundaries of the tracking system's accuracy and using it to their benefit. This is in line with Chalmers and MacColl's discussions of *seamful design* and the value of taking advantage of the *seams* that exist in UbiComp environments, such as the limitations in a device's sensing abilities, suggesting that these seams leave room to later be utilized for different purposes [127].

Over the years, AR and UbiComp researchers have developed multiple user interfaces aimed at creating a richer and more interactive experience. Lee et al. [128] pointed out means for enriching AR interfaces with ubiquitous computers but also noted that there still exists a lack of haptic feedback from virtual objects. To enhance such interactions, they developed a system that takes advantage of mediated vibrotactile feedback on the user's hands. Han et al. [129] presented an AR haptics interface called the *HydroRing*, taking into account the importance of ubiquitous interaction with the environment. Using HydroRing, one can sense vibration, pressure, and temperature from various physical and virtual objects, such as the unseen wiring behind a wall or menu items in an AR experience.

Tracking is another area of research that is necessary for creating a continuous and accurate placement of virtual entities in the physical world. Newman et al. [130] pointed out the current tracking limitations of ubiquitous AR approaches since they are mostly developed to be application dependent with limited generalizability. To address this limitation, they introduced their tracking architecture called the *UBITRACK* aimed at covering the needs of different applications within their proposed Milgram-Weiser continuum – considering the spectrum of real-virtual environments and monolithic and ubiquitous computing approaches. Singh and Mantri reviewed the different tracking methods used for AR applications and discussed their limitations, proposing a ubiquitous hybrid tracking approach taking advantage of both vision-based and sensor-based methods [131].

Similar to the convergence of UbiComp and AI, advances in IoT technology and its popularity resulted in a sudden increase of research and new interaction metaphors for AR interfaces [20]. The majority of the research in the AR and IoT convergence area viewed and utilized AR as an interface to communicate with IoT devices and in some cases comparing them with previously used traditional interaction approaches. This point of view was reflected by Gimenez and Pous [7] as well, noting, "AR has recently been touted as one of the ideal interfaces for IoT." For instance, Garcia-Macias et al. [132] described the idea of *sentient visors* for the IoT paradigm, as it comprises of both virtual and physical things, allowing identification of smart objects, relaying information about their services, and providing means to interact with them. They captured this idea in their mobile-based prototype called the *UbiVisor* and illustrated its application in a scenario where a plant pot was equipped with humidity and temperature sensors. Users could access the sensor information and receive situated notifications for watering the plant by pointing their mobile devices at it. Similarly for mobile-based interactions, Wirtz et al. [133] noted that issues, such as the need to communicate with smart objects through mediating apps, restricts the ubiquity of such interactions. They addressed both Internet connectivity and interaction limitations through an AR prototype, where the interaction paradigms were provided directly by each smart object through situated graphical user interfaces. In a similar approach, Heun et al. [134] pointed out that 2D remote interfaces can add unwanted complexity to the user experience in many scenarios for interactions with smart objects. They proposed *The Reality Editor*, allowing users to control smart objects through situated UIs instead.

Several researchers conducted studies to understand the impact of 2D remote interfaces and their situated counterparts in AR environments [15, 135]. Liu et al. [135] emphasized the importance of receiving feedback from the situated UI, designing three task difficulty levels, and comparing them through four different interfaces: text, 2D interface, situated AR overlay without feedback, and situated AR overlay with feedback. Their findings indicated the importance of feedback in interface design as participants performed significantly better in the situated AR overlay with feedback condition compared to others. Similarly, Jo et al. [15] compared situated interactions with traditional 2D interfaces, for instance, turning the lights on and off. Interactions with the situated UI had several benefits such as ease of use, naturalness, and speed, although there were drawbacks such as fatigue, which can be explained by the novelty of the interaction. In search for more universal interfaces for smart and actuated devices, Kasahara et al. [17] designed the *exTouch* system, where by pointing their mobile devices at physical actuated objects, users were able to manipulate them through manipulation of the virtual counterpart.

Other researchers utilized AR glasses instead of mobile AR devices to facilitate the IoT interaction space. In a survey of smart glasses, Lee and Hui [136] described and categorized the capability of smart glasses in facilitating different modes of interaction, such as on-device touch or hands-free input. For instance, Zhang et al. [137] and Kollee et al. [138] presented head-based and gesture-based methods utilized through HMDs to interact with smart objects.

While smart home environments are often considered for the applications in the convergence of AR and IoT, the benefits of the convergence have been actively explored in smart factory environments as well. Hao and Helo et al. [139] proposed using AR as a means to enhance the interaction between human and machine in smart manufacturing environments. They presented a device maintenance scenario where operators can benefit from situated support using smart glasses, such as viewing the device's datasheets.

Interestingly, this convergence area's main focus has been on the utilization of AR interaction metaphors for IoT devices, and less attention has been given to the utilization of IoT for enhancement of AR experiences. Although in various AR projects, researchers have developed Internet-connected prototypes for input/output purposes [140, 141], only a few have leveraged IoT devices and interaction schemes in the realization of their AR experiences [15, 142, 143]. This limitation might be because most AR experiences are envisioned and developed to be ego-centric with the necessary devices (i.e., sensors and displays) collocated with the user and not distributed in the environment. Also, the perception of potential performance complications by utilizing IoT standards might be another contributing factor in the slow adoption of IoT devices in the service of AR experiences [144]. However, in recent years, more research avenues have opened up for realizing ubiquitous and distributed AR experiences fueled by the increase in popularity and use of IoT devices among the general public.

We predict that over the next years, further advances in AR user interfaces and technologies toward everyday use cases in a smart home environment will give rise to the standardized integration of IoT devices into AR frameworks and the development of new IoT devices specifically for the purpose of enhancing AR experiences with improved sensing and multimodal sensations, e.g., driven by the gaming industry.

32.3.2 Transreality and The Augmented Reality Internet of Things

We covered current convergence research with different combinations of AR, AI, and UbiComp technologies by describing the concepts and summarizing some prior literature in the previous section. Here, we introduce the concept of *transreality*, where both physical and virtual objects can sense and understand the user's or other peripheral activities in the environment and dynamically and seamlessly interact with each other as a whole integration of the three research thrusts, while describing other related concepts.

The term transreality often appears in gaming context, e.g., transreality games, which is a type/mode of video games that combines playing the game in a virtual environment with game-related physical experience in the real world [145]. Beyond the gaming context, such transreality could take advantage of various pervasive/ubiquitous, mobile, location-based, and AR/VR/MR technologies to extend the implications to our daily lives aiming at more effective and efficient human-computer interactions. Martin and LaViola [18] introduced a *Transreality Interaction Platform (TrIP)* to "enable service ecosystems which situate virtual objects alongside virtualized physical objects and allow for novel ad-hoc interactions between humans, virtual, and physical objects in a transreality environment" and presented a proof-of-concept implementation while describing the system architecture, data model, and query language.

Similar to transreality, such concepts combining the real and virtual worlds in the domain of UbiComp and AR/MR/VR have been proposed and researched for more than a decade [146–148]. For example, Kim et al. defined *ubiquitous virtual reality (U-VR)* as "a concept of creating ubiquitous VR environments which make VR pervasive into our daily lives and ubiquitous by allowing VR to meet a new infrastructure, i.e. UbiComp" [146].

Paradiso and Landay also presented a concept of *cross-reality* – ubiquitous MR environment that comes from the fusion of technologies, such as ubiquitous sensor/actuator networks and shared online virtual worlds [147]. Following this concept of real-virtual convergence (i.e., cross/dual reality), Lifton et al. [148, 149] proposed several prototypes bridging the physical and virtual worlds through networked sensors and actuators. One example is the *ShadowLab*, inspired by Second Life [150], where the data/information from 35 sensors distributed in a floor of the Media Lab (MIT) were used to visualize features, such as the amount of current drawn from outlets and activity levels detected (i.e., level of sound, motion, and vibration) in the virtual environment.

Mirror world, which is originally introduced in David Gelernter's 1991 book *Mirror Worlds* [151], is another yet similar paradigm described by Ricci et al. [152] to capture the convergence of UbiComp, AI, and AR. In this paradigm, the new interaction space provided by AR and IoT can facilitate operations outside of the limited environments, such as smart rooms, to operations in much larger scales, for example, smart cities and communities. In the larger spaces supported by this paradigm, objects share the counterparts in the real or the mirror world, capable of sensing information from either world and be actuated by the inhabitants of those worlds (i.e., humans or agents).

Another related concept is *hybrid reality* that Seo et al. described in a smart home context [153]. They particularly employed the hybrid reality to overcome the difficulties in the evaluation of smart homes and user experience, such as differences in the layout of smart homes, and the number of smart appliances. In the paradigm, real objects are superimposed with virtual ones, and the virtual environment allows for adding new sensors and actuators. This approach provides opportunities to both developers and potential residents of such spaces to design the features of the space based on the needs of the residents and resolve any potential issues before building the physical space.

The concept of *Digital Twins* is also one of the realizations in this transreality trend. Although the concept was anticipated before, e.g., in the book *Mirror Worlds* [151], the term "digital twin" became widely acknowledged in publications by Michael Grieves who applied the concept in industry manufacturing [154]. The digital twin is basically an integration/connection between the physical product (or living/nonliving entity) and its digital/virtual replica in the virtual world. Through the interactions with the virtual replica, users can have more flexibility to understand and control the physical world beyond the spatial and temporal limitations in the real world. The concept, which aims to create the exact same virtual twin of the physical part, is compared to other cross-reality concepts that generally focus on the mere synchronization of the physical world and the digital representation, such as an abstraction of some aspects of the physical world. Glaessgen and Stargel [155] pointed out the difficulty of appropriate maintenance and sustainable management of complex vehicle systems in NASA and Air Force and proposed the *digital twin* paradigm, which is a high-fidelity simulation space containing a digital replica of the real object equipped with all the necessary features, such as the physical model, sensor information, and maintenance data captured from the real twin, to resolve these challenges. Tao et al. [156] also emphasized the importance of the seamless integration of the physical and virtual worlds to realize more effective digital twin environments and introduced a digital twin framework using VR/AR technologies.

Despite the variations of the concepts addressed above, the essence of their goals and approaches is a space supporting ubiquitous communication between continuously connected physical and virtual entities in an intelligent manner, even for those focused on specific domains. For the realization of such transreality environments, where AR, UbiComp, and AI are synergistically converging, the integration of AR and IoT is essential and invaluable. Applin and Fischer [157] pointed out that the convergence of IoT and AR can better allow people to make changes to the world around them and discussed cultural implications of this blended physical-virtual reality technologies in an anthropological perspective.

Given this timely interest and need, we will describe a generalizable AR-IoT framework and interaction spaces in the following sections while emphasizing the importance for achieving more effective and efficient human-computer interactions in the future of transreality.

32.4 AR-IoT Framework and Interaction Design

Several AR-IoT frameworks have been proposed previously considering the general-purpose AR-IoT services [6, 158] and specific scenarios, such as smart home environment [80, 153]. Here, we present an AR-IoT framework that suggests three key aspects briefly mentioned in Sect. 32.1, i.e., (1) distributed and object-centric AR data management, (2) AR-IoT object-guided tracking, and (3) context-based AR interaction and content interoperability, and describe the subsequent benefits of the convergence of AR and IoT, which could introduce novel smart and physically interactive AR services. An example of such a high-level AR-IoT framework that blends in smart and interactive mixed environments is shown in Fig. 32.3. Figure 32.4 presents a potential architecture design, which describes five steps of data flow in the framework: (1) AR-IoT service providers and manufacturers upload the data for the AR-IoT products to the cloud server, which can be utilized for the user's interaction with the products, such as tracking features and services that the products can offer; (2) the local edge server near the AR-IoT clients (or users) synchronizes necessary data of the AR-IoT products, which are available in the local AR-IoT client environment with the cloud server while considering the user's profile and context; (3) each AR-IoT product in the client environment also synchronizes the data with the edge server, so that the AR-IoT clients can identify and interact with the products; (4) the users interact or control the AR-IoT products through natural and intuitive AR interfaces; and (5) the user experience, profile, and situational context are continuously updated in the edge server, so that the AR-IoT products can provide timely appropriate and relevant service to the users.

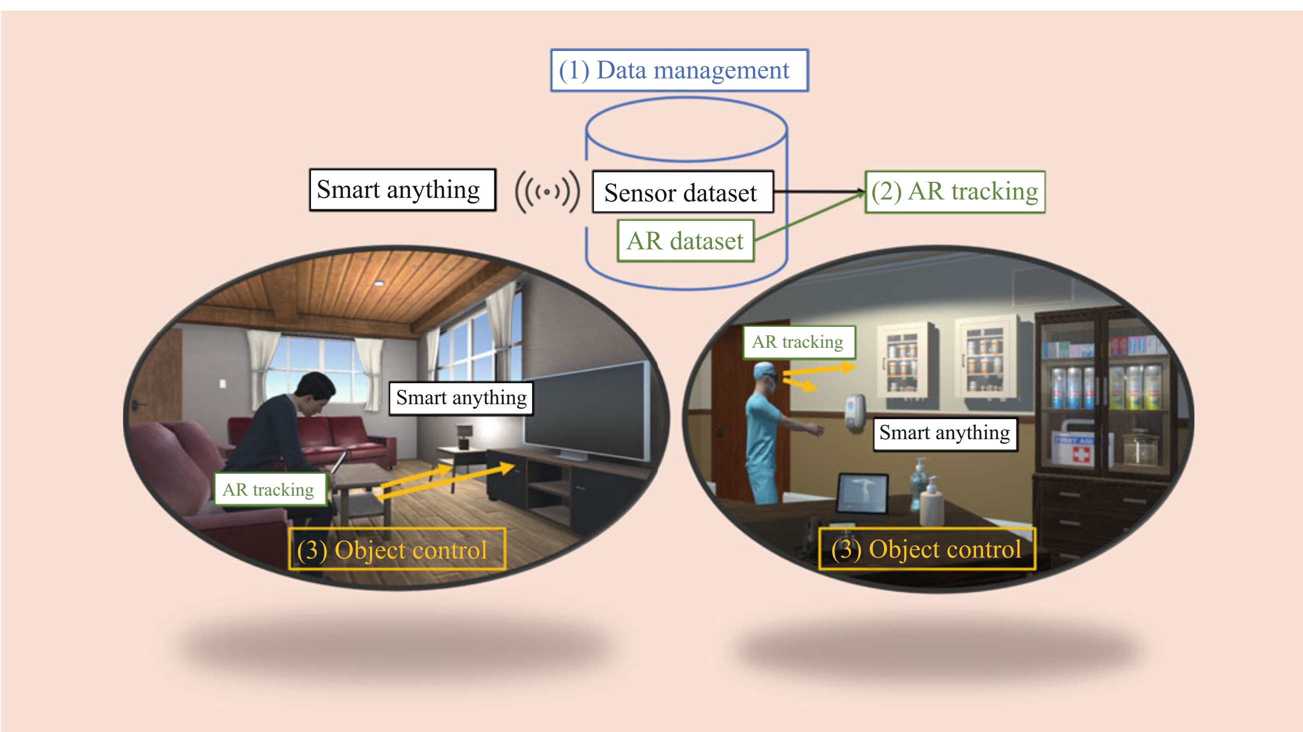


Fig. 32.3 AR-IoT framework in the transreality paradigm: smart IoT objects can store and provide AR-necessary data to the users for smart and interactive AR experiences, and *anywhere* and *anything* AR interactions will be available for the users

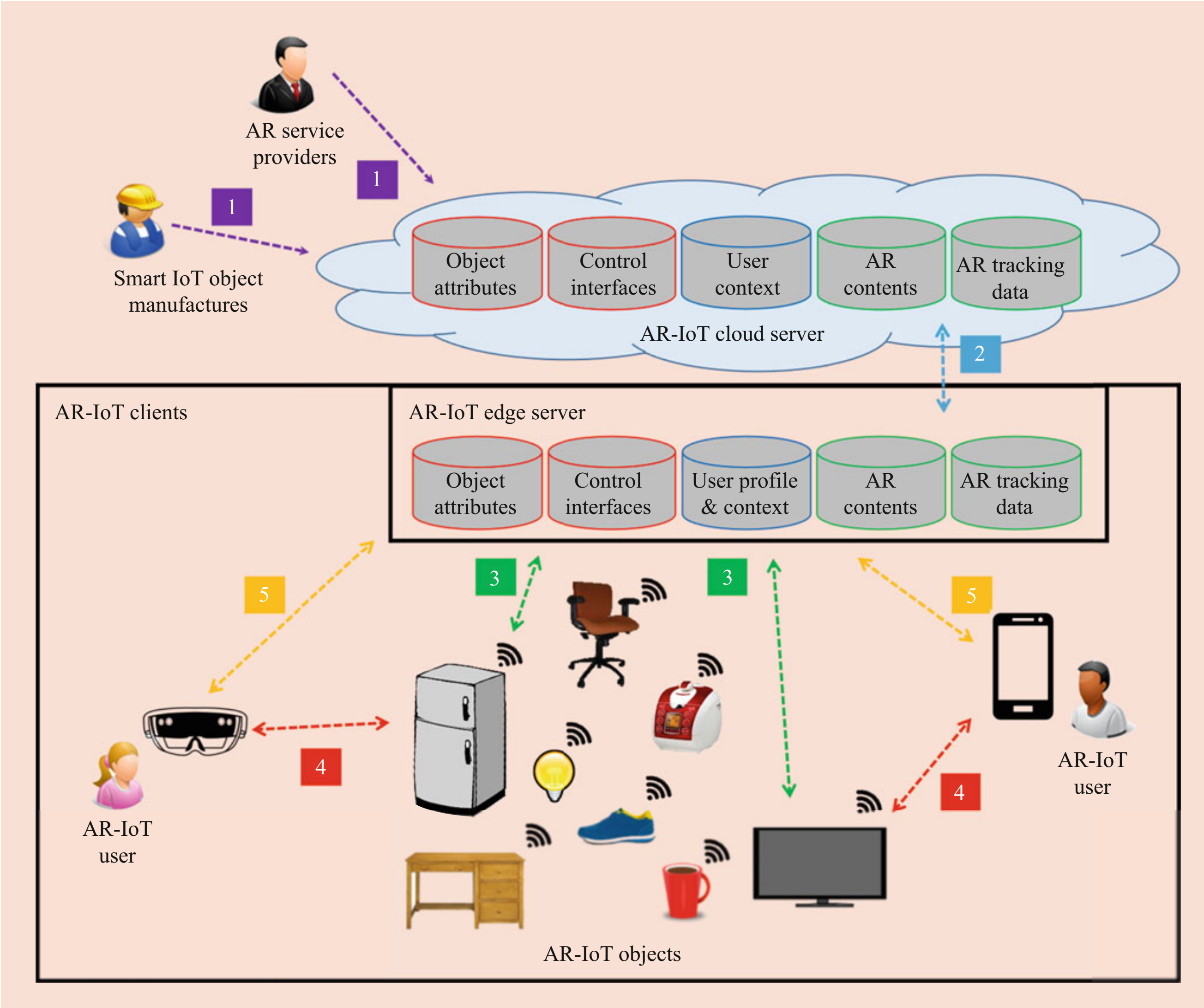


Fig. 32.4 An example of the AR-IoT architecture

The IoT objects will be accessible and interactable to the AR users in a distributed manner and can provide appropriate and necessary data for the AR experience, such as tracking or context-relevant service contents [5]. For the seamless AR experience, robust techniques for tracking the physical environment and registering virtual contents are required. The tracking process should be achieved in real time, which normally needs a high computational power, e.g., for extracting, describing, and matching 2D/3D visual features from the scene. This computational burden can be reduced using the AR-IoT framework, which can store such tracking information in distributed IoT objects and provide them to the AR clients through the network directly. Also, virtual contents provided to the users should be context-appropriate. Here the context could include the AR use cases, for example, displaying information for consumer products such as appliance control and instruction manual, or the heterogeneous AR devices that the users are using, for instance, mobile interfaces, a projection-based Spatial AR (SAR), HMDs, and even voice or other audio effects. The AR-IoT framework should be adaptive and flexible to store and provide such context-relevant contents to the users through the IoT objects. Additionally through the AR-IoT framework, the users can have the capability, in terms of efficient operation of IoT objects, to interact with the IoT devices through the AR interfaces associated with the smart environment. This is based on the help of both the IoT communication capability and the graphical AR environment. Such an AR-IoT framework can be applied to any valuable applications and services to connect everyday IoT objects and provide the user's AR experience, such as an AR manual, training, control, and instructions. It will offer comprehensive interfaces for the users to access and interact with AR entities and IoT objects that could engage the users in the shared physical-virtual space [159]. Note that this typical approach can also be realized with the filtering feature that the AR clients can identify IoT target objects nearby and augment the candidate virtual entities while obtaining context-relevant datasets from the object [15].

An example interaction mechanism that employs the AR-IoT paradigm in human-environment interactions is shown in Fig. 32.5c while comparing with conventional interactions with smart IoT objects and AR contents (Fig. 32.5a and 32.5b, respectively). We should note that a smart IoT object is made to be available in the physical environment seamlessly as a ubiquitous computer, so the users will interact with such smart objects based on the user's prior interaction experience and intuition in that environment, which sometimes conflict with the actual affordances that the smart objects can provide. AR interaction can be more informative in that aspect because the users can see additional information about the objects and the environment through augmented virtual contents, which can be updated adaptively in real time. AR-IoT interaction taking advantage of both physical IoT objects and adaptive AR contents means that the users can control their surrounding environment more effectively and efficiently using all the affordances that IoT objects and AR contents can offer [160]. Jo et al. [15] proposed an architecture for such an AR-IoT interaction combining an AR interface with the IoT for shopping services. They developed a proof-of-concept prototype of the AR framework tested on IoT lamps and an in situ interaction method to support control directly with the IoT object.

In the following sections, we will describe the three key aspects of the AR-IoT in detail while drawing a clearer picture of what the IoT-enabled AR of the future will look like, how it will operate, and what it will be capable of providing.

32.4.1 Object-Centric AR-IoT Data Management

AR-IoT frameworks and services along with dataset management for everyday objects are described briefly in this section. Additionally, we summarize several architectures, data processes, data structures, and content representation for the physical objects to interact with AR in the published research. AR-IoT services commonly need to manage generic/specific data and service content for their constituent objects or augmentation targets, for example, visual features for AR recognition and tracking, virtual content and information about the object itself, control interface, and organized additional contents for the operation. Having IoT objects that can communicate with the AR clients in the environment can allow the users to access such data and content through the AR-IoT interaction. The communication and information exchange between the AR client and the IoT objects can occur directly between them as well as through the regional IoT server [5]. Herein, we review current approaches to manage such physical object data for the use of AR, e.g., architecture and data handling, and discuss how they can be extended and scaled to the level of IoT [15].

To interact with real objects using AR interfaces in the early days, people were concerned with an AR framework that was capable of executing ubiquitous communication between physical objects and the AR device [161]. Recent works have attempted to deal with mapping the sensor-object relationship and filtering approaches to reduce the search space in the near space [5, 162]. Specifically, in one notable framework with respect to AR, Iglesias et al. [163] suggested an intelligent selection of resources by the user's attributes, user-object proximity, relative orientation, resource visibility, and AR interaction connecting the object. They developed an object browser based on AR with context-aware representation of resources. Similarly, Ajanki et al. [164] constructed an AR interface with contextual information and defined context-sensitive virtual information about people, offices, and artifacts. They suggested a filtered AR concept to visualize selected information about teaching and research project for visitors in a university.

Figure 32.6 shows the possible data management/process flow for AR-IoT frameworks based on the physical objects, considering AR datasets with scalability. The AR service users can retrieve fiducial visual markers or natural features from the surrounding objects, which are necessary for the relationship between the physical and virtual objects to be defined in advance. The users, holding their AR devices, are able to view the filtered AR objects, which are corresponding to the nearby IoT-capable objects based on their relative distance or direction from the user. Then, the client AR system directly receives the "feature" and "content" information for the object with the attached sensor. The AR users can experience an efficient transreality environment (e.g., IoT control interface) to mix the virtual object by donning an HMD or using a mobile phone with an attached camera module [5].

The *cloud* as a computing resource has been investigated in many studies considering the pervasive and ubiquitous AR experience [165]. The use of the cloud is beneficial for computation time to reduce the heavy work by matching the large quantity of features for the poor computing capability of mobile devices. Recent research has focused on the process to register and manage AR datasets with a cloud computing device, so they are concerned with how to fit tracking information and AR presentation datasets from the remote cloud server. For example, the users receive the collected AR attributes and tracking information, which are shared via the cloud server, to access and interact with the surrounding objects. Then, the AR users, who are only with

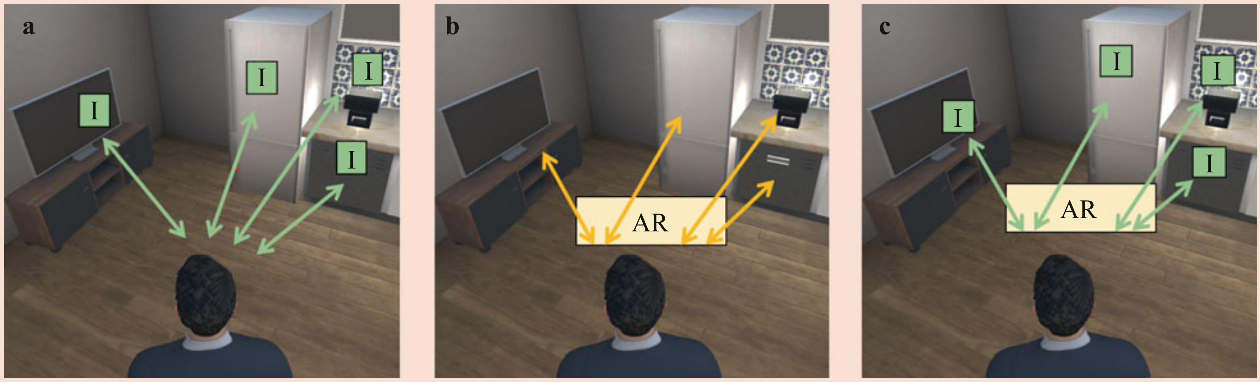


Fig. 32.5 Illustration of different interaction paradigms with smart objects: (**a**) ubiquitous computer interaction, (**b**) AR interaction, and (**c**) AR-IoT interaction. The *I* is used to indicate smart IoT objects

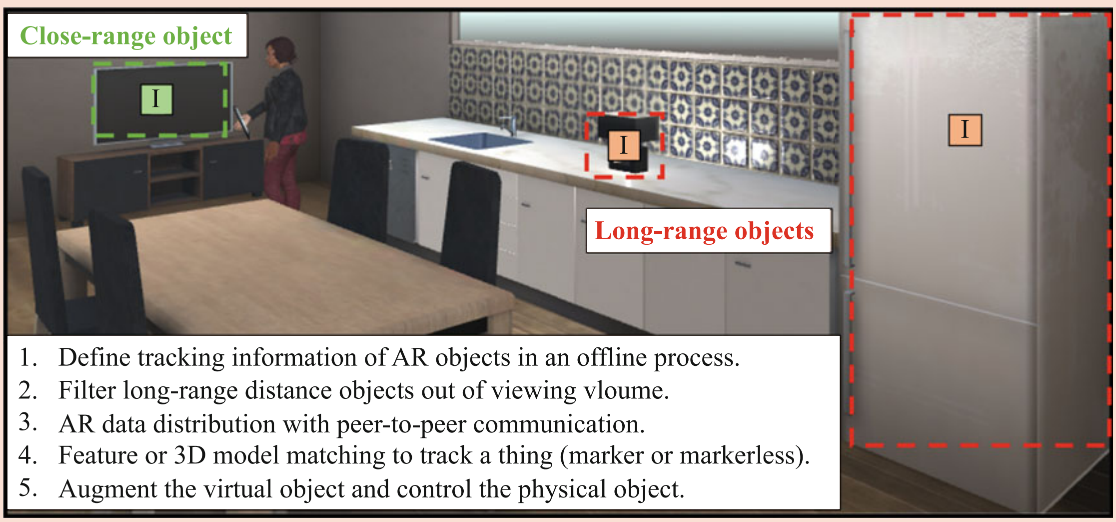


Fig. 32.6 Distributed data management scheme for "anywhere" AR services and interactions with "anything," i.e., physical IoT objects. The letter *I* is used to indicate *IoT objects*

the AR browser program without tracking information and AR presentation datasets in the device, can easily connect all of the things with the prebuilt relationship mapping among the physical objects and relevant virtual contents.

A more recent trend is to improve the way to use the adjacent computing resources in the user's surroundings rather than enabling computing services on the end of the network at a long distance, such as a remote server or the cloud [5]. It will still be difficult for cloud services to support scalability to the level of "everywhere." An alternative may be to connect to a single area server, which only covers a particular local area, e.g., a room or home, while managing a limited number of objects [165]. The adjacent computing approach can be used to solve problems such as a bottleneck assignment and detection of moving objects by a remote server. This approach is similar to fog computing architecture in the domain of a sensor network that emphasizes latency reduction with high-quality service and handles datasets at the network edge [166–169]. The AR users can connect directly to the objects in the surrounding area because the sensors attached to the object can detect it in real time to consider the user's position in certain ranges [166].

Some research has proposed details of the content structure for resource management as another issue to provide AR datasets on the Web-based platforms. In some ways, the contexts on the Web-based platforms are similar to the AR environment with IoT services and content, where most IoT services are implemented as applications. One promising direction is to use the Web to support interactions with physical objects, as exemplified by Google's Physical Web [170].

Here, the smart objects have particular URLs and can exhibit their own dynamic and cross-platform contents, represented in standard languages such as HTML and JavaScript. We can envision a future where various types of IoT services, including even AR, will be available under a unified Web framework, that is, the "webization" of things. Ahn and coworkers, for example, presented a content structure as an extension to HTML5 for building webized mobile AR applications [171]. This allows a physical object to be referenced and associated with its virtual counterpart. Kim et al. [172] also presented an AR content structure for building mobile AR applications in HTML5, as on the Web. They used an extended presentation logic of HTML to apply current Web architecture and a referencing method with matching between physical and virtual resources. As a similar AR data structure, Muller et al. [173] introduced a custom XML-based format to define AR manual structures for home appliances. Considering the concept of our AR-IoT framework, we present an example of associating a smart window (physical object) with virtual objects (augmentation) about weather information in Fig. 32.7. We augment the smart window with sensor data fed from physical weather sensor stations while receiving the related information from the Web at the same time.

In our AR-IoT framework for transreality, there are many different types of IoT devices in the user's environment, so we should consider the characteristics of AR contents according to the related IoT devices. This concept is similar to the website components that have the different configurations depending on the platforms, such as in mobile or desktop computing devices. Thus, to make the IoT-enabled AR platform to be naturally applied anywhere, depending on the nature of the IoT device, it should adaptively control the degree of AR content representations.

32.4.2 Scalable AR-IoT Recognition and Tracking

AR research primarily focuses on overlaying a virtual 3D model, that is, on how to realistically integrate augmentation with the real world. The most important technical challenges for AR would be fast, accurate, and stable recognition and spatial tracking and registration (or pose estimation) of objects in the environment. AR tracking technologies using the context of physical objects have gradually increased not only for the robust tracking quality but also for the efficient data management and the scalability of AR experience [130].

There are many research works on AR tracking methods that use computer vision algorithms and data from distributed sensors in the environment (see Fig. 32.8). Claros et al. proposed a fiducial marker-based AR medical system using a wireless sensor network (WSN) to monitor real-time information with collected biometric signals from patients [175]. The WSN was used to process semantic information collected from distributed sensors and a marker ID overlapped with the real world to visualize perceptual information (e.g., temperature and humidity). Mihara et al. implemented a light-emitting diode (LED) AR marker with the procedure of reading LED blink patterns attached to a TV rather than fiducial markers [176]. Despite the ease in using the fiducial markers or pattern IDs, one of the problems with them is that they must have individual markers corresponding to each physical object, which makes the scalable AR experience difficult. For example, if there are a large number of objects in the environment around the AR user, the same or even more number of markers will be needed, which could make the environment unnatural and cluttered. Thus, scaling such marker-based methods for millions of

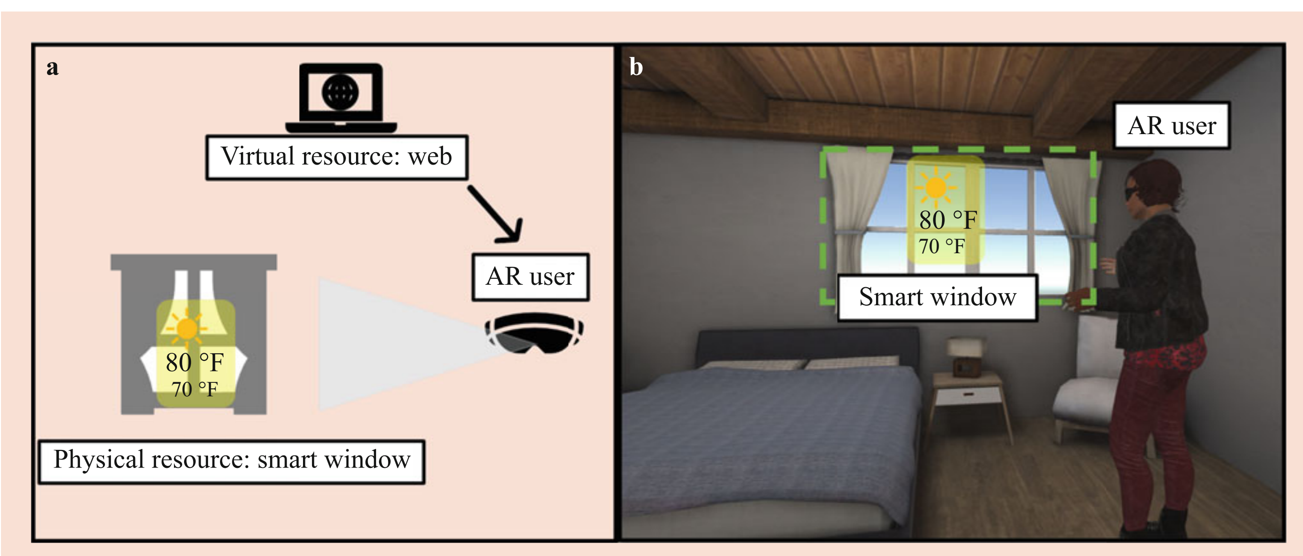


Fig. 32.7 Webized AR content representation in which virtual data are associated with a Web-accessible physical resource [170]: (**a**) virtual and physical resources of webized AR content and (**b**) an example associating a physical sensor dataset [174]

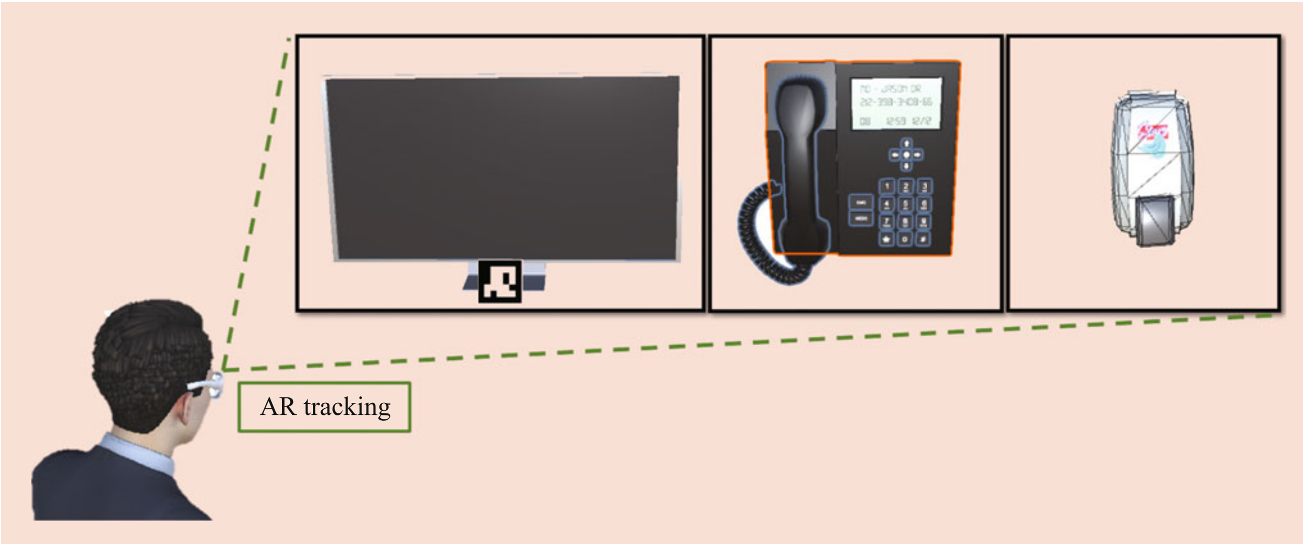


Fig. 32.8 Computer vision-based object recognition and tracking solutions for AR using (**a**) a fiducial marker, (**b**) natural features, and (**c**) a 3D model-based approach

IoT objects would be even more difficult – either the level of accuracy or the capability of real-time response is likely to suffer. Placing and attaching markers to thousands of everyday objects also is not a practical solution. Additionally, because there is no single universal recognition and tracking method to cover all types of objects, a multitude of algorithms should be used collectively. Therefore, in a typical situation, before the object is recognized, one cannot determine a priori about which algorithm would be the best to apply to its recognition in the first place. Also, all algorithms must be attempted exhaustively, which will again result in significant latency.

Another popular tracking technique is a natural feature-based method that identifies an object and computes its pose using primitive geometric visual features and properties of the object detected by a camera sensor [177]. Different feature-based techniques vary in their robustness and may be applicable to certain classes of objects. Such feature-based methods require a high volume of features (and data) and a strong computing power to establish a robust feature extraction and matching and often require a preliminary learning phase to handle difficult matching conditions, for example, an angled view, dark lighting, and occlusions [178].

In case the target object has few visual features, e.g., textureless objects, template image matching approach is often used; however, this method has some disadvantages for use in robust 3D tracking [164]. Different approaches to solve the problem of tracking textureless 3D objects have been developed for certain situations, such as under poor lighting conditions, during partial occlusions, and against cluttered backgrounds, but the quality of tracking still remains relatively low [178]. Some of the researchers proposed 3D model-based tracking methods that attempt to recognize and track target objects by matching and fitting a 3D wireframe model to the edges extracted from a camera image [179]. However, such model-based approaches also have drawbacks, for example, the methods require a good initial solution with a lengthy convergence time, and it is not clear which reference 3D wireframe model would be the most suitable. Despite such complications, they remain viable alternatives, given the absence or lack of feature information in some specific scenarios.

After all for the robust and visually unobtrusive tracking method, the currently dominant AR tracking is based on simultaneous localization and mapping (SLAM) technique, which denotes the computational technique that creates and updates a map of an unknown space where the user is located, while simultaneously tracking their location in it [3]. SLAM was originally proposed in the field of robotics but now is known as a great alternative to traditional AR tracking approaches because it avoids the necessity for prior information, such as reference images or 3D models. Apparently, SLAM has overcome many limitations that the previous tracking methods had and has been enhanced in terms of the robustness of tracking over the past 10 years, although there are still typical limitations for SLAM, such as a high computational cost to deal with tracking and mapping simultaneously and tracking loss caused by fast camera motions.

Our proposed AR-IoT framework with multiple network connections in the transreality environments will address this scalability issue using the distributed and object-oriented manner.

32.4.3 Context-Based AR-IoT Interactions and Content Interoperability

The use of the AR-IoT framework can also enhance the intuitiveness of physical object attributes and interfaces among a large amount of objects and data in the environment by providing context-relevant visual clues that can guide possible physical-virtual interactions. The most prevalent IoT applications often involve object control interfaces in situ or remote scenarios. IoT objects consist of sensors for incoming data, networking modules for wireless capability, and actuators to control the object's functionality [179]. Expressly, actuators can operate with the user's interaction, such as the decision of object properties, e.g., turning on/off lights or opening/closing doors.

Researchers presented interaction approaches to control and manipulate the physical world through digital interfaces or virtual replicas. Jackson et al. showed the operation method for such object behaviors, e.g., controlling electronic devices or the states of physical objects, in a simulated environment for home automation [180]. A few research papers have also shown different mechanisms or visualizations of interconnected simulation results with sensors and/or actuators embedded in an everyday environment. For example, Lifton and Paradiso presented a dual reality system that generates an interplay between the simulated world and a bunch of sensors, such as an electrical power strip in the real world [148]. In their simulated dual reality, the users can explore a variety of experiences interacting with both physical and virtual objects; this opportunity is concerned with not only mutually reflected sensor browsing but also interaction techniques with the interconnected sensor through sensor/actuator networking. Lu also proposed a bidirectional mapping technique for IoT-enhanced information visualization while presenting a system to realize eco-feedback for energy saving [181]. When the user turns on a home appliance, such as a TV, in the real environment, for example, the attributes of the deployed sensor to detect the user's activity are transmitted to the simulated world. Then, the monitored virtual world can generate a counterpart representation in the real world.

In the scope of AR research and applications, some also focused on such an AR simulation for the control of the object's function and usability measurement. For example, AR can be used to visualize simulations of applied control for previewing or training purposes [182]. Since AR can provide intuitive and immediate virtual information of in situ local objects and even for remote objects, e.g., using a remote-controlled camera, AR is an excellent visualization method for object control considering user's context [5]. With recent advances of AR technology, there have been a few attempts to merge the two, i.e., using AR as the control and simulation interface for IoT objects. Researchers have tried direct control for everyday objects in the AR environment. Rekimoto and Ayatsuka proposed a visual tagging system called *CyberCode*, which is based on an AR 2D barcode to identify and detect objects [183]. The system has an operation mode to manipulate physical objects, in which the user performs a natural interaction among the target objects, for example, "drag and drop" from one object to another target object after selecting the manipulating object. Then, the target object would carry out the particular operation with context-relevant information in the first selected manipulating object. Müller et al. also suggested an AR manual to convey step-by-step instructions [173]. In the AR manual system, they defined a user markup manual language (UMML) file to generate sequential operations with corresponding steps. Greg Tran presented an AR 3D architecture system based on contextual relationships with real geometry for interrelationships between digital and physical objects [184] while describing "Mediating Mediums," which explores the future of mixed reality and the relationship with physical form and environments, particularly about architecture. The system could provide simulated geometries that are projected into videos of physical spaces. Consequently, in the near future, IoT objects will connect with each other, and AR users in the operation environments will be able to intuitively manipulate the context of objects, with immediate in situ support from intelligent AR-IoT technologies.

Moreover, it is important to note that the user experience and perception in the AR-IoT interactions will depend vastly on the AR devices or displays. Kruijff el al. provided a classification of perceptual issues in AR and suggested predominant issues for a specific device (e.g., head-worn display, handheld mobile device, projector-camera system) [185]. Most previous works mainly used smartphones to provide images that synthesize real and virtual environments, but they did not consider the presentation of synthesized images directly to the human eye. More recently, AR devices with a helmet-type HMD (or head-worn display) that synthesizes spatially registered virtual objects overlaying a user's view have been introduced and becoming popular. These helmet-type AR devices are mainly divided into optical and video see-through HMDs, depending on whether actual images are viewed directly by the user or via a video input. For example, the optical see-through HMDs provide the means to present AR visual information to users normally using an additive light model approach, e.g., by projecting light onto a surface which is then reflected into the user's eyes [186], whereas the video see-through HMDs capture the real environment scene through camera modules and render the processed imagery with superimposed virtual contents. While different types and form factors of AR devices are proposed and introduced, the interaction mechanisms between the users and the environment or among the physical and virtual content should be adapted and diversified depending on the context in AR-IoT-enabled transreality. In that sense, we aim at emphasizing the importance of a framework for AR-IoT interactions considering the object characteristics and the impact of various kinds of AR devices.

32.4.4 AR-IoT Framework Evaluation

To evaluate the proposed AR-IoT framework, here we develop a proof-of-concept implementation of the framework and experimentally assess the user's satisfaction level and usability through a couple of preliminary user studies in a shopping service context [15]. The underlying assumption is that our proposed AR-IoT interaction approach would be useful and well-received and create more effective user experience compared to the traditional GUI-based approach.

We developed a proof-of-concept prototype incorporating IoT-enabled smart clocks and lamps (see Fig. 32.9). We used a Raspberry Pi 3 Model B (RPi) and had beacons integrated into the smart clocks and lamps to pose them. The RPi board is equipped with a quad-core 1.2-GHz 64-bit CPU, a 1-GB RAM, a 100 Base Ethernet, 4 USB 2.0 ports, a HDMI port, and a MicroSD port and has a small storage capacity and wireless Internet communication (BCM43438). The AR client was implemented on a smartphone that connects to the IoT products directly through the beacons. The proposed AR-IoT framework was applied to this prototype, which means in addition to storing generic data and virtual content in the IoT objects, individual and different-typed IoT objects of interest in the vicinity of the AR client can communicate the information for the users to recognize and track the objects, e.g., broadcasting the tracking features, algorithm type, and current physical state of the product, such as location and distance to the client or other companion reference objects. The datasets for AR features were extracted from the images of the IoT objects for recognition and tracking, and the TCP/IP protocol was used for the communication to exchange data between the AR client and IoT devices.

For the studies, we prepared a situated environment mimicking a shopping center, where participants could receive a list of available IoT-enabled smart objects nearby when they entered the environment and see the AR-capable objects, which were filtered based on distances between the participant and the object. The AR client system, which the participant used, received information about the smart objects, such as AR tracking and control interfaces (e.g., buttons). Based on the communicated information, the object-relevant virtual contents and control interface were spatially overlaid on the target product in the AR-IoT interaction condition, whereas traditional 2D GUI on the smartphone screen was used in the control condition. To develop control UI menus and AR contents, we used Unity3D C# scripting language and PTC Vuforia AR tracking engine.

The first experiment looked at the level of user's satisfaction by comparing two different types of interfaces: (a)

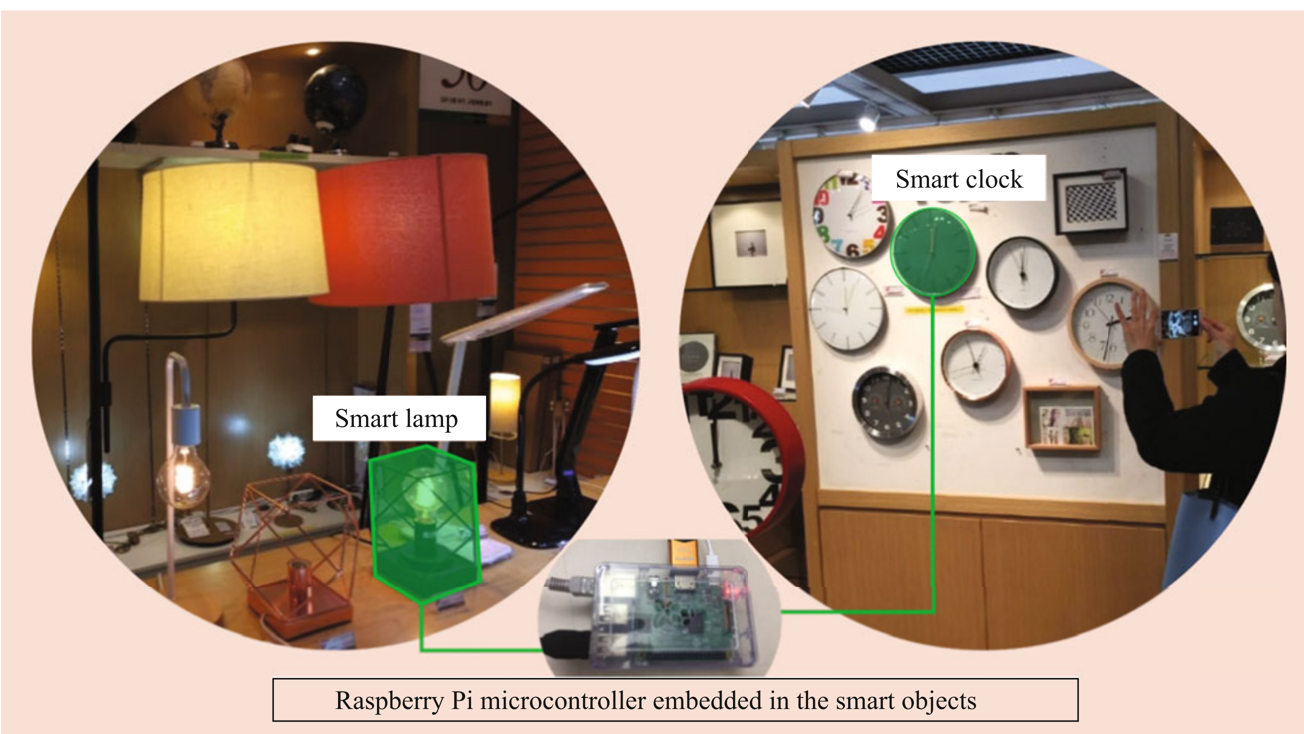


Fig. 32.9 AR-IoT object prototypes for a proof-of-concept study: a smart clock and a smart lamp with Raspberry Pi embedded for IoT features

Table 32.1 Usability questionnaire measuring the ease of use, naturalness, fatigue, task completion speed, and preference

Measure	Question
Q1 (Ease-of-use)	How accessible did you find the interface to be? (1: very difficult – 7: very easy)
Q2 (Naturalness)	How intuitive and natural did you find the interface to be? (1: very contrived – 7: very natural)
Q3 (Fatigue)	How fatigued were you after using the interface? (1: very fatigued – 7: not fatigued at all)
Q4 (Speed)	How fast did you feel you were able to complete the task? (1: very slowly – 7: very fast)
Q5 (Preference)	How did you like the interface that you used? (1: don't like at all – 7: very like)

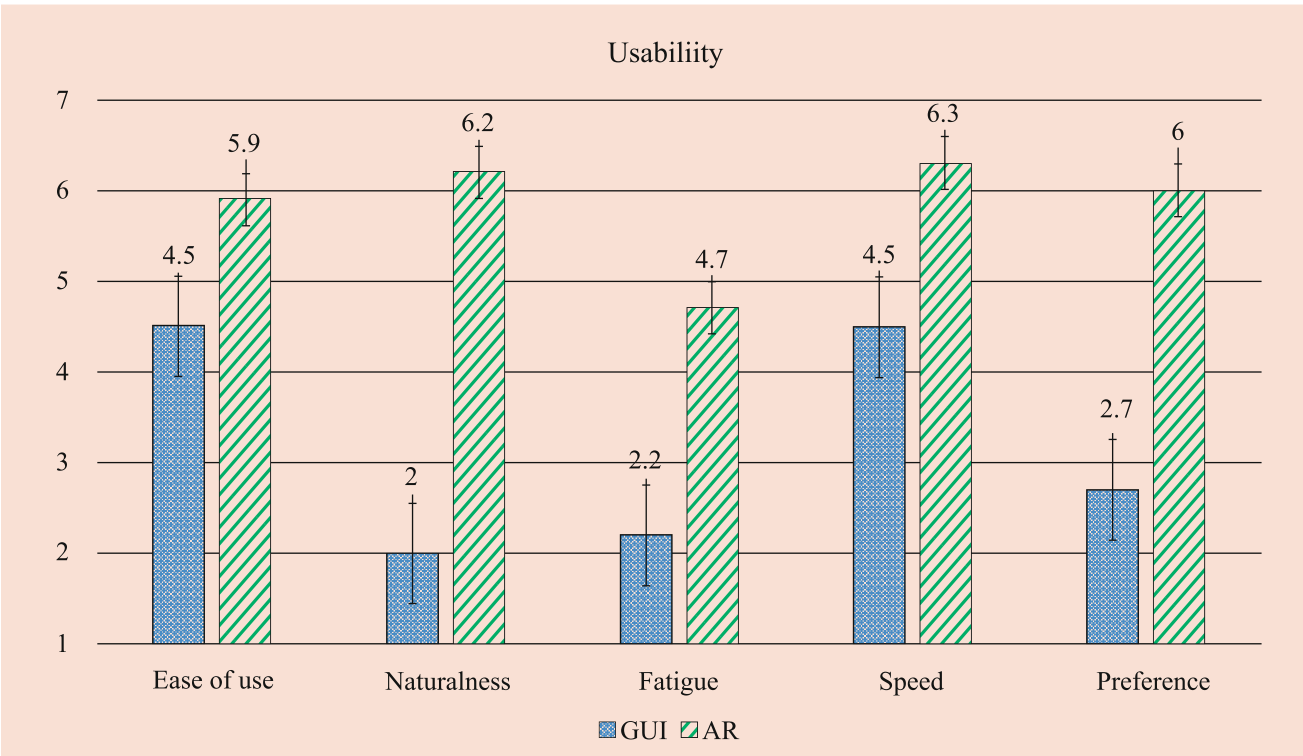


Fig. 32.10 Usability evaluation results

conventional Web-based and (b) AR-based. Ten participants (age $M = 36.0$, $SD = 7.5$; female: 2) experienced both interfaces in a within-subject design and answered a question asking their general satisfaction with the interfaces in a 7-point Likert scale. The satisfaction scores were analyzed through the nonparametric Wilcoxon test for paired samples, which revealed that the mean satisfaction score was significantly higher ($Z = 2.871$, $p < 0.05$) with the AR-based interface ($M = 6.2$, $SD = 0.8$) than the conventional interface ($M = 3.5$, $SD = 1.2$).

The second experiment investigated the usability of the AR-IoT interface. Similar to the first experiment, 16 participants (age $M = 37.0$, $SD = 6.6$; female: 4) were asked to switch the IoT-enabled lamps on and off using (1) conventional GUI using the touch screen on a smartphone and (2) AR-based interface with spatial registration of virtual GUIs to the target IoT device. Participants performed the light control tasks multiple times with different IoT lamps and were asked to answer a usability survey asking their perception of the interaction and the performance in a 7-point Likert scale, such as ease of use, naturalness, fatigue, speed, and simple preference (see Table 32.1). The result showed that the participants reported higher scores for the AR-based interface in all the usability categories compared to the conventional GUI (see Fig. 32.10).

We also evaluated the error rates in the control of the IoT devices between the two interfaces, e.g., the number of incorrect selection and operation of the device. The results showed that the participants made 32 errors in average (out of 80 trials) for the GUI-based condition, while they only made 3 errors in average for the AR-based condition, which suggests that the participants took advantages of the spatial and visual affordance of the AR interface for more intuitive and direct object control in the physical environment.

32.5 Opportunities of Embodied Interactions in AR-IoT Environments

While we described the concept of transreality and the convergence of AR-IoT and summarized the related research in previous sections, we covered various examples of 2D/3D user interfaces – mostly for controlling the physical IoT objects or the environments. The most common interaction style is to use typical GUIs with Windows, Icons, Menus, and Pointers (WIMP) on 2D screens, e.g., PCs, smartphones, and tablets. Involving reliable and effective touchscreen techniques, the affordances of the interaction have been increased in more "natural" ways, such as selecting the target object by tapping the screen or an intuitive drag and drop [183, 187]. In addition, while VR/AR technology is considered as a valid platform for such interactions with IoT objects, the consideration of spatial context, e.g., searching or navigating the potential IoT objects in the environment and interacting with other adjacent objects, becomes more and more important [188]. Some examples of such AR-IoT interaction mechanisms are illustrated in Fig. 32.11.

In particular, as immersive wearable AR displays like smart glasses and see-through HMDs are grabbing the public's attention and becoming more and more popular, a variety of 3D user interfaces and interaction techniques using ad hoc VR/AR controllers or natural body gestures, which are traditionally researched in the VR/AR community for decades [189], can be applied to the AR-IoT interactions considering the spatial environment context. Recent research by Lages and Bowman [190] presented an adaptive AR interface that can change the position and the form depending on the environment while the user is walking. For example, a set of virtual windows in the AR environment could follow the user, rotate themselves, and automatically align to the wall for better user experience in the AR interactions. They conducted a user study that investigated how such contextual adaptations can contribute to the system usability and the user's behavior and revealed that participants particularly appreciated the ability of the adaptive interface to automatically follow and position the information to the user's view. Such intelligent and context-adaptive AR interactions are one of the key elements to achieve more effective and efficient AR-IoT interactions among the users and the physical and virtual objects in the transreality environment.

While considering the intelligent and context-adaptive AR interactions, we see unique opportunities of embodied interactions through virtual avatars and agents with visual appearances in AR-IoT environments. In the remainder of this section, we will first present some of the affordances of such virtual avatar/agent-based embodied interactions in the AR-IoT transreality paradigm, give a few examples of prototypes that utilized the idea of AR-IoT for their intelligent agents, present findings on fundamental research on embodied AR agents that can shape the realization of embodied agents in the AR-IoT transreality paradigm, and describe two use cases where the intelligent embodied virtual agents can be adapted to enhance the human-computer interaction experience in our daily lives and professional settings.

32

32.5.1 Embodied Interactions in AR-IoT Environments

As we described before in Sect. 32.3 regarding the transreality paradigm our daily lives of the future will become more complex with large volumes of heterogeneous data and information from a variety of sources such as millions of smart objects [59, 62]. The importance of pervasive AI and ubiquitous IoT objects will continue to grow, and consequently, the research on the efficiency of communication

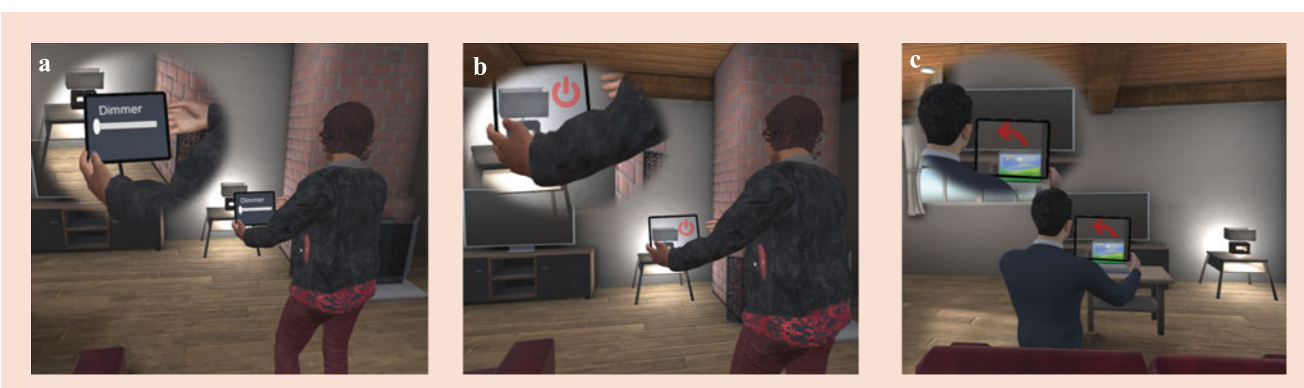


Fig. 32.11 Illustrations of various interactions with IoT objects: (**a**) an in situ/remote operation with traditional graphical user interface (GUI) button, (**b**) an AR interface that could recognize the target object and present appropriate GUIs, (**c**) a metaphorical natural interaction (e.g., drag and drop) to invoke an object function interacting in a virtual/augmented space to affect the physical world

with them and trust in the validity of information conveyed will be important. Embodied interactions in AR-IoT environments can go beyond users controlling IoT-enabled devices via 2D/3D interfaces, which were made available to them through AR devices, by extending to multimodal interactions through embodied virtual avatars and agents in AR and facilitating physical-virtual interactions in the AR-IoT transreality paradigm – in particular considering the efficiency and trust in the interactions with the smart IoT objects and physical environments.

Nonverbal behavior plays a crucial function in regulating communication and providing critical social and contextual information [191]. For example, some nonverbal behaviors, such as pointing and gesturing, can directly convey information that would otherwise require significant verbal explanation. This is true for both seeing and carrying out such behaviors. Socially intelligent nonverbal behaviors in the embodied interactions can convey situational and information awareness, which can be critical for establishing common ground in communication. For example, if a user is looking around a scene, information conveyed is assumed to be relevant to that scene, without any verbal explanation. Furthermore, information conveyed about the scene will likely carry more weight (increase trust), as the provider will be perceived as having direct awareness of the events/objects in the scene. Nonverbal communication can also be used to increase the transparency of the confidence in, and even the provenance of, the information, which in turn can increase trust and efficiency [192].

In the vision of pervasive connected AR environments (i.e., the transreality paradigm), humans and virtual entities are aware, can seamlessly interact with, and influence each other through different modalities, suggesting a bidirectional relationship between the physical and virtual worlds. For interactions to be seamless, virtual entities are envisioned to capture the features of their real counterparts, such as appearance and behavior, and exhibit awareness and understanding of their environment. In such circumstances, the visual embodiment and nonverbal behaviors of the virtual entities mean the increase of the chances for successful communication among the users and the AI or smart objects/environment through the richer and efficient communication channels. Here, we describe some of the findings from our own research efforts in the past and other relevant works that emphasize the potential benefits of embodied interactions assuming seamless physical-virtual interactivities are available in transreality with the convergence of AR-IoT.

32.5.2 Embodied AR-IoT Agent Prototypes

To our knowledge, we have only found a few prototypes where researchers utilized embodied AR agents as a means to facilitate interactions with smart objects. In the context of embodied IVAs capable of controlling smart objects, Amores et al. [193] introduced the idea of an embodied AR agent that can change the state of a smart lamp, for example, the agent walks toward the lamp and switches it on or off. Kim et al. [143] further extended this idea to investigate the influence of the embodiment and locomotion behavior in more detail, where these factors were varied in a human-agent interaction scenario inspired by common use cases of voice-based assistants, such as Amazon Alexa. The embodied AR agent was presented to the users through and optical see-through HMD (see Fig. 32.12a) and extended the contribution of Haesler et al. [194] by exploring a more diverse interaction scenario. In the interaction scenario studied by Kim et al. [143], some tasks involved the virtual agent controlling physical objects in the same room as the participant, while others presented opportunities where the embodied agent were asked to control physical objects or gain/relay information to and from the physical environment

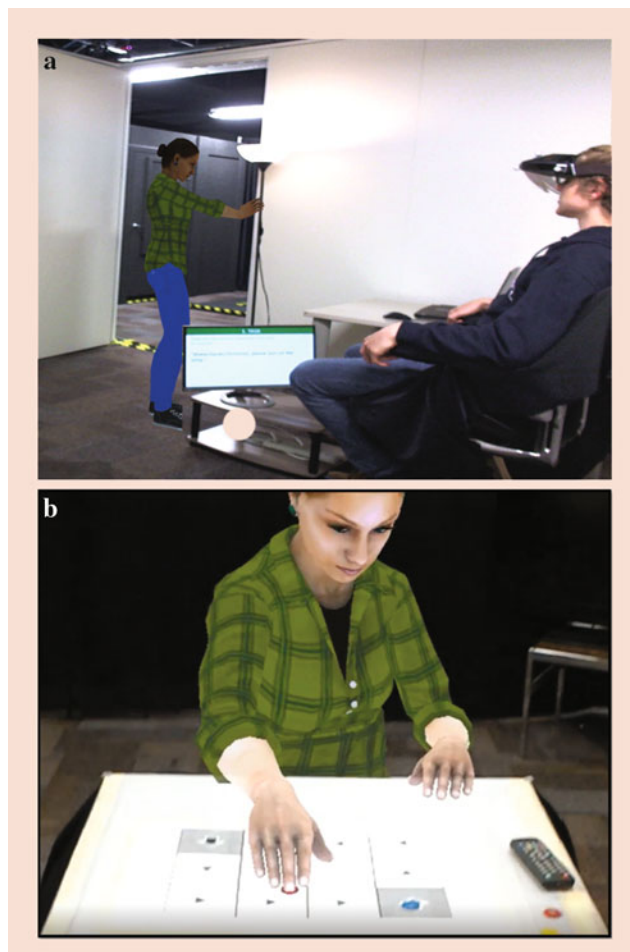


Fig. 32.12 Virtual humans interacting with the physical objects: (**a**) a virtual human turning on a floor lamp using IoT-enabled light bulb and (**b**) a virtual human moving a physical token in a board game scenario

(i.e., objects and people). Additionally, some tasks revolved around the concept of user's privacy needs and their level of inclination to share their private information with the agent. The task diversity in this scenario allowed for various opportunities where the effects of agent's embodiment and locomotion behavior were explored, such as the embodied AR agent walking to a physical lamp in the room to turn it on, compared to having no locomotion behavior and no embodiment conditions. They found that the embodied virtual agent with abilities to be aware of its surrounding environment and physically influence it could positively impact the participants' sense of privacy-preserving and their confidence in the agent's activities and abilities.

32.5.3 Embodied AR Agent Insights

Advances in AR technology have facilitated increasing research on the development and understanding of embodied AR agents. As embodied AR-IoT interactions evolve, such findings can shape how AR-IoT agents are realized and understood. One area that has received an increasing attention is embodied AR agent's physical-virtual interactivity [195]. Lee et al. [196, 197] studied the effects of a virtual human's ability to move a physical token on participants' perceptions of their interaction in a table top game setup. They found that participants felt more co-present with the virtual human when she moved a real game token compared to moving a virtual one, and interestingly, this effect changed their overall perception of the virtual human's ability with regard to moving/affecting other physical objects (see Fig. 32.12b).

In consideration of the effects of tactile physical-virtual interactivity in the embodied interactions in AR/MR, Lee et al. [198] developed an AR physical-virtual table, where a virtual human and a participant occupied the virtual and physical ends of the table. In one condition, leaning on either side of the table resulted in a wobble on the other side as if both physical and virtual sides of the table were seamlessly connected, while it's not the case in the other condition. They found that participants felt more socially present with the virtual human when the table wobbled compared to the condition that it did not. In another study where participants shared a runway with a virtual human, Lee et al. [199] simulated the effects of the virtual human's footsteps when walking and jumping using vibrotactile feedback (see Fig. 32.13). They found that participants experienced a higher sense of co-presence (i.e., sense of being together) with the virtual human and perceived it as more physical compared to conditions where this effect was absent. Such vibrotactile feedback in these examples can be easily achieved in smart environments with IoT objects embedded to improve the user experience of the interaction with virtual entities in AR-IoT environments.

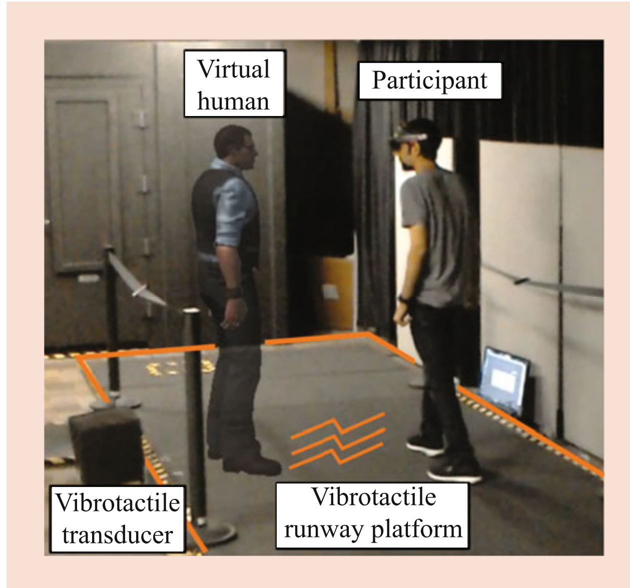


Fig. 32.13 A vibrotactile platform for generating the sensation of a virtual human's footsteps on the floor when walking and jumping

Similar to many real-world experiences, sometimes the changes in the environment are more subtle and only affecting the ambience of the main interaction. As physical entities, humans are used to detecting and responding to such changes, for example, airflow or environmental noise; thus, it is important to understand the influence of similar behaviors when portrayed by virtual entities. Kim et al. [200, 201] investigated how air blowing from a real fan, which moves virtual papers and attracts the attention of a virtual human, influences participants' sense of co-presence with the virtual human through different AR setups. They found that participants felt more co-present with the virtual human when the virtual papers and curtains fluttered due to the real airflow and the virtual human exhibited awareness of these events, by holding onto the fluttering paper and looking toward the fan compared to a condition where virtual entities were unaffected by the real fan blowing.

Advances in AR technology and continuous research efforts, such as those mentioned above, are paving the way for embodied interaction of virtual humans in various fields of application, such as intelligent virtual assistants, caregivers, and collaborators, allowing for more engaging experiences.

Kim et al. [202] and Wang et al. [203] explored the influence of an agent's embodiment in collaborative problem-solving scenarios. Comparing embodied virtual human assistance, with voice assistance, and no assistance in a desert survival task, Kim et al. [202] found that receiving assistance, regardless of form (i.e., embodied and voice-only), enhanced participants' performance. However, the embodied assistance provided a richer experience and less task load compared with voice assistance. Wang et al. [203] also varied the agent's embodiment and appearance in a collaborative search task, where participants worked with the agent and asked for hints to find hidden objects. Their findings showed that participants gazed more at the humanoid agents compared to the one that looked like a virtual Alexa. Healthcare is another important area that can benefit from intelligent agents and connectivity of objects. Kim et al. [142] designed scenarios involving health-related and daily life activities, comparing real and virtual human assistants in embodied and voice-only forms. Their results indicated that the virtual counterparts are not at the point of creating a similar experience to a real human; however, both embodied interactions created a more engaging experience than their voice-only counterparts.

Although virtual humans have been researched more extensively than other form factors, they are not the only entities capable of influencing users' perceptions. Virtual animals have also been researched in a number of domains, with encouraging findings that suggest their potentials for roles involving physical-virtual connectivity (see Fig. 32.14).

Johnsen and Ahn et al. [204] developed an AR setup where children interacted with a virtual pet dog and their exercise levels influenced the state of the virtual dog and the tricks it could learn. For instance, the pet became more fit as the child exercised more. Compared to a goal-based interface without the virtual dog, children's exercise levels significantly increased when interacting with the virtual dog. Norouzi et al. [205] gave participants a virtual pet dog and created a scenario where another person walked over the participant's virtual dog. They varied the response of the virtual dog to the physical collision (i.e., not doing anything or falling over) and the awareness of the other person (i.e., aware of the dog or unaware of the dog). They found that when the virtual dog exhibited the falling over behavior, participants felt more co-present with it and also gave lower affect scores to the other person regardless of their awareness level.

Overall, these findings indicate that a virtual character's awareness of both the physical and the virtual world and their ability to influence and be influenced by them is vital for their effective realization. The convergence of AR-IoT in transreality will increase the potential of such physical-virtual interactions using prevalent smart IoT devices, and the embodied interactions through/with virtual entities will benefit from such seamless mixed environments while offering unique opportunities to enhance social realism and influence. The findings addressed above can also be viewed as potential guidelines when designing embodied interactions in the transreality space.

32.5.4 Potential Use Cases

In this section, we describe a couple of use cases that capture the potential of AR-IoT convergence in our daily lives. These examples illustrate the integration of the ambient intelligence paradigm with the main aspects of the AR-IoT space that we addressed in Sect. 32.4: (1) distributed and object-centric data management, (2) IoT object-guided tracking, and (3) seamless interaction and content interoperability. Through this convergence, we describe the notion of an object-centric interconnection of physical and virtual things capable of exhibiting awareness of their environment, including the user, in a singular or a collective manner and accordingly facilitate AR interactions and experiences.

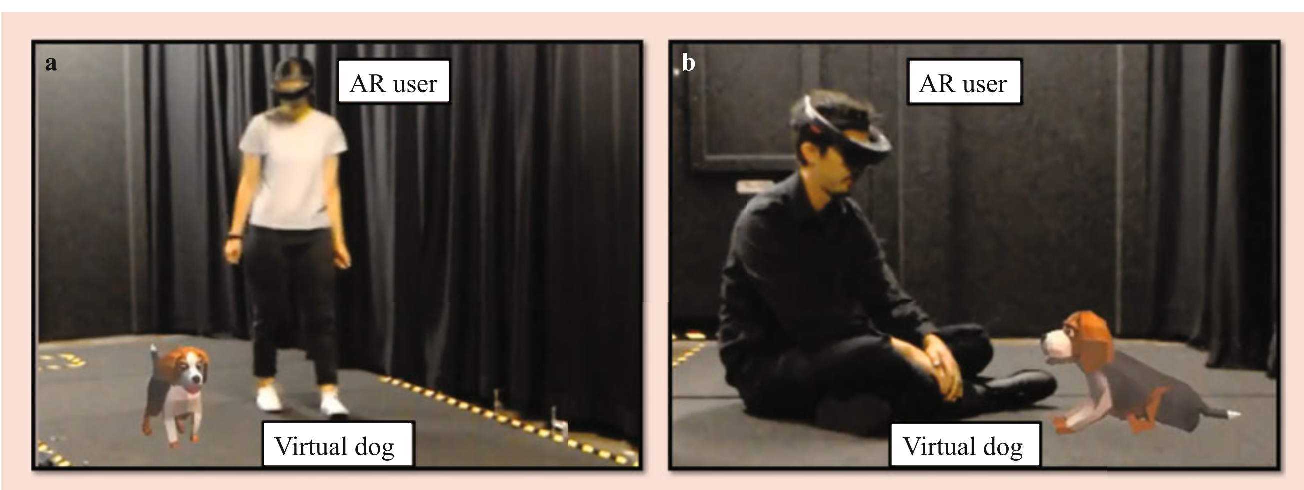


Fig. 32.14 Illustrations of example interactions with a virtual dog in AR. The users can see and interact with the virtual dog, for example, walking the dog together and playing fetch with it

Scenario 1 Steve is planning to buy a new high-end gaming personal computer (PC) and has already narrowed down the important criteria for the choice of PC, e.g., GPU/CPU requirements, display resolution, etc. Some of the options were based on the suggestions of his intelligent AR assistant, which are pervasively embedded in his mobile phone, wearable glasses, and home. At the electronics store, Steve goes to one of the store's guiding stations and gets virtual landmarks for directions to the devices on his list. To do so, his smart AR glasses directly connect to the guiding station, which opens up the virtual interaction space situated over the station itself. Steve asks his intelligent AR assistant for the device list they had devised earlier. The list virtually appears in his field of view through the AR glasses, and he interacts with it by dragging and dropping it on the virtual window of the guiding station. Now he gets appropriate directions for every item on his list and virtual landmarks for each item, which efficiently guide him on where to go and what to look for. Similar to the interaction with the guiding station, Steve's smart AR glasses can directly connect to, gather information from, and interact with objects of interest in the store without needing to connect to a main server, which is in line with the main components of the AR-IoT object-centric data management vision. For instance, situated with each item of interest, he can view the product information and control the device through AR interaction metaphors, such as playing a video from his own library to test the display quality by means of a virtual user interface and gestures. After some time, Steve has narrowed down his choice but is stuck between two options. His intelligent AR assistant appears and helps him to make the best decision by visualizing a comparison table and suggesting certain items for him. Depending on the amount of data, it might be cumbersome for the user to read all the visualized information in AR. In such cases, the AR assistant with visual embodiment can remind the user of each device's capability and also the preferences of the user while exhibiting appropriate gestural and facial expressions. The example above illustrates circumstances where the user is connected to each object separately or collectively through his intelligent AR assistant. Also, with the context of ambient intelligence in mind, the intelligent AR assistant already knows the preferences of the user, the form factor it should assume based on context, and the type of information the user might be interested in based on the profile data collected from previous interactions with him.

Scenario 2 Nowadays, it is a common practice for many professions, especially healthcare, to be trained using physical and virtual simulators with varying degrees of fidelity. Alice is an instructor at a university and has recently purchased a high-fidelity physical patient simulator with a human-like appearance to train medical students. Although Alice has used a number of simulators before, she is not sure how to interact with this model. Her smart AR glasses can detect the smart IoT-enabled simulator and directly connect to it. This leads to the appearance of the simulator's interaction metaphor visualized in her field of view situated next to the physical simulator. To get familiarized with the system, Alice can choose her preferred interface among several interface options, for example, a traditional graphical user interface or a virtual twin (i.e., in line with the digital twin paradigm [206] of the physical simulator, to educate herself. She chooses the virtual twin interface by gesturing toward that option. The virtual twin overlaid on the physical patient simulator describes all the different capabilities of the simulator step by step and encourages Alice to try each one. For instance, the virtual twin starts with the pulse capability of the simulator and asks her to touch the marked spot on the physical patient simulator's wrist to sense the pulse and activate the virtual interface for adjusting its value and potential changes for different medical scenarios. Later, it guides Alice to the add-on capabilities of the physical simulator, such as smart instrumented moulage devices that can be added onto the physical simulator to replicate combat casualty care scenarios [207]. Moreover, she can add the students' profiles in her class to the database of the simulator, and the intelligent interface of the system keeps track of the students' academic/training progress through observing the students' interactions with the simulator. For example, the simulator and the user's AR glasses in the AR-IoT environment monitor the students' activities, such as what symptoms of the simulated patient they missed in certain training scenarios and whether they did not treat the patient appropriately (e.g., not paying attention to the patient), which could lead to a wrong diagnosis, mistreatment, or unsatisfactory patient experience.

32.6 Conclusions

While AR/MR technology is experiencing a renaissance of development and consumer interest, there are still many obstacles in the way of the widespread adoption of ubiquitous AR in our everyday life. Azuma outlined the important challenges to overcome for ubiquitous AR, e.g., precise tracking anywhere and anytime, wide field of view of optical see-through near-eye displays, innovative interfaces, and semantic understanding of real-world objects [208]. Recent trends toward a merger of AR with UbiComp and advanced AI algorithms (i.e., the transreality paradigm) have the potential to resolve some of the challenges by using the distributed smart objects for data management and interactions while even making virtual content even more intelligent and interactive up to the point where they may be perceived as true social entities – such as socially influential virtual human avatars or agents [142, 143, 209]. In such transreality environments, we expect more natural and seamless interaction between the virtual content and the physical environment. To meet these expectations, advanced AR technology that enables more dynamic physical-virtual interaction has to be devised, and rigorous user studies are needed to understand how and in what ways the surrounding physical-virtual environment is contributing to human perception and interaction. Recent dramatic increase of evaluation research in AR supports this claim [3].

In this chapter, we described the concept of transreality, where physical and virtual worlds are highly merged and connected to each other, and presented the AR-IoT framework which can benefit from the realization of such transreality environments while discussing the main components in this realm to allow for a scalable, "anywhere" with "anything" interaction space in AR-IoT embedded environments. We presented a brief history of AR, IoT, and relevant concepts while covering the broader scopes of UbiComp and AI in consideration of synergies with pervasive AR technology. We further provided detailed descriptions of previous convergence research among them and literature that described related paradigms that envisioned and, in some cases, implemented the notion of pervasive and intelligent AR interactions with both real and virtual things.

We particularly described the three main components of the AR-IoT framework in the transreality paradigm, i.e., (1) distributed and object-centric AR-IoT data management, (2) scalable AR-IoT object guided tracking, and (3) context-based AR-IoT interaction and content interoperability, and described work that addressed the design and implementation of the components above.

Finally, we emphasized the potential benefits of embodied interactions in AR using virtual avatars and agents to improve the overall user experience and exert social influence over the users. The use of the AR-IoT framework, which can bring seamless interactions among the physical IoT objects and AR content, provides synergistic impacts on the user experience and behavior in human-agent interactions in the transreality space. We presented research that exemplifies such benefits and technical requirements for the "anywhere" with "anything" interaction space and described use cases that capture the essence of the transreality paradigm applied to a wide range of applications.

Together with upcoming high-speed 5G technologies and advanced Cloud and Edge computing paradigms, the interconnection between the virtual and real worlds will be more dynamic and immediate [210, 211]. As our real world is more and more equipped with IoT devices, we are moving toward a "physical Web," where information is tied to tangible physical objects, locations, and spaces [208]. Such smart devices will continuously monitor human activities and intelligently recognize the context for better user experience [212]. The AR-IoT framework provides a universal interface for us to access, retrieve, and absorb this information.

The notion of *digital living*, which stands for a paradigm of living without the bounds of place and time as presented by Negroponte in his book "Being Digital" [213], is being realized with this new convergence of technologies. We anticipate widespread AR in such digital living environments [214]; it is timely and important to research and develop the synergistic convergence of AR with other relevant domain technologies, such as the AR-IoT framework, while also considering other ethical issues like privacy and data security.

Acknowledgments This material includes work supported in part by the U.S. National Science Foundation under Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the U.S. Office of Naval Research under Award Numbers N00014-21-1-2578, N00014-18-1-2927, and N00014-21-1-2882 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch). This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2021R1I1A3060198). We also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2022-03294].

References

1. Azuma, R.T.: A survey of augmented reality. Presence Teleoper. Virtual Environ. **6**(4), 355–385 (1997)
2. Sutherland, I.E.: A head-mounted three dimensional display. In: Proceedings of the AFIPS '68 (Fall, Part I), pp. 757–764 (1968)
3. Kim, K., Billinghurst, M., Bruder, G., Duh, H.B.L., Welch, G.F.: Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008–2017). IEEE Trans. Vis. Comput. Graph. **24**(11), 2947–2962 (2018). https://doi.org/10.1109/TVCG.2018.2868591
4. Schwab, K.: The Fourth Industrial Revolution. Crown Publishing Group, USA (2017)
5. Jo, D., Kim, G.J.: Local context based recognition + internet of things: complementary infrastructures for future generic mixed reality space. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, p. 196 (2015). https://doi.org/10.1145/2821592.2821647
6. White, G., Cabrera, C., Palade, A., Clarke, S.: Augmented reality in IoT. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11434, pp. 149–160 (2019)
7. Gimenez, R., Pous, M.: Augmented reality as an enabling factor for the Internet of Things. In: W3C Workshop: Augmented Reality on the Web 2010, pp. 2–4 (2010)
8. Phupattanasilp, P., Tong, S.R.: Augmented reality in the integrative Internet of Things (AR-IoT): application for precision farming. Sustainability **11**(9), 2658 (2019). https://doi.org/10.3390/su11092658
9. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Sensing as a service model for smart cities supported by internet of things. Trans. Emer. Telecommun. Technol. **25**(1), 81–93 (2014)
10. Zhang, H., Zhang, Z., Zhang, L., Yang, Y., Kang, Q., Sun, D.: Object tracking for a smart city using IoT and edge computing. Sensors **19**(9), 1987 (2019)
11. Srimathi, B., Janani, E., Shanthi, N., Thirumoorthy, P.: Augmented reality based IoT concept for smart environment. Int. J. Intellect. Adv. Res. Eng. Comput **5**, 809–812 (2019)

12. Badouch, A., Krit, S.D., Kabrane, M., Karimi, K.: Augmented reality services implemented within smart cities, based on an internet of things infrastructure, concepts and challenges: an overview. In: Proceedings of the Fourth International Conference on Engineering & MIS 2018, pp. 1–4 (2018)
13. Brooks, F.P.: The computer scientist as toolsmith II. Commun. ACM **39**(3), 61–68 (1996). https://doi.org/10.1145/227234.227243
14. Radkowski, R.: Augmented reality to supplement work instructions. Model-Based Enterprise Summit (2013)
15. Jo, D., Kim, G.J.: IoT + AR: pervasive and augmented environments for "Digi-log" shopping experience. Hum.-Centric Comput. Inform. Sci. **9**(1), 1–17 (2019). https://doi.org/10.1186/s13673-018-0162-5
16. Lee, S., Lee, G., Choi, G., Roh, B.h., Kang, J.: Integration of OneM2M-based IoT service platform and mixed reality device. In: 2019 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–4. IEEE, Piscataway (2019)
17. Kasahara, S., Heun, V., Lee, A.S., Ishii, H.: Second surface: multiuser spatial collaboration system based on augmented reality. In: SIGGRAPH Asia 2012 Emerging Technologies, pp. 1–4 (2012)
18. Martin, K.A., Laviola, J.J.: The transreality interaction platform: enabling interaction across physical and virtual reality. In: Proceedings of the IEEE International Conference on Internet of Things (iThings); IEEE Green Computing and Communications (GreenCom); IEEE Cyber, Physical, and Social Computing (CPSCom); IEEE Smart Data (Smart Data), pp. 177–186. IEEE, Piscataway (2016). https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.54
19. Jo, D., Kim, G.J.: AR enabled IoT for a smart and interactive environment: a survey and future directions. Sensors **19**(19), 4330:1–17 (2019). https://doi.org/10.3390/s19194330
20. Norouzi, N., Bruder, G., Belna, B., Mutter, S., Turgut, D., Welch, G.: A systematic review of the convergence of augmented reality, intelligent virtual agents, and the Internet of Things. In: Springer Transactions on Computational Science Computational Intelligence, pp. 1–24 (2019). https://doi.org/10.1007/978-3-030-04110-6_1
21. Steuer, J.: Defining virtual reality: dimensions determining telepresence. J. Commun. **42**(4), 73–93 (1992). https://doi.org/10.1111/j.1460-2466.1992.tb00812.x
22. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality. Found. Trends® Hum.-Comput. Interact. **8**(2–3), 73–272 (2014)
23. Baudrillard, J.: Simulacra and Simulation. University of Michigan Press, Ann Arbor (1994)
24. Suzuki, K., Wakisaka, S., Fujii, N.: Substitutional reality system: a novel experimental platform for experiencing alternative reality. Sci. Rep. **2**(459), 1–9 (2012)
25. Sutherland, I.E.: The ultimate display. In: Proceedings of the Congress of the Internation Federation of Information Processing (IFIP), pp. 506–508 (1965)
26. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inform. Syst. **E77-D**(12), 1–15 (1994)
27. Mann, S.: Mediated reality with implementations for everyday life. Presence Connect, MIT Press Journal PRESENCE: Teleoperators and Virtual Environments (2002)
28. Mann, S., Furness, T., Yuan, Y., Iorio, J., Wang, Z.: All reality: virtual, augmented, mixed (X), mediated (X,Y), and multimediated reality. arXiv: 1804.08386 (2018)
29. Azuma, R.: Making Augmented Reality a Reality. Imaging and Applied Optics 2017 p. JTu1F.1 (2017)
30. Microsoft: What is mixed reality? (2018). https://docs.microsoft.com/en-us/windows/mixed-reality/mixed-reality
31. Keebler, J.R., Patzer, B.S., Wiltshire, T.J., Fiore, S.M.: Augmented reality systems in training. In: Brown, K.G. (ed.) The Cambridge Handbook of Workplace Training and Employee Development, Cambridge Handbooks in Psychology, pp. 278–292. Cambridge Handbooks in Psychology. Cambridge University Press, Cambridge (2017)
32. Althoff, T., White, R.W., Horvitz, E.: Influence of Pokémon go on physical activity: study and implications. J. Med. Internet Res. **18**(12), e315 (2016)
33. Digi-Capital: Record over 3b ar/vr investment in 2017 (2018). https://www.digi-capital.com/news/2018/01/record-over-3b-ar-vr-investment-in-2017-1-5b-in-q4
34. Buchanan, B.G.: A (very) brief history of artificial intelligence. AI Mag. **26**(4), 53–60 (2005)
35. Zhang, Q., Yang, L.T., Chen, Z., Li, P.: A survey on deep learning for big data. Inform. Fusion **42**(August 2017), 146–157 (2018)
36. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
37. Weiser, M., Brown, J.S.: Designing Calm Technology, pp. 1–5 (1995)
38. Weiser, M.: The computer for the 21st century. Sci. Am. **265**(3), 94–104 (1991)
39. Etzrodt, K., Engesser, S.: Ubiquitous tools, connected things and intelligent agents: disentangling the terminology and revealing underlying theoretical dimensions. First Monday (2019). https://doi.org/10.5210/fm.v24i9.9700
40. Poole, D.L., Mackworth, A., Goebel, R.G.: Computational intelligence and knowledge. In: Computational Intelligence: A Logical Approach, pp. 1–22. Oxford University Press, Oxford (1998)
41. Mangina, E., Carbo, J., Molina, J.M.: Agent-Based Ubiquitous Computing. Atlantis Ambient and Pervasive Intelligence, vol. 1. Atlantis Press, Paris (2010)
42. Mauldin, M.L.: ChatterBots, tinyMuds, and the turing test entering the loebner prize competition. In: Proceedings of the National Conference on Artificial Intelligence, pp. 16–21 (1994)
43. Bradeško, L., Mladenić, D.: A Survey of Chatbot Systems through a Loebner Prize Competition. In: Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies, vol. 2, pp. 34–37 (2012)
44. Klopfenstein, L.C., Delpriori, S., Malatini, S., Bogliolo, A.: The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In: Proceedings of the ACM Conference on Designing Interactive Systems, pp. 555–565 (2017). https://doi.org/10.1145/3064663.3064672
45. Tulshan, A.S., Dhage, S.N.: Survey on virtual assistant: Google Assistant, Siri, Cortana, Alexa. In: Advances in Signal Processing and Intelligent Recognition Systems, pp. 190–201 (2019). https://doi.org/10.1007/978-981-13-5758-9_17
46. Dubiel, M., Halvey, M., Azzopardi, L.: A Survey Investigating Usage of Virtual Personal Assistants (1), 2–6 (2018)
47. Kim, S., Lee, J., Gweon, G.: Comparing Data from Chatbot and Web Surveys. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems—CHI '19, pp. 86:1–12. ACM Press, New York (2019). https://doi.org/10.1145/3290605.3300316
48. Damacharla, P., Dhakal, P., Stumbo, S., Javaid, A.Y., Ganapathy, S., Malek, D.A., Hodge, D.C., Devabhaktuni, V.: Effects of voice-based synthetic assistant on performance of emergency care provider in training. Int. J. Artif. Intell. Educ. **29**(1), 122–143 (2019). https://doi.org/10.1007/s40593-018-0166-3
49. Chung, H., Iorga, M., Voas, J., Lee, S.: "Alexa, can i trust you?" Computer **50**(9), 100–104 (2017). https://doi.org/10.1109/MC.2017.3571053
50. Asthon, K.: That 'Internet of Things' Thing. RFID J. 4986 (2009)
51. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
52. Wortmann, F., Flüchter, K.: Internet of Things. Bus. Inform. Syst. Eng. **57**(3), 221–224 (2015)
53. Li, S., Da Xu, L., Zhao, S.: The Internet of Things: a survey. Inform. Syst. Front. **17**(2), 243–259 (2015)

54. Noura, M., Atiquzzaman, M., Gaedke, M.: Interoperability in Internet of Things: taxonomies and open challenges. Mob. Netw. Appl. **24**(3), 796–809 (2019)
55. Auto-ID Labs. https://www.autoidlabs.org/
56. Peña-López, I., et al.: ITU internet report 2005: the Internet of Things (2005)
57. Infso, D.: Networked enterprise & RFID INFSO G. 2 micro & nanosystems, in co-operation with the working group RFID of the ETP EPOSS, Internet of Things in 2020, roadmap for the future [R]. Information Society and Media, Tech. Rep., vol. 10 (2008)
58. Dunkels, A., Vasseur, J.: IP for smart objects, internet protocol for smart objects (IPSO) alliance, white paper# 1 (2008)
59. Rayes, A., Salam, S.: Internet of Things (IoT) overview. In: Internet of Things From Hype to Reality, pp. 1–35. Springer, Berlin (2019)
60. Li, H., Ota, K., Dong, M.: Learning IoT in edge: deep learning for the Internet of Things with edge computing. IEEE Netw. **32**(1), 96–101 (2018)
61. Diro, A.A., Chilamkurti, N.: Distributed attack detection scheme using deep learning approach for Internet of Things. Fut. Gener. Comput. Syst. **82**, 761–768 (2018)
62. Rothmuller, M., Barker, S.: IoT—The Internet of Transformation 2020 (2020)
63. Turgut, D., Boloni, L.: Value of information and cost of privacy in the Internet of Things. IEEE Commun. Mag. **55**(9), 62–66 (2017)
64. Mayle, A., Bidoki, N.H., Masnadi, S., Boeloeni, L., Turgut, D.: Investigating the value of privacy within the Internet of Things. In: IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE, Piscataway (2017)
65. Bölöni, L., Turgut, D.: Value of information based scheduling of cloud computing resources. Fut. Gener. Comput. Syst. **71**, 212–220 (2017)
66. Al-Turjman, F.: Information-centric framework for the Internet of Things (IoT): traffic modeling & optimization. Fut. Gener. Comput. Syst. **80**, 63–75 (2018)
67. Qi, J., Yang, P., Min, G., Amft, O., Dong, F., Xu, L.: Advanced Internet of Things for personalised healthcare systems: a survey. Pervasive Mob. Comput. **41**, 132–149 (2017)
68. Al-Turjman, F., Alturjman, S.: Context-sensitive access in Industrial Internet of Things (IIoT) healthcare applications. IEEE Trans. Ind. Inform. **14**(6), 2736–2744 (2018)
69. Indrakumari, R., Poongodi, T., Suresh, P., Balamurugan, B.: The growing role of Internet of Things in healthcare wearables. In: Balas, V.E., Solanki, V.K., Kumar, R. (eds.) Emergence of Pharmaceutical Industry Growth with Industrial IoT Approach, pp. 163–194. Academic Press, Cambridge (2020). https://doi.org/ https://doi.org/10.1016/B978-0-12-819593-2.00006-6
70. Stojkoska, B.L.R., Trivodaliev, K.V.: A review of Internet of Things for smart home: challenges and solutions. J. Cleaner Prod. **140**, 1454–1464 (2016). https://doi.org/10.1016/j.jclepro.2016.10.006
71. Iannizzotto, G., Nucita, A., Fabio, R.A., Caprì, T., Lo Bello, L.: More intelligence and less clouds in our smart homes. In: Economic and Policy Implications of Artificial Intelligence, pp. 123–136. Springer, Berlin (2020). https://doi.org/10.1007/978-3-030-45340-4_9
72. Kallmann, M., Thalmann, D.: Modeling objects for interaction tasks. In: Computer Animation and Simulation, pp. 73–86 (1999)
73. Cook, D.J., Augusto, J.C., Jakkula, V.R.: Ambient intelligence: technologies, applications, and opportunities. Pervasive Mob. Comput. **5**(4), 277–298 (2009)
74. Cook, D., Das, S.: Smart Environments: Technology, Protocols and Applications. In: Cook, D.J., Das, S.K. (eds.) Wiley, Hoboken (2004). https://doi.org/10.1002/047168659X
75. Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of Things: vision, applications and research challenges. Ad Hoc Netw. **10**(7), 1497–1516 (2012)
76. Olson, N., Nolin, J.M., Nelhans, G.: Semantic web, ubiquitous computing, or internet of things? A macro-analysis of scholarly publications. J. Doc. **71**(5), 884–916 (2015)
77. Brdiczka, O., Crowley, J.L., Langet, M., Maisonnasse, J.: Detecting human behavior model from multimodal observation in a smart Home. Autom. Sci. Eng. **6**(4), 588–597 (2008). https://doi.org/10.1109/TASE.2008.2004965
78. Smirek, L., Zimmermann, G., Beigl, M.: Just a smart home or your smart home—a framework for personalized user interfaces based on eclipse smart home and universal remote console. Proc. Comput. Sci. **98**, 107–116 (2016). https://doi.org/10.1016/j.procs.2016.09.018
79. Jacobsson, A., Boldt, M., Carlsson, B.: A risk analysis of a smart home automation system. Fut. Gener. Comput. Syst. **56**, 719–733 (2016). https://doi.org/10.1016/j.future.2015.09.003
80. Kum, S.W., Kang, M., Park, J.I.: IoT delegate: smart home framework for heterogeneous IoT service collaboration. KSII Trans. Internet Inform. Syst. **10**(8), 3958–3971 (2016). https://doi.org/10.3837/tiis.2016.08.029
81. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
82. Lee, B., Kwon, O., Lee, I., Kim, J.: Companionship with smart home devices: the impact of social connectedness and interaction types on perceived social support and companionship in smart homes. Comput. Hum. Behav. **75**, 922–934 (2017). https://doi.org/10.1016/j.chb.2017.06.031
83. Morvaj, B., Lugaric, L., Krajcar, S.: Demonstrating smart buildings and smart grid features in a smart energy city. In: Proceedings of the 3rd International Youth Conference on Energetics, pp. 2–9. IEEE, Piscataway (2011)
84. Radziwon, A., Bilberg, A., Bogers, M., Madsen, E.S.: The smart factory: exploring adaptive and flexible manufacturing solutions. Proc. Eng. **69**, 1184–1190 (2014)
85. Tang, S., Shelden, D.R., Eastman, C.M., Pishdad-Bozorgi, P., Gao, X.: A review of building information modeling (BIM) and the internet of things (IoT) devices integration: present status and future trends. Autom. Constr. **101**, 127–139 (2019)
86. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Sensing as a service model for smart cities supported by Internet of Things. Trans. Emer. Telecommun. Technol. **25**(1), 81–93 (2014)
87. NSF: Growing Convergence Research (2020).
88. NSF: Growing Convergence Research (2019)
89. Ducatel, K., européenne. Technologies de la société de l'information, U., européenne. Institut d'études de prospectives technologiques, U., européenne. Société de l'information conviviale, U.: Scenarios for ambient intelligence in 2010 (2001)
90. Remagnino, P., Foresti, G.L.: Ambient intelligence: a new multidisciplinary paradigm. IEEE Trans. Syst. Man Cybern. A: Syst. Hum. **35**(1), 1–6 (2004)
91. Ramos, C., Augusto, J.C., Shapiro, D.: Ambient intelligence—the next step for artificial intelligence. IEEE Intell. Syst. **23**(2),15–18 (2008)
92. Plötz, T., Hammerla, N.Y., Olivier, P.L.: Feature learning for activity recognition in ubiquitous computing. In: Twenty-Second International Joint Conference on Artificial Intelligence (2011)
93. Atallah, L., Yang, G.Z.: The use of pervasive sensing for behaviour profiling—a survey. Pervasive Mob. Comput. **5**(5), 447–464 (2009)
94. Doctor, F., Hagras, H., Callaghan, V.: A type-2 fuzzy embedded agent to realise ambient intelligence in ubiquitous computing environments. Inform. Sci. **171**(4), 309–334 (2005)
95. Patterson, D.J., Etzioni, O., Fox, D., Kautz, H.: Intelligent ubiquitous computing to support Alzheimer's patients: enabling the cognitively disabled. In: Adjunct Proceedings, p. 21 (2002)
96. Georgievski, I., Aiello, M.: Automated planning for ubiquitous computing. ACM Comput. Surv. **49**(4), 1–46 (2016)

97. DAngelo, G., Rampone, S., Palmieri, F.: An artificial intelligence-based trust model for pervasive computing. In: 2015 10th International Conference on P2p, Parallel, Grid, Cloud and Internet Computing (3pgcic), pp. 701–706 (2015)
98. Soda, S., Nakamura, M., Matsumoto, S., Izumi, S., Kawaguchi, H., Yoshimoto, M.: Implementing virtual agent as an interface for smart home voice control. In: 2012 19th Asia-Pacific Software Engineering Conference, vol. 1, pp. 342–345 (2012)
99. López, G., Quesada, L., Guerrero, L.A.: Alexa vs. Siri vs. Cortana vs. Google assistant: a comparison of speech-based natural user interfaces. In: International Conference on Applied Human Factors and Ergonomics, pp. 241–250 (2017)
100. Reis, A., Paulino, D., Paredes, H., Barroso, J.: Using intelligent personal assistants to strengthen the elderlies' social bonds. In: International Conference on Universal Access in Human-Computer Interaction, pp. 593–602 (2017)
101. Reis, A., Paredes, H., Barroso, I., Monteiro, M.J., Rodrigues, V., Khanal, S.R., Barroso, J.: Autonomous systems to support social activity of elderly people a prospective approach to a system design. In: 2016 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW), pp. 1–5 (2016)
102. Druga, S., Williams, R., Breazeal, C., Resnick, M.: " Hey Google is it OK if I eat you?" Initial explorations in child-agent interaction. In: Proceedings of the 2017 Conference on Interaction Design and Children, pp. 595–600 (2017)
103. Knote, R., Janson, A., Eigenbrod, L., Söllner, M.: The what and how of smart personal assistants: principles and application domains for IS research. Multikonferenz Wirtschaftsinformatik 1083–1094 (2018)
104. Austerjost, J., Porr, M., Riedel, N., Geier, D., Becker, T., Scheper, T., Marquard, D., Lindner, P., Beutel, S.: Introducing a virtual assistant to the lab: a voice user interface for the intuitive control of laboratory instruments. SLAS Technol. **23**(5), 476–482 (2018)
105. Lee, W., Cho, S., Chu, P., Vu, H., Helal, S., Song, W., Jeong, Y.S., Cho, K.: Automatic agent generation for IoT-based smart house simulator. Neurocomputing **209**, 14–24 (2016)
106. Helal, A., Cho, K., Lee, W., Sung, Y., Lee, J., Kim, E.: 3D modeling and simulation of human activities in smart spaces. In: 2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing, pp. 112–119 (2012)
107. Chung, H., Iorga, M., Voas, J., Lee, S.: Alexa, can I trust you? Computer **50**(9), 100–104 (2017)
108. Chung, H., Park, J., Lee, S.: Digital forensic approaches for Amazon Alexa ecosystem. Digital Invest. **22**, S15–S25 (2017)
109. Campagna, G., Ramesh, R., Xu, S., Fischer, M., Lam, M.S.: Almond: the architecture of an open, crowdsourced, privacy-preserving, programmable virtual assistant. In: Proceedings of the 26th International Conference on World Wide Web, pp. 341–350 (2017)
110. Holz, T., Campbell, A.G., O'Hare, G.M.P., Stafford, J.W., Martin, A., Dragone, M.: MiRA—mixed reality agents. Int. J. Hum.-Comput. Stud. **69**(4), 251–268 (2011). https://doi.org/10.1016/j.ijhcs.2010.10.001
111. Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The alive system: full-body interaction with autonomous agents. In: Proceedings Computer Animation'95, pp. 11–18 (1995)
112. Anabuki, M., Kakuta, H., Yamamoto, H., Tamura, H.: Welbo: An embodied conversational agent living in mixed reality space. In: CHI'00 Extended Abstracts on Human Factors in Computing Systems, pp. 10–11 (2000)
113. Balcisoy, S., Thalmann, D.: Interaction between real and virtual humans in augmented reality. In: Proceedings. Computer Animation '97 (Cat. No.97TB100120), pp. 31–38. IEEE Coput. Soc. Press (1997). https://doi.org/10.1109/CA.1997.601037
114. Kim, K., Welch, G.: Maintaining and enhancing human-surrogate presence in augmented reality. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Workshop on Human Perception and Psychology in Augmented Reality, pp. 15–19 (2015). https://doi.org/10.1109/ISMARW.2015.13
115. Barakonyi, I., Psik, T., Schmalstieg, D.: Agents that talk and hit back: animated agents in augmented reality. In: Third IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 141–150 (2004)
116. Barakonyi, I., Weilguny, M., Psik, T., Schmalstieg, D.: Monkeybridge: autonomous agents in augmented reality games. In: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology, pp. 172–175 (2005)
117. Chekhlov, D., Gee, A.P., Calway, A., Mayol-Cuevas, W.: Ninja on a plane: automatic discovery of physical planes for augmented reality using visual slam. In: 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 153–156 (2007)
118. Azad, S., Saldanha, C., Gan, C.H., Riedl, M.O.: Mixed reality meets procedural content generation in video games. In: Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference (2016)
119. Hantono, B.S., Nugroho, L.E., Santosa, P.I.: Review of augmented reality agent in education. In: 2016 6th International Annual Engineering Seminar (InAES), pp. 150–153 (2016)
120. Wagner, D., Billinghurst, M., Schmalstieg, D.: How real should virtual characters be? In: Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, pp. 57–es (2006)
121. Holstein, K., McLaren, B.M., Aleven, V.: Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In: International Conference on Artificial Intelligence in Education, pp. 154–168 (2018)
122. Cavazza, M., Charles, F., Mead, S.J., Martin, O., Marichal, X., Nandi, A.: Multimodal acting in mixed reality interactive storytelling. IEEE MultiMedia **11**(3), 30–39 (2004)
123. Dow, S., Mehta, M., Harmon, E., MacIntyre, B., Mateas, M.: Presence and engagement in an interactive drama. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 1475–1484 (2007)
124. Cheok, A.D., Yang, X., Ying, Z.Z., Billinghurst, M., Kato, H.: Touch-space: mixed reality game space based on ubiquitous, tangible, and social computing. Pers. Ubiquitous Comput. **6**(5-6), 430–442 (2002)
125. Montola, M.: A ludological view on the pervasive mixed-reality game research paradigm. Personal Ubiquitous Comput. **15**(1), 3–12 (2011)
126. Vogiazou, Y., Raijmakers, B., Geelhoed, E., Reid, J., Eisenstadt, M.: Design for emergence: experiments with a mixed reality urban playground game. Personal Ubiquitous Comput. **11**(1), 45–58 (2007)
127. Chalmers, M., MacColl, I.: Seamful and seamless design in ubiquitous computing. In: Workshop at the Crossroads: The interaction of HCI and Systems Issues in UbiComp, vol. 8 (2003)
128. Lee, J.Y., Rhee, G.W., Seo, D.W.: Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment. Int. J. Adv. Manuf. Technol. **51**(9–12), 1069–1082 (2010)
129. Han, T., Anderson, F., Irani, P., Grossman, T.: Hydroring: supporting mixed reality haptics using liquid flow. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, pp. 913–925 (2018)
130. Newman, J., Bornik, A., Pustka, D., Echtler, F., Huber, M., Schmalstieg, D., Klinker, G.: Tracking for distributed mixed reality environments. In: Workshop on Trands and Issues in

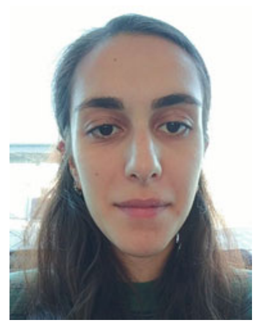
32

Tracking for Virtual Environments at the IEEE Virtual Reality Conference (VR'07). Citeseer (2007)
131. Singh, G., Mantri, A.: Ubiquitous hybrid tracking techniques for augmented reality applications. In: 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), pp. 1–5 (2015)
132. Macias, J.A.G., Alvarez-Lozano, J., Estrada, P., Lopez, E.A.: Browsing the Internet of Things with sentient visors. Computer **44**(5), 46–52 (2011)
133. Wirtz, H., Rüth, J., Serror, M., Bitsch Link, J.Á., Wehrle, K.: Opportunistic interaction in the challenged Internet of Things. In: Proceedings of the 9th ACM MobiCom workshop on Challenged networks, pp. 7–12 (2014)
134. Heun, V., Hobin, J., Maes, P.: Reality editor: programming smarter objects. In: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, pp. 307–310 (2013)
135. Liu, C., Huot, S., Diehl, J., Mackay, W., Beaudouin-Lafon, M.: Evaluating the benefits of real-time feedback in mobile augmented reality with hand-held devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2973–2976 (2012)
136. Lee, L.H., Hui, P.: Interaction methods for smart glasses: a survey. IEEE Access **6**, 28712–28732 (2018)
137. Zhang, B., Chen, Y.H., Tuna, C., Dave, A., Li, Y., Lee, E., Hartmann, B.: Hobs: head orientation-based selection in physical spaces. In: Proceedings of the 2nd ACM Symposium on Spatial User Interaction, pp. 17–25 (2014)
138. Kollee, B., Kratz, S., Dunnigan, A.: Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing. In: Proceedings of the 2nd ACM Symposium on Spatial User Interaction, pp. 40–49 (2014)
139. Hao, Y., Helo, P.: The role of wearable devices in meeting the needs of cloud manufacturing: a case study. Robot. Comput.-Integr. Manuf. **45**, 168–179 (2017)
140. Schmidt, S., Nunez, O.J.A., Steinicke, F.: Blended agents: Manipulation of physical objects within mixed reality environments and beyond. In: Symposium on Spatial User Interaction, pp. 1–10 (2019)
141. Lee, M., Norouzi, N., Bruder, G., Wisniewski, P.J., Welch, G.F.: Mixed reality tabletop gameplay: social interaction with a virtual human capable of physical influence. IEEE Trans. Vis. Comput. Graph. **24**(8), 1–12 (2019)
142. Kim, K., Norouzi, N., Losekamp, T., Bruder, G., Anderson, M., Welch, G.: Effects of patient care assistant embodiment and computer mediation on user experience. In: Proceedings of IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), pp. 17–24 (2019). https://doi.org/10.1109/AIVR.2019.00013
143. Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., Welch, G.F.: Does a digital assistant need a body? the influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In: 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 105–114 (2018)
144. Salman, T., Jain, R.: A survey of protocols and standards for Internet of Things. arXiv preprint arXiv:1903.11549 (2019)
145. Lindley, C.A.: Game space design foundations for trans-reality games. In: Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, pp. 397–404. ACM Press, New York (2005). https://doi.org/10.1145/1178477.1178569
146. Kim, S., Suh, Y., Lee, Y., Woo, W.: Toward ubiquitous VR: when VR meets ubiComp. In: The 4th International Symposium on Ubiquitous VR, pp. 1–4 (2006)
147. Paradiso, J.A., Landay, J.A.: Guest editors' introduction: cross-reality environments. IEEE Pervasive Comput. **8**(3), 14–15 (2009)
148. Lifton, J., Paradiso, J.A.: Dual reality: merging the real and virtual. In: International Conference on Facets of Virtual Environments, pp. 12–28 (2009)
149. Lifton, J., Laibowitz, M., Harry, D., Gong, N.w., Mittal, M., Paradiso, J.A.: Metaphor and manifestation—cross-reality with ubiquitous sensor/actuator networks. IEEE Pervasive Comput. **8**(3), 24–33 (2009)
150. Second life
151. Gelernter, D.: Mirror Worlds or the Day Software Puts the Universe in a Shoebox: How It Will Happen and What It Will Mean. Oxford University Press, Oxford (1991)
152. Ricci, A., Piunti, M., Tummolini, L., Castelfranchi, C.: The mirror world: preparing for mixed-reality living. IEEE Pervasive Comput. **14**(2), 60–63 (2015). https://doi.org/10.1109/MPRV.2015.44
153. Seo, D.W., Kim, H., Kim, J.S., Lee, J.Y.: Hybrid reality-based user experience and evaluation of a context-aware smart home. Comput. Ind. **76**, 11–23 (2016). https://doi.org/10.1016/j.compind.2015.11.003
154. Grieves, M.: Product Lifecycle Management: Driving the Next Generation of Lean Thinking: Driving the Next Generation of Lean Thinking. McGraw-Hill Education (2005)
155. Glaessgen, E., Stargel, D.: The digital twin paradigm for future NASA and U.S. Air Force vehicles. In: 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, pp. 1–14. American Institute of Aeronautics and Astronautics (2012). https://doi.org/10.2514/6.2012-1818
156. Tao, F., Zhang, M., Nee, A.Y.C.: Chapter 11—digital twin and virtual reality and augmented reality/mixed reality. In: Tao, F., Zhang, M., Nee, A.Y.C. (eds.) Digital Twin Driven Smart Manufacturing, pp. 219–241. Academic Press (2019). https://doi.org/10.1016/B978-0-12-817630-6.00011-4
157. Applin, S.A., Fischer, M.: A cultural perspective on mixed, dual and blended reality. In: IUI Workshop on Location Awareness for Mixed and Dual Reality, pp. 11–14 (2011)
158. Blanco-Novoa, Ó., Fraga-Lamas, P., Vilar-Montesinos, M.A., Fernández-Caramés, T.M.: Creating the internet of augmented things: an open-source framework to make IoT devices and augmented and mixed reality systems talk to each other. Sensors **20**(11), 1–26 (2020)
159. Michalakis, K., Aliprantis, J., Caridakis, G.: Visualizing the Internet of Things: naturalizing human-computer interaction by incorporating ar features. IEEE Consum. Electron. Mag. **7**(3), 64–72 (2018)
160. Choi, M.I., Park, L.W., Lee, S., Hwang, J.Y., Park, S.: Design and implementation of hyper-connected IoT-VR platform for customizable and intuitive remote services. In: 2017 IEEE International Conference on Consumer Electronics (ICCE), pp. 396–397 (2017)
161. Rekimoto, J., Nagao, K.: The world through the computer: Computer augmented interaction with real world environments. In: Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology, pp. 29–36 (1995)
162. Huo, K., Cao, Y., Yoon, S.H., Xu, Z., Chen, G., Ramani, K.: Scenariot: Spatially Mapping Smart Things Within Augmented Reality Scenes. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 219:1–13 (2018). https://doi.org/10.1145/3173574.3173793
163. Iglesias, J., Bernardos, A.M., Casar, J.R.: An evidential and context-aware recommendation strategy to enhance interactions with smart spaces. In: International Conference on Hybrid Artificial Intelligence Systems, pp. 242–251 (2013)

164. Ajanki, A., Billinghurst, M., Gamper, H., Järvenpää, T., Kandemir, M., Kaski, S., Koskela, M., Kurimo, M., Laaksonen, J., Puolamäki, K., Ruokolainen, T.: An augmented reality interface to contextual information. Virtual Reality **15**(2–3), 161–173 (2011)
165. Huang, B.R., Lin, C.H., Lee, C.H.: Mobile augmented reality based on cloud computing. In: Anti-Counterfeiting, Security, and Identification, pp. 1–5 (2012)
166. Jesudian, S.S., Jabez, J.: An integration of IoT and mixed reality using fog computing concepts (2018)
167. Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp. 13–16 (2012)
168. Ghobaei-Arani, M., Souri, A., Rahmanian, A.A.: Resource management approaches in fog computing: a comprehensive review. J. Grid Comput. **18**, 1–42 (2020)
169. Tao, F., Zhang, M., Nee, A.: Digital twin and cloud, fog, edge computing. In: Digital Twin Driven Smart Manufacturing, pp. 171–181 (2019)
170. Google: The Physical Web (May 2020)
171. Ahn, S., Ko, H., Yoo, B.: Webizing mobile augmented reality content. New Rev. Hypermedia Multimedia **20**(1), 79–100 (2014)
172. Kim, M., Choi, S.H., Park, K.B., Lee, J.Y.: User interactions for augmented reality smart glasses: a comparative evaluation of visual contexts and interaction gestures. Appl. Sci. **9**(15), 3171 (2019)
173. Müller, L., Aslan, I., Krüßen, L.: Guideme: a mobile augmented reality system to display user manuals for home appliances. In: International Conference on Advances in Computer Entertainment Technology, pp. 152–167 (2013)
174. Kim, J., Lee, J., Yoo, B., Ahn, S., Ko, H.: [poster] view management for webized mobile AR contents. In: 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 271–272 (2014)
175. Claros, D., De Haro, M., Domínguez, M., Trazegnies, C.D., Urdiales, C., Sandoval, F.: Augmented reality visualization interface for biometric wireless sensor networks. In: International Work-Conference on Artificial Neural Networks, pp. 1074–1081 (2007)
176. Mihara, S., Sakamoto, A., Shimada, H., Sato, K.: Augmented reality marker for operating home appliances. In: 2011 IFIP 9th International Conference on Embedded and Ubiquitous Computing, pp. 372–377 (2011)
177. Uchiyama, H., Marchand, E.: Object detection and pose tracking for augmented reality: recent approaches (2012)
178. Diao, P.H., Shih, N.J.: Marins: A mobile smartphone ar system for pathfinding in a dark environment. Sensors **18**(10), 3442 (2018)
179. Park, Y., Yun, S., Kim, K.H.: When IoT met augmented reality: visualizing the source of the wireless signal in ar view. In: Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, pp. 117–129 (2019)
180. Jackson, B., Watten, P., Newbury, P., Lister, P.: Rapid authoring for VR-based simulations of pervasive computing applications (2005)
181. Lu, C.H.: IoT-enhanced and bidirectionally interactive information visualization for context-aware home energy savings. In: 2015 IEEE International Symposium on Mixed and Augmented Reality-Media, Art, Social Science, Humanities and Design, pp. 15–20 (2015)
182. Zhu, Z., Branzoi, V., Wolverton, M., Murray, G., Vitovitch, N., Yarnall, L., Acharya, G., Samarasekera, S., Kumar, R.: AR-mentor: Augmented reality based mentoring system. In: 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 17–22 (2014)
183. Rekimoto, J., Ayatsuka, Y.: Cybercode: designing augmented reality environments with visual tags. In: Proceedings of DARE 2000 on Designing augmented reality environments, pp. 1–10 (2000)
184. Tran, G.: Mediating Mediums (August 2011)
185. Kruijff, E., Swan, J.E., Feiner, S.: Perceptual issues in augmented reality revisited. In: 2010 IEEE International Symposium on Mixed and Augmented Reality, pp. 3–12 (2010)
186. Rolland, J.P., Fuchs, H.: Optical versus video see-through head-mounted displays in medical visualization. Presence Teleoper. Virtual Environ. **9**(3), 287–309 (2000)
187. Kasahara, S., Niiyama, R., Heun, V., Ishii, H.: exTouch: spatially-aware embodied manipulation of actuated objects mediated by augmented reality. In: Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, pp. 223–228 (2013)
188. Petersen, N., Stricker, D.: Continuous natural user interface: reducing the gap between real and digital world. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 23–26 (2009). https://doi.org/10.1109/ISMAR.2009.5336502
189. LaViola, J.J., Kruijff, E., McMahan, R.P., Bowman, D.A., Poupyrev, I.: 3D User Interfaces: Theory and Practice. Addison-Wesley Usability and HCI Series. Addison-Wesley (2017)
190. Lages, W.S., Bowman, D.A.: Walking with adaptive augmented reality workspaces. In: Proceedings of the ACM International Conference on Intelligent User Interfaces, pp. 356–366 (2019)
191. Ekman, P.: Emotional and Conversational Nonverbal Signals. In: Language, Knowledge, and Representation, vol. 99, pp. 39–50. Springer, Dordrecht (2004)
192. Haeuslschmid, R., Von Buelow, M., Pfleging, B., Butz, A.: Supporting trust in autonomous driving. In: Proceedings of International Conference on Intelligent User Interfaces (IUI), pp. 319–329 (2017)
193. Amores, J., Benavides, X., Comin, M., Fuste, A., Pla, P., Miralles, D.: Smart avatars: using avatars to interact with objects. In: Proceedings of ACM CHI, pp. 1–3 (2012)
194. Haesler, S., Kim, K., Bruder, G., Welch, G.: Seeing is believing: improving the perceived trust in visually embodied alexa in augmented reality. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 204–205 (2018)
195. Norouzi, N., Kim, K., Bruder, G., Erickson, A., Choudhary, Z., Li, Y., Welch, G.: A systematic literature review of embodied augmented reality agents in head-mounted display environments. In: In Proceedings of the International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments, p. 11 (2020)
196. Lee, M., Norouzi, N., Bruder, G., Wisniewski, P., Welch, G.: Mixed reality tabletop gameplay: social interaction with a virtual human capable of physical influence. IEEE Trans. Vis. Comput. Graph. **8**, 3534–3545 (2021)
197. Lee, M., Norouzi, N., Bruder, G., Wisniewski, P.J., Welch, G.F.: The physical-virtual table: exploring the effects of a virtual human's physical influence on social interaction. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, pp. 1–11 (2018)
198. Lee, M., Kim, K., Daher, S., Raij, A., Schubert, R., Bailenson, J., Welch, G.: The wobbly table: Increased social presence via subtle incidental movement of a real-virtual table. In: 2016 IEEE Virtual Reality (VR), pp. 11–17. IEEE (2016)
199. Lee, M., Bruder, G., Höllerer, T., Welch, G.: Effects of unaugmented periphery and vibrotactile feedback on proxemics with virtual humans in AR. IEEE Trans. Vis. Comput. Graph. **24**(4), 1525–1534 (2018)
200. Kim, K., Schubert, R., Hochreiter, J., Bruder, G., Welch, G.: Blowing in the wind: increasing social presence with a virtual

human via environmental airflow interaction in mixed reality. Comput. Graph. **83**, 23–32 (2019)
201. Kim, K., Bruder, G., Welch, G.F.: Blowing in the wind: increasing copresence with a virtual human via airflow influence in augmented reality. In: ICAT-EGVE, pp. 183–190 (2018)
202. Kim, K., de Melo, C.M., Norouzi, N., Bruder, G., Welch, G.F.: Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 529–538 (2020)
203. Wang, I., Smith, J., Ruiz, J.: Exploring virtual agents for augmented reality. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2019)
204. Johnsen, K., Ahn, S.J., Moore, J., Brown, S., Robertson, T.P., Marable, A., Basu, A.: Mixed reality virtual pets to reduce childhood obesity. IEEE Trans. Vis. Comput. Graph. **20**(4), 523–530 (2014)
205. Norouzi, N., Kim, K., Lee, M., Schubert, R., Erickson, A., Bailenson, J., Bruder, G., Welch, G.: Walking your virtual dog: Analysis of awareness and proxemics with simulated support animals in augmented reality. In: 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 157–168 (2019)
206. Canedo, A.: Industrial IoT lifecycle via digital twins. In: Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis, CODES '16. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2968456.2974007
207. Taylor, G., Deschamps, A., Tanaka, A., Nicholson, D., Bruder, G., Welch, G., Guido-Sanz, F.: Augmented reality for tactical combat casualty care training. In: Augmented Cognition: Users and Contexts, pp. 227–239. Springer, Berlin (2018)
208. Azuma, R.T.: The most important challenge facing augmented reality. Presence Teleoper. Virtual Environ. **25**(3), 234–238 (2016)
209. Norouzi, N., Kim, K., Hochreiter, J., Lee, M., Daher, S., Bruder, G., Welch, G.: A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In: ACM International Conference on Intelligent Virtual Agents, pp. 17–22 (2018). https://doi.org/10.1145/3267851.3267901
210. Bastug, E., Bennis, M., Medard, M., Debbah, M.: Toward interconnected virtual reality: opportunities, challenges, and enablers. IEEE Commun. Mag. **55**(6), 110–117 (2017). https://doi.org/10.1109/MCOM.2017.1601089
211. Torres Vega, M., Liaskos, C., Abadal, S., Papapetrou, E., Jain, A., Mouhouche, B., Kalem, G., Ergüt, S., Mach, M., Sabol, T., Cabellos-Aparicio, A., Grimm, C., De Turck, F., Famaey, J.: Immersive interconnected virtual and augmented reality: A 5G and IoT perspective. J. Netw. Syst. Manag. **28**(4), 796–826 (2020). https://doi.org/10.1007/s10922-020-09545-w
212. Zehtabian, S., Razghandi, M., Boloni, L., Turgut, D.: Predictive caching for AR/VR experiences in a household scenario. In: Proceedings of the International Conference on Computing, Networking and Communications, pp. 591–595 (2020). https://doi.org/10.1109/ICNC47757.2020.9049692
213. Negroponte, N.: Being Digital. Alfred A. Knopf (1995)
214. Welch, G.F., Bruder, G., Squire, P., Schubert, R.: Anticipating Widespread Augmented Reality: Insights from the 2018 AR Visioning Workshop. Tech. rep., University of Central Florida and Office of Naval Research (2019)

Kangsoo Kim is an Assistant Professor in the Department of Electrical and Software Engineering at the University of Calgary. He earned his Ph.D. in Computer Science from the University of Central Florida in 2018. His research interests broadly cover Human-Computer Interaction and Augmented/Mixed Reality (AR/MR), particularly Virtual Avatars/Agents, and Perception and Cognition in MR.



Nahal Norouzi earned her PhD and master's in Computer Science from the University of Central Florida in 2021 and 2020 and her bachelor's in Electrical Engineering from Tehran Polytechnic University. She is currently a Post-Doctoral Scholar in the Synthetic Reality Lab at the University of Central Florida. She is primarily researching different areas of Virtual and Augmented Reality, including virtual animals and humans, augmented and enhanced perception, and virtual collaboration.

Dongsik Jo received his Ph.D. degree in Computer Engineering from Korea University in 2017. In 2004, he joined the virtual reality team, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. Since 2021, he has been a Professor of the School of IT Convergence, University of Ulsan. His research interests include Virtual Reality, Augmented Reality, IoT Sensing, and Avatar Motion.

Gerd Bruder is a Research Assistant Professor at the Institute for Simulation and Training at the University of Central Florida. He received his Habilitation in Computer Science from the University of Hamburg in 2017 and his Ph.D. in Computer Science from the University of Münster in 2011. His research interests include Mixed Reality, Perception and Cognition, and Human-Computer Interaction.

Gregory F. Welch is a Computer Scientist with appointments in UCF Nursing, Computer Science (CS), and the Institute for Simulation and Training. Welch earned his B.S. in E.E.T. from Purdue University and his M.S. and Ph.D. in CS from UNC Chapel Hill. Previously, he worked at UNC, NASA, and Northrop. His research interests include Virtual and Augmented Reality and Medical Applications.

Convergence of IoT and Augmented Reality

33

Alexis Morris and Georgina Yeboah

Contents

33.1 Introduction 831
33.1.1 Mixed Reality: A Parallel to the IoT Ecosystem 832
33.1.2 Toward XRI: A Growing Trend in Hybrid Mixed Reality IoT Systems 833
33.1.3 Chapter Overview 834

33.2 A Multidisciplinary Taxonomy for XRI Systems 834
33.2.1 XRI Taxonomy: Thematic Literature Review 834
33.2.2 Relating the XRI Taxonomy to XRI System Design Needs from a Multi-disciplinary Perspective 840

33.3 Contextual Reality (CoRe) Frameworks for XRI System Design 841
33.3.1 Designing Proof-of-Concept XRI Systems with CoRe 843
33.3.2 CoRe Framework Insights in Relation to the XRI Taxonomy Perspectives 844

33.4 Discussion 848
33.4.1 Directions for Future Work 848

33.5 Conclusion 849

References 849

Abstract

This chapter contends that the Internet of Things (IoT), which consists of technologies for networked embedded devices and decentralized software applications, must also have rich interface solutions that both inform and engage its users. This is a new frontier in IoT human-computer interaction (HCI), and meeting these challenges requires interface designs that present information to users in a context-sensitive manner, which can inform users while allowing them to remain engaged in their environmental tasks. Mixed reality, combined with context sensing, also known as X-Reality (XR), allows the display of 3D content in situ, providing novel interface design possibilities for IoT edge devices. These XR-IoT (XRI) hybrid systems can become more personal, immersive, embedded, information-rich, decentralized, multiuser, and agent-driven. This repositions the IoT from the passive background embedded environment into active, engaging, foreground information infrastructures. This chapter highlights the challenges and considerations to meet the convergence of the IoT with mixed reality, toward a transformation of IoT devices into hybrid virtual and physical objects that adapt to their user's context, and presents information in engaging ways. Such a merger will contribute to the IoT broadly and impact HCI across existing and future IoT deployments.

Keywords

Internet-of-Things · Human-computer interaction · Mixed reality · Virtual and augmented reality · Multi-agent systems

33.1 Introduction

The Internet of Things is a major computing paradigm with global impact and a maturing research community. It is poised to connect people with smart devices embedded in their environment, enabling them, as users and information consumers, to access these device properties and interact with their functions. The expected impact of the IoT interconnected environment is well-known, particularly as it spurs Industry 4.0, where connected devices offer economic benefits, cost savings, and better information, leading to better decision-making, and efficiency. This is estimated within the trillions of dollars (conservative estimates range between \$2.7trillion and \$6.2 trillion by 2025, for example, with typical estimates above \$14 trillion by 2025) [55], across varied industries, ranging from healthcare to smart vehicles, smart cities, and consumer applications. The IoT is surging, with demand for smarter environments, deployments of sen-

A. Morris (✉) · G. Yeboah
Adaptive Context Environments Lab, OCAD University, Toronto, ON, Canada
e-mail: amorris@ocadu.ca; gyeboah@ocadu.ca

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_33

sor equipment, smart space assistant software, commercially available devices, and industrial sensing devices. The use cases are broadly applicable, and the benefits are impactful in terms of cost savings, situational awareness, and information content accessibility.

However, modern IoT systems have considerable human-computer interaction (HCI) challenges that are not presently being addressed by current IoT interface designs; for instance, current interfaces for the IoT make extensive use of 2D dashboards with "drill-down" metaphors [51] and smartphone and smartwatch interfaces, some of which provide haptics [2]; and recently, voice-based interfaces have become viable [40]; these all may involve multiple users running multiple services simultaneously, forming a challenging complex systems design and interaction problem [19]. These kinds of interaction are also challenged by the medium they use; for instance, voice interaction enables verbal commands and audio information reports but makes it difficult to summarize information for rapid comprehension, unlike information in visual form factors.

As the IoT evolves to meet the many application use cases where it is being deployed, it also must address the many important human-computer interaction challenges [60] and sociotechnical system needs inherent to its deployment. These are difficult to quantify and incorporate and hence are often overlooked [52]; but they have significant influence on the successful adoption of the technology. Considering the IoT from a human factors' lens, such as using the physical, psychological, social, organizational, and political lens [59], leads to insights into where to focus the development of a more human-centered IoT. For instance, IoT devices are often without a clear user interface and are often invisible in the background, which raises issues of transparency in IoT-enabled smart environments, where it is not always apparent as to how to correctly identify, operate, and access these devices across the many contexts in which they are deployed [54]. This is a recent research gap in the IoT community, one which has an impact on cognitive overload and task efficiency for humans involved. Further, there remains the need to explore interaction metaphors for IoT environments that go beyond existing desktop metaphors and 2D dashboards [51] and which can incorporate both explicit and implicit forms of interaction – as IoT devices are often operating based on sensory inputs and contextual sense-making in addition to user commands. The IoT research community is beginning to address these topics [30, 46, 56], while the industry continues the deployment of IoT systems – the landscape is in a state of rapid evolution.

33.1.1 Mixed Reality: A Parallel to the IoT Ecosystem

The domain of mixed reality is also surging, with a healthy and growing ecosystem; hardware availability and suitability for immersive experiences is growing in terms of its ability to provide quality experiences (with high degrees of presence, interactivity at room scale, hand presence, object anchoring and tracking, and reasonably wide field of view). The hardware is also now more affordable, and the software development frameworks are now becoming mature, leading to a breadth of mixed reality software experiences now in the hands of consumers and enterprise – showing benefits in terms of education and training, cost savings, and more [14].

The ongoing advances in the domain of immersive media using mixed reality have produced practical, capable toolsets for providing visual and multimodal interaction solutions and have paved the way for new kinds of graphical user interfaces that can be projected or overlaid into the viewer's environment, through conventional handheld mobile device form factors and head-mounted form factors [50]. The domain of mixed reality is now growing at a similar pace with the IoT and is itself expected to be a multibillion-dollar industry that will transform the mobile computing landscape in the current decade [14]. This chapter is based on the premise that mixed reality displays, like smart glasses, are beneficial as IoT interfaces (versus current traditional displays like smartphones) for interaction, control, adaptation, and collaboration within context management systems for IoT applications – a growing trend in the literature [23, 61]. Mixed reality (MR) refers to the "subclass of Virtual Reality (VR) related technologies that involve the merging of real and virtual worlds," wherein a spectrum of immersive devices have been explored [38], as in Fig. 33.1.

This includes augmented reality (AR), wherein "an otherwise real environment is augmented by means of virtual (computer graphic) objects" [38], and even augmented virtuality (AV), where the viewer's perspective of a virtual world is augmented by real-world content. Virtual reality refers to environments where "the participant observer is totally immersed in, and able to interact with, a completely synthetic world" [38].

Recent state-of-the art surveys in mixed reality evidence the growth and maturity of the field, as it begins to achieve mainstream; for example, Billinghurst, et al. [5] take a comprehensive look at nearly 50 years of augmented reality

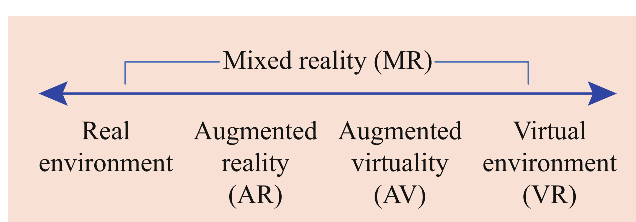


Fig. 33.1 Milgram's reality-virtuality continuum, as a standard description of the mixed reality domain, based on [39]

(AR) research, highlighting AR-related concepts in display, input and interaction technologies, design guidelines, interface patterns, AR development tools and evaluations of AR systems, and other AR and VR technology requirements needed for AR and VR to replace real environment elements and the evaluation of such systems. This includes themes like Scene Generation, Display Device, and Tracking Sensing and present-day applications within education, architecture, and marketing among future research directions.

Hence, designs of hardware and software for projects all across this spectrum are all becoming mainstream [14], and the continued application of mixed reality displays (particularly in the form of smart glasses and head-mounted displays (HMDs)) to the domain of IoT systems is inevitable; however, the potential effects and benefits, especially to the context management systems in IoT applications [61], remain open for exploration.

33.1.2 Toward XRI: A Growing Trend in Hybrid Mixed Reality IoT Systems

The gap in HCI for the IoT is a long-standing open research challenge, as current interfaces remain insufficient at being contextually relevant, easily understandable, and non-distracting [30, 46]. There is a need to bring together design thinking, context awareness, computer vision, IoT sensor networking, and mixed reality into a single system framework, which also aligns with the definition of cross-reality, or X-Reality (XR), as a ubiquitous MR environment [47], and in this chapter as mixed reality IoT, or XRI, emphasizing the fusion of XR with the IoT. Research opportunities arise for merging mixed reality with the IoT, where advances in MR and VR must be designed to fit modern IoT systems and frameworks. In fact, the synergies between these two technologies are being increasingly recognized within the most recent literature, such as [6], highlighting the complementarity of mixed reality needs versus IoT needs, particularly in terms of critical problem overlaps, such as AR and IoT data management, object-guided tracking, object control, and interface designs [23]. Smart glasses for IoT are promising, as they are audiovisual and can have reasonably large field-of-view "effective" display size and resolution, along with stereoscopic depth [31,32,58]. Also, the form factor, wearable properties, and projected advancement toward augmented reality and mixed reality systems [33] make these forms of display a potential avenue for immersive, adaptive IoT interfaces and interaction with embedded IoT devices.

Likewise, with the current wearable and wireless form factors of existing equipment, it becomes clear that both technologies address a unique aspect of the future human-computer interface. Mixed reality expands the canvas of information content beyond the display, enveloping and bringing the user more closely in line with virtual objects and enhancing experiences and awareness. The IoT expands environmental intelligence using embedded sensors and enhances situational awareness and response behavior of smart devices and the decision-making of humans-in-the-loop (HitL). It can allow for multisectoral perspectives of situational awareness across varied domains in real time, contributing to a common operating picture of events, enabling dynamic real-time information access, as well as capability to control IoT-enabled resource devices. Research gaps in the IoT include limited interfaces, networking improvements, security and privacy concerns, interoperability of devices and frameworks (often from multiple sources), and data management, for example. The gaps in mixed reality research involve limited real-world connectivity and interaction with physical objects and devices in the environment (as opposed to high interaction with virtual objects), object tracking and content anchoring, placement, occlusion, and user interaction mechanisms that still need to mature to reach desktop and mobile computing productivity and interaction standards. As such, these two domains have a complementary solution space, where a convergence can achieve the best outcomes of each.

Both the IoT and mixed reality communities (including virtual and augmented reality) have been active communities of research for decades (IoT research can be considered as extending from the work of the ubiquitous computing and wireless sensor networking communities of the early 2000s, whereas much mixed reality research stems from the mid-1990s into the early and late 2000s, culminating in modern head-mounted displays for VR and AR in the 2010's) [4,15]. Recent work, post-2016, has begun to consider the utility and synergies of merging mixed reality and IoT [21], highlighting needs for object tracking, feature information sharing, interaction services, and designs for augmented appliances. This work showed the complementarity and synergies between AR and IoT but only focused on textual information overlay, as AR manuals for appliances [22]. Further work considered the use of mixed reality for teleconferencing and presence, indicating collaborative uses [25]. Other works include [16] with a focus on gaze-based interaction with an embodied notification service character, providing an early usability study. This work is limited in terms of ability to scale to larger use cases and scenarios.

Current work, post-2019, has begun to show use cases where IoT and AR can provide visualization in situ, such as for precision farming [49], but is also not scalable or generalizable to other applications, although it proves the outdoor use case. The latest surveys on the merger of IoT and augmented reality show that the problems of IoT and AR are complementary [23], in terms of data management, viewer and display devices, and interfaces and interaction methods needed, highlighting possible gaps and opportunities, categorized as (i) AR data management, (ii) object-guided tracking, and (iii) AR-based object control and interfaces. Further, most recent forms of AR and IoT applications, such as for shopping use cases [24], present similar approaches to addressing object data management, access, control mechanisms, and content exchange across AR and IoT systems. These current applications are promising, in that they show subjects experience higher usability through AR-based IoT interactions for the shopping problem. This provides useful proofs of concept, although much more is needed to investigate the design of unique AR interfaces and embodied agents. In most recent compendiums of this new field [53], context awareness is highlighted as being critical, and there remains multiple design use cases to be investigated.

33.1.3 Chapter Overview

The above sections have highlighted the growing trend toward convergence of the IoT with mixed reality, as a synergetic and complementary paradigm that transforms IoT devices into hybrid virtual and physical objects, allowing new applications to emerge, such as XR systems that adapt to their user's context and IoT systems that present information in immersive and engaging ways. The contribution of this chapter involves (i) proposing an interdisciplinary taxonomy to consider XRI systems; (ii) an examination of XRI-related literature using the taxonomy lens; (iii) a presentation of core components, for example, XRI system designs; and (iv) a discussion of research gaps. These contributions are presented as follows: Section 33.2 proposes a new taxonomy and highlights the landscape where mixed reality and the IoT can connect, exploring the literature related to this convergence from a multidisciplinary perspective. Section 33.3 outlines the need for core XRI system architectures and frameworks and presents the authors' early designs and research addressing XRI prototypes and insights related to the taxonomy. Section 33.4 provides a discussion and presents opportunities for future research. Section 33.5 summarizes the chapter.

33.2 A Multidisciplinary Taxonomy for XRI Systems

Over the last decade, the Internet of Things has been expanding in its use of adaptive environments and thoughtfully placed sensors to communicate not just with other devices but also with the users within those environments [48, 51]. The tasks and deployment of information have always been at the forefront, thanks to devices that shared relevant information, tasks, and goals with one another. Now users are integrated more prominently within this unique form of accessing, interpreting, and utilizing adaptive environmental information [1, 28, 51]. Technology itself is becoming more ubiquitous and less out of place [35]. Context, commonly referring to any information that describes the situation of an entity, is our first element of understanding the kind of information that is being dealt with [11]. IoT information relates to this across multiple outlets [1, 9]. How that information is delivered and interpreted is often affected by the context and also raises surrounding issues, such as information privacy and security [1,51,62]. Mixed reality harmonizes both context and IoT practices with the creative use of technical devices, visual techniques, user interface design, and resulting systems [37, 50]. Overall, these factors play an important role in creating spaces that use and harmonize an environment's information.

Designing systems which must meet the domains of both the Internet of Things and mixed reality introduces the need for a multidimensional perspective on these technologies. Within this work, a high-level XRI-themed taxonomy is proposed, which relates to the following four multidisciplinary perspectives, as shown in Fig. 33.2:

- *Reality-virtuality continuum:* Milgram's taxonomy (reality-virtuality continuum) [38] highlights the need to consider the XRI system based on the expected virtualization environment, whether some class of AR, or AV, or even fully VR.
- *Presence:* The need to clearly design to maximize the degree of "presence" and immersion required for the domain of interest impacts the selection of how much mixed reality features and content to present to the viewer and how to maintain engagement and immersion within the IoT setting, without becoming distracting.
- *Human-computer interaction and user experience:* The selection of approaches that follow established best practices in user interaction design and interface design is critical to consider, to allow the user within the system to remain effective and productive.
- *Human factors:* The human factor dimension must also be considered, particularly the "human-tech" dimensions of Vicente [59], namely, the physical, psychological, social, organizational, and political gaps that must be bridged in terms of any XRI system design, in order to maintain a clear fit between the stakeholder needs and system runtime behaviors.

33.2.1 XRI Taxonomy: Thematic Literature Review

We have collected a set of recent state-of-the-art papers (Note that this is not presented as an exhaustive selection of literature but only a representative set of primarily state-of-the-art literature papers, used to show fit with the proposed taxonomy and its themes.) and highlight them in Table 33.1.

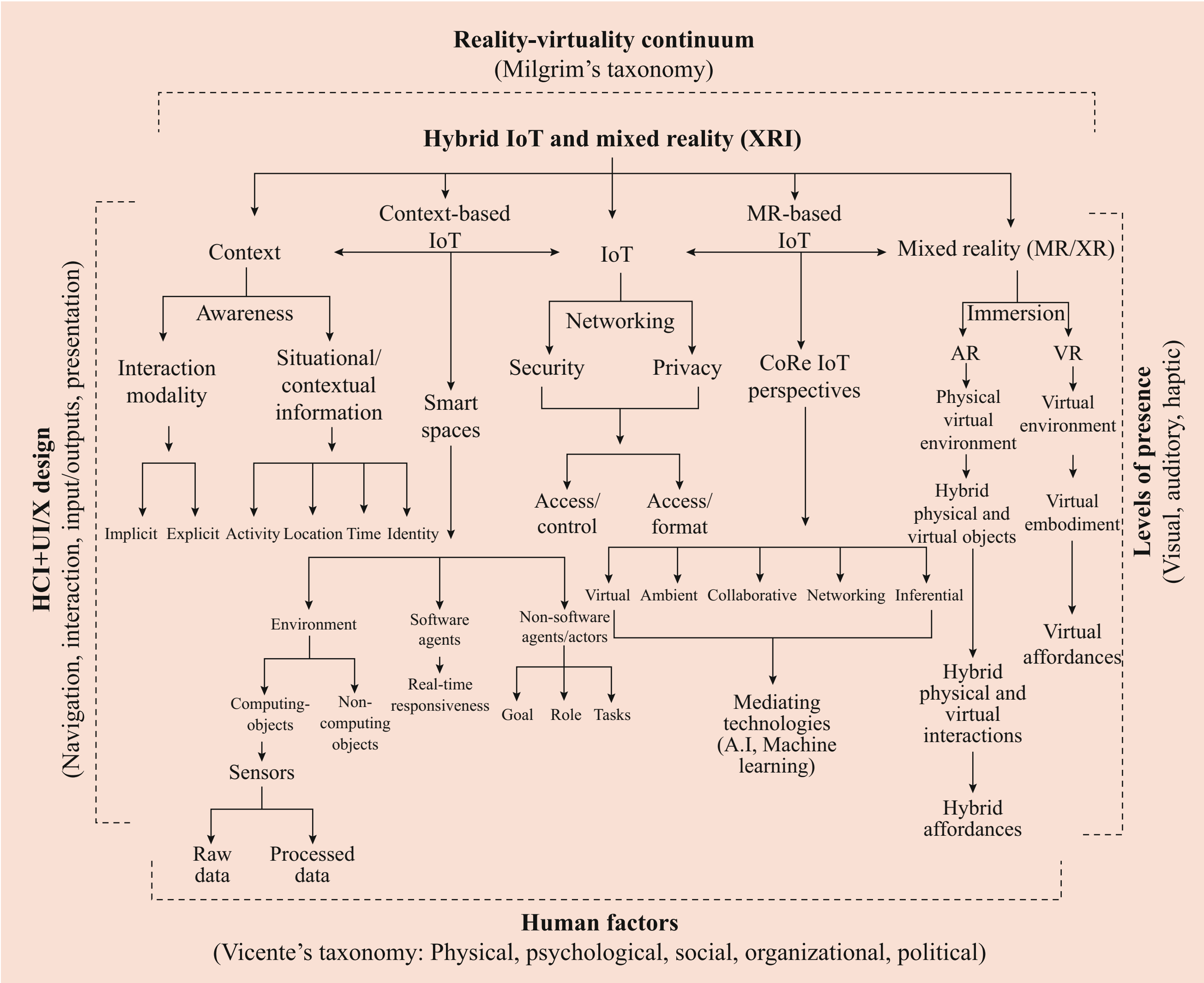


Fig. 33.2 A hybrid taxonomy diagram of concepts needed to investigate XRI systems, including context, IoT, and mixed reality factors. This considers the reality-virtuality continuum [39], human-computer interaction, human factors [59], and the notions of presence needed in XRI

We looked at the following topics that were used in the CoRe framework in Sect. 33.3: namely, context, UI design, IoT, smart spaces, mixed reality, virtual reality (VR), augmented reality (AR), human-object relationships (HOR), and human and organizational factors (HOF).

According to our preliminary research, context is being applied to specific role scenario situations (spatial, temporal, and environmental) rather than being treated as a definition [27]. Access control deals with what agents are eligible for situational-based information outside of the system's regular scenario and what types of networks could handle accessing this information while being secure and adhering to privacy [41]. UI design research highlights what activities are well suited for a given interface as an interaction component, what types of UI design translates to users within a given context, and what could be improved [3,44,45]. Other papers we have included in our collection have focused on the influence of UI design, their techniques, and how they impact users and cross-platform technologies beyond the screen [9, 12]. In HCI, this research aims to update and introduce challenges that can facilitate a discussion surrounding the uses and common patterns seen in UI design. Defining smart spaces and context broadens the possibility for ongoing investigation of what constitutes as a smart space and how definitions for various contexts, tools, and designs can help recognize its features.

IoT research continues to be evolving products of context presentations of information and how that information is handled. For example, Perera presented a decade of context-aware devices and raw sensor data in order to understand old and new middleware solutions [48]. This is still prevalent as later research is still molding a definition of today's state-

Table 33.1 Selected literature and descriptive analysis for XRI. This is a representative set of literature that is highly correlated with the taxonomy (a more thorough selection is left for future work)

Taxonomy	Selected literature	Descriptive analysis
Context	A Survey of Context-Aware Access Control Mechanisms for Cloud and Fog Networks: Taxonomy and Open Research Issues (2020) [27]	Surveys security, privacy, and access control for dynamically changing contexts. Highlights specific concerns in a wide range of contextual conditions (e.g., spatial, temporal, and environmental) and explores FOG networks and cloud computing
UI design	UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception (2020) [12]	Examines malicious UI and how app interfaces disguised interactive prompts in order to manipulate user interaction and behavior. 240 popular apps were analyzed, and 589 users were surveyed on how they perceived DPs (Malicious UI tactic: Dark Pattern). On average, 95% of apps contained one or more DPs, where popular apps contain at least seven different types. Most users do not recognize DPs but would change their behavior once they knew about them
UI design	Cross Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices (2019) [9]	Surveys and examines 510 papers in the cross-device computing domain that looks at trends and unified terminology within UI design across devices
Internet of Things (IoT)	Context aware computing for the internet of things (2015) [48]	Taxonomy of IoT context-aware devices and raw sensor data are surveyed in the span of a decade to understand old and new middleware solutions. The context lifecycle, their own taxonomy, and evaluation are presented
Smart spaces	Surveying human habit modelling and mining techniques in smart spaces (2019) [34].	Proposes comparative framework for techniques and approaches for modelling and extracting models to be employed with modern smart spaces. Taxonomy provides understanding in the suitability of a specific technique for a specific setting. Uses specific systematic action sequences to showcase delivery of goals in the form of ontology diagrams, networks, and events
Smart spaces	Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges (2016) [1]	Presents a taxonomy of IoT-based smart environments such as smart homes, health, industry, grid, city, building, and transport. The taxonomy considers the use of communication enablers, network types, technologies, local area, wireless standards, objectives, and characteristics. Challenges include managing big IoT data, investments, and security and privacy
Mixed reality	Utilizing the Mixed Reality Cube Taxonomy for Interactive Documentary Research (2016) [13]	Focuses on interactive documentaries and MR affordances as an interaction diagram. Compares literature and projects with their taxonomy for future research and understanding. Considers how interactive documentaries (IDs) will become locative with stories that progress in context and user behavior through devices such as virtual reality. Presents a framework for creating MR applications and MR development
Mixed reality	A Review on Mixed Reality: Current Trends, Challenges and Prospects (2020) [50]	Main focuses are mixed reality architecture breakdowns, where it fits in the realm of realities and addressing security and privacy measures
Mixed reality	A recent review and a taxonomy for hard and soft tissue visualization-based mixed reality (2020) [57]	Created a DVV (data, visualization processing, and view) taxonomy to evaluate and better understand the use and implementation of MR being used in image-guided surgeries (IGS) in real time. Uses taxonomy to effectively visualize and incorporate user interactions through raw sensor data and context
Virtual reality	All Reality: Values, taxonomy, and continuum, for Virtual, Augmented, eXtended/MiXed (X), Mediated (X,Y), and Multi-mediated Reality/Intelligence (2018) [36]	Evaluates the multimedia forms of all reality including extended reality, VR, AR, MR, etc. and their rights yet poor responsibility of security, privacy, and trust. Looks into humanistic intelligence as an important inclusion of "ethically aligned realities" through an evolution of all-reality diversity scope and timeline

(continued)

Table 33.1 (continued)

Taxonomy	Selected literature	Descriptive analysis
Virtual reality	A Systematic Review of a Virtual Reality System from the Perspective of User Experience (2020) [29]	Explores the evaluation of UX within VR. Proposes a systematic taxonomy based on HCI and UX research. 393 articles were collected, and 65 were selected via systematic reviews and meta-analysis methodology. Provides a definition for VR. Focuses on types of input and output devices, hand or non-hand interactions, and the issues and challenges that arise
Virtual reality	Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions (2015) [43]	Proposes a taxonomy on virtual reality systems and mental immersion. Uses the CAVE system as the perspective's foundation of low-high forms of virtual reality immersion. Not much research has been done on CAVE immersion due to costs
Augmented reality	Application of augmented reality technologies for preparation of specialists of new technological era (2020) [18]	Discusses the importance of training new professionals into newly assimilated AR technologies for education and future work opportunities. More research on how to use AR technologies in this context is required. AR industry, education, cost, and research are the main factors of proper information assimilation
Augmented reality	Visualization Techniques in Augmented Reality: A Taxonomy, Methods and Patterns (2020) [63]	Provides a taxonomy on existing AR visualization techniques from 67 existing works on the subject. Through this taxonomy, the authors provide effective ways of applying visual techniques for AR environments
Human-object-relationship (HOR) and human and organizational factors (HOF)	Investigation of Human Factors Using HFACS Framework—A Case Study for Unintended Reactor Trip Events in NPP(2020) [26]	Discusses HFACS (Human Factors Analysis and Classification System) and how it can be used to classify and understand human error in complex work systems such as the nuclear industry. A taxonomy of previously reported accidents is categorized in HFACS and identified by an expert. This causes limitations for basic understanding of errors that the paper iterates the need for further study due to liability or untruthful reports of accidents
Human-object-relationship (HOR) and human and organizational factors (HOF)	The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems (2019) [10]	Proposes a taxonomy that extracts knowledge from human-in-the-loop machine learning (ML) for hybrid intelligent systems (the collaboration of human intelligence and AI mechanics)

of-the-art smart space and its use of IoT through comparing frameworks and techniques [1, 34]. The uses of mixed reality have become integrated with visually deploying context-related information and fall between the mediums of AR and VR technologies [36]. As the sharing of access control for situational-based information becomes vital to larger groups and systems, the use of mixed reality and its components becomes a topic of discussing cost, assimilation, and trust [29, 36]. Most frameworks and surveys looking at MR, VR, and AR mainly try to articulate and educate researchers of what these mediums are used for, the use of their interfaces, and immersive properties and place them within a visual-immersion reality timelines that fit the visual and interaction expectations of the reality [18]. They also recognize the challenges each reality poses and prompt discussions of their use in comparison to realities that do not immerse virtual worlds with real ones like AR, VR, and possibly concepts like extended reality [13, 29, 36].

The human-object-relationship focuses on fabricating a language that is understood among agents when designing interfaces and tools to deploy certain tasks [7, 8, 17, 59]. These works have tried to paint a picture of what our relationships with tools, objects, and interfaces are like. They have also provided us with concerns in relation to the use of complex systems and our habit of being prone to human error. The political, social, and economic factors of these have made way for the considerations of HFACS (Human Factors Analysis and Classification System) and human and organizational factors to help compensate human error through machine learning (ML) and artificial intelligence (AI) [26].

In terms of XRI, this selected literature has been considered, according to the following themes, as seen in Table 33.2: (i) the degree to which conceptual models related to XRI are described, (ii) the degree to which software and sensor use related to XRI is described, (iii) the degree to which the level of interaction related to XRI is described; and (iv) the degree to which networks and protocols related to XRI are described. This provides a comparative look at this broad literature.

Table 33.2 A comparison of the selected literature related to the XRI taxonomy

Selected papers	Conceptual models related to XRI	Software and sensor use related to XRI	Level of interaction related to XRI	Network and protocols related to XRI
Kayes et al. [27]	Taxonomy of contextual conditions; Taxonomy of authorization models	Cloud computing; Context-sensitive access control mechanisms; IoT devices; Cloud-services	Fog and Cloud interactions to support IoT; Global Data Centre. Fog servers organizing access to app, services and contents	Context aware access control (CAAC) protocols
Di Geronimo et al. [12]	Taxonomy of Dark Patterns applied to mobile apps	30 apps for selected 8 categories	E-commerce websites that trigger buying; classification of dark patterns implemented for mobile apps. Walk-through protocol of navigating apps (from logging in to using the app to logging out) and reviewing user experience	No specific network or protocols related to XRI highlighted
Brudy et al. [9]	Taxonomy of cross-device design space dimensions. Ontology of cross-device research terminology	Tangibles; IoT devices; wearables; AR/VR headsets; smart glasses; networked devices	People to device relationships. Interaction techniques for cross-device computing: Configuration, content engagement and disengagement	Cross-device; multi-device; distributed in addition to specific interaction formats
Perera et al. [48]	Taxonomy of functionality commonly supported in existing research prototype systems. Context-Life Cycle	Sensors; actuators; middleware solutions in relation to IoT. Effective and efficient sensing and communication of data. Level of context awareness	Event detection: machine-to-machine (M2M) and machine-to-person communication. Context discovery and annotation. Security and Privacy within IoT. Context reasoning and querying	Mobile data processing between low- and high-end computational devices
Leotta et al. [34]	Comparative SLR (Systematic Literature Review) framework. Taxonomy-based Ontology Models for Contextual and Activity classes and properties	Embedded Systems; powerful computing devices; sensing technologies and ambient intelligence (AmI)	Sensor measurements of user behavior in the environment. Modeling human habits and environmental dynamics. Recognizing context at runtime and acting within the environment	Contextual data modelling framework of environments (or humans)
Ahmed et al. [1]	Taxonomy of IoT based smart Environments	Sensing and communication technologies; data fusion; emerging computing and information security	Minimal user interaction as an IoT objective	Communication technology, enablers, wireless standards, and network types. Software-defined network controllers (SDN)
Fisher [13]	The MiRA Cube Taxonomy; MRID framework	Mobile devices; sensor tracking; HoloLens; AR object. VR headsets	Active adaptive interactions, passive adaptive interactions. Level of interaction is classified between weak agency and strong agency. Expansive interactions (narratives)	Social networks where user collaboration is the primary network domain
Rokhsaritalemi et al. [50]	General Architecture of a Mixed Reality System	Mixed reality software, toolkits, and application programming interface	Mixed reality UI interaction methods. Security and privacy approaches for MR (user interface protection)	Architecture of mixed reality systems and their issues; security and privacy mechanisms. Mixed reality network-oriented applications
Tuladhar et al. [57]	Data, Visualization processing and View (DVV) taxonomy to evaluate current Mixed Reality systems. Conceptual system overview following a model-view-controller paradigm	Software-based telesurgical visualization. Mixed reality environments (MREs) for a surgical scenario	Component view attributes and instances; interaction tools; hardware and data manipulation. Hand gestures, audio input, and sensor-enabled surgical tools. Display devices; HoloLens; other displays	No specific network and protocols related to XRI highlighted

(continued)

Table 33.2 (continued)

Selected papers	Conceptual models related to XRI	Software and sensor use related to XRI	Level of interaction related to XRI	Network and protocols related to XRI
Mann et al. [36]	Mediated Reality (X,Y) Continuum Taxonomy	Sensory attenuation technologies. Multimediated reality. Technologies; multi-veillant technologies	Extended human sensory perception. Multisensory synthetic synesthesia. Human-in-the-loop AI	No specific network and protocols related to XRI highlighted
Kim, et al. [29]	UX framework of a VR system. Device taxonomy in VR. User activity-related component taxonomy in VR. Environment taxonomy in VR	VR Engine. HMD. Non-hand input with tracking. Hand input with tracking	Single users interact within the virtual world and environment. Multiuser interaction within the virtual world and environment (co-located) or remotely. Movable/riding environment	No specific network and protocols related to XRI highlighted
Muhana [43]	A taxonomy of virtual reality systems	Cave Automatic Virtual Environments (CAVE). Tracking sensors. Sensory feedback and sensory immersion	Direct manipulation: (1) direct user control, (2) physical control, (3) virtual control, (4) agent control. Navigation in CAVE. Meta-commands	No specific network and protocols related to XRI highlighted
Iatsyshyn et al. [18]	No conceptual models highlighted	Visual design software programs; AR technology to help support STEM learning	Education: visualization of teaching materials. Blending AR technology with learning strategies in STEM; social interaction via communication, entertainment, and games	Social networks for communication in AR highlighted
Zollmann et al. [63]	Taxonomy of AR Visualization Techniques	Camera image. Registration data. Geometric data. Context data. Coordinate systems.	Visual cues and spatial relationships	AR visualization pipelines for extracting image-based physical cues, model-based physical cues, external virtual cues, mapping distance to appearance, information filtering
Karthick et al. [26]	HFACS framework; organizational influences; unsafe supervisions; precondition for unsafe acts	No software and sensor use related to XRI	No specific levels of interaction related to XRI	No specific network and protocols related to XRI
Dellermann et al. [10]	Hybrid Intelligence Systems; Taxonomy of a development method. Taxonomy of hybrid intelligence design	Business applications of hybrid intelligence systems	Human-in-the-loop mechanisms. Socio-technology. Human to AI interaction; machine teaching. AI-human interaction. Machine feedback and interpretability	No specific network and protocols related to XRI highlighted

In terms of the XRI taxonomy, the context literature identified approaches to analyze and access contextual sources of information through Context-Aware Access Control (CAAC). With security and privacy being one of their main focuses, they address ad hoc networking among IoT devices and cloud networking services for better access and use of organizing contextualized information.

The user interface design literature dealt with pattern configuration and cross-device configuration that looked at how users interact with a device's UI and how the devices themselves interact with each other. Malicious UI research explored how behavior is influenced when users engage with mobile apps, whereas cross-device research presented intricate interactive relationships not only with users but with multiple devices in social group settings.

The Internet of Things literature presented taxonomies and frameworks that supported prototypes and systems related to context-aware IoT middleware solutions. The paper(s) highlight the steps and levels of context awareness through their proposed taxonomies and frameworks to address their appropriate uses and actions within IoT. The smart space literature presented comparative taxonomies and frameworks that highlight systematic protocols of how a smart space is designed for specific and unique settings. The use of information and processed data from raw data is applied in different smart space scenarios and shared technologies. Security and privacy are also discussed in terms of IoT data and access within the context of smart spaces.

The mixed reality literature presented taxonomies that delve into the understanding of mixed reality environments, where the architecture of MR is evaluated in great depth through comparative literature. Mixed reality agents scaling from weak to strong agency, with both virtual and physical corporal presence, were presented. The process of visualizing raw data was examined, exploring the use of sensors and software applications, and examples of interaction methods were provided. The virtual reality literature presented evaluation taxonomies of VR along with other forms of reality such as augmented reality and extended reality, alongside definitions, what constituted as a VR environment, and its forms of interaction and engagement. The VR taxonomy highlights topics of context intelligence, input methods within VR spaces, and environmental characteristics in comparison with automated virtual environments (CAVE) technology. The augmented reality literature explored pipelines for visual techniques that can be implemented in AR for better use and understanding of its virtual properties in physical spaces. The taxonomies and frameworks presented also shows how users with little to no background or understanding with AR technologies could learn and apply it within a variety of fields including educational or industry domains.

The human-to-object relationship and human organizational factors literature identified behavioral influences that cause human error and unsafe acts through the evaluation of the human factors. The literature also examined solutions of how to combat human error by proposing hybrid systems where humans and machines learn from one another and share the load of task management.

33.2.2 Relating the XRI Taxonomy to XRI System Design Needs from a Multi-disciplinary Perspective

This proposed XRI taxonomy is used in this chapter to cross-reference literature in the above domains. In order to justify these categorizations as useful to XRI systems, the following relationships between the XRI taxonomy, its multidisciplinary perspectives, and XRI system design needs are here presented, followed by a thematic literature review:

In Terms of the Reality-Virtuality Continuum *Context* – XRI must account for context management across the variety of displays and technologies within the RV Continuum. *UI Design* – XRI must apply appropriate UI designs for navigation, interaction, inputs/outputs, and presentation of content in a manner that fits the display level and technologies within the RV continuum. *IoT* – XRI must address the networking, edge device sensing and control, access control, privacy, and application design constructs while streamlining to fit the level of display and technologies within the RV continuum. *Smart Spaces* - XRI-deployed systems must fulfill the needs and requirements of smart space designs, incorporating sensing, pattern recognition, and device control together with appropriate visualization and interaction with objects at the level of the technologies within the RV continuum. *Mixed Reality* – XRI must address the requirements of the mixed reality spectrum, accounting for both AR and AV, but not reality and full virtuality. *Virtual Reality* – XRI must address the requirements of the virtual reality spectrum and its hardware, displays, inputs, and design requirements. *Augmented Reality* – XRI must address the requirements of the augmented reality spectrum and its hardware, displays, inputs, and design requirements. *Human-Object-Relationship & Human-organizational Factors* – XRI must consider the human-factor interrelationships between humans in the loop and the objects in their environment, especially when these objects are IoT edge devices, and the XRI system must account for social networked interrelationships between humans-humans, humans-agents, and agent organizations.

In Terms of the Level of Presence *Context* – XRI must apply contextual information related to a person's situational state in a way that preserves or increases the level of presence, using visual, auditory, or haptic technologies. *UI Design* – XRI must apply UI designs for navigation, interaction, inputs/outputs, and presentation of content in a manner that preserves or increases the level of presence of the user. *IoT* – XRI must apply IoT networking, edge device sensing and control, access control, privacy and application design constructs in a manner that preserves or increases the level of presence of the user. *Smart Spaces* – XRI-deployed systems must fulfill the needs and requirements of smart space designs while maintaining appropriate levels of presence. *Mixed Reality* – XRI must consider the level of presence of users within the system across the entirety of the mixed reality spectrum (AR and AV, but not focused solely on either reality or virtuality). *Virtual Reality* – XRI must address the requirements of the virtual reality end of the spectrum while preserving the level of presence of users within the application. *Augmented Reality* – XRI must address the requirements of the augmented reality portion of the spectrum while preserving the level of presence of users within the application. *Human-Object-Relationship & Human-organizational Factors* – XRI must consider how to retain the presence of users within the system while also accounting for the human-factor interrelationships between humans-in-the-loop, and the objects in their environment, and the specific kinds of networked interrelationships between humans-humans, humans-agents, and agent organizations.

In Terms of Human-Computer Interaction and User Experience *Context* – XRI system interfaces must design for managing and interaction of context information, accounting for the appropriate navigation, interaction, inputs/outputs, and presentation of this situational information. *UI Design* – XRI system interfaces must apply UI design principles and best practices for navigation, interaction, inputs/outputs, and presentation of content to its users, accounting for HCI needs within the XR and IoT domains. *IoT* – XRI system interfaces must recognize IoT networking, edge device sensing and control, access control, privacy and application design and interaction needs while maintaining user interaction requirements within both XR and IoT domains. *Smart Spaces* – XRI system interfaces must be designed to address the needs of smart spaces that incorporate sensing, pattern recognition, device control, visualization, and interaction while maintaining standards from within both XR and IoT domains. *Mixed Reality* – XRI system interfaces must be designed to fulfill visualization and interaction requirements from both XR and IoT domains, accounting for both AR and AV, but not focused solely on either reality or virtuality. *Virtual Reality* – XRI system interfaces must be designed to fulfill visualization and interaction requirements from both XR and IoT domains, accounting for virtual reality devices and interaction technologies. *Augmented Reality* – XRI system interfaces must be designed to fulfill visualization and interaction requirements from both XR and IoT domains, accounting for augmented reality devices and interaction technologies. *Human-Object-Relationship & Human-organizational Factors* – XRI system interfaces must consider appropriate human factor interrelationships between humans-in-the-loop and the objects in their environment while accounting for HCI and UX standards.

In Terms of Human Factors *Context* – XRI must consider the human factor requirements when acquiring and managing situational context information, from the individual psychophysiological through to social, organizational, and political lens. *UI Design* – XRI must consider human factor requirements when negotiating appropriate system interaction and interface design practices and guidelines. *IoT* – XRI must consider human factor requirements that overlap with IoT networking, edge device sensing and control, access control, privacy, and application designs for both XR and IoT domains. *Smart Spaces* – XRI must consider human factor requirements within smart space designs, accounting for multidimensional factors throughout the sensing, pattern recognition, device control, visualization, and interaction needs of the XR and IoT system. *Mixed Reality* – XRI must consider the human factor requirements that impact the mixed reality spectrum, accounting for both AR and AV, but not focused solely on reality and virtuality. *Virtual Reality* – XRI systems must consider human factor requirements that impact the virtual reality end of the spectrum, from both XR and IoT perspectives. *Augmented Reality* – XRI systems must consider human factor requirements that impact the augmented reality end of the spectrum, from both XR and IoT perspectives. *Human-Object-Relationship & Human-organizational Factors* – XRI systems must consider the human factor dimensions (such as physical, psychological, social, organizational, political) while accounting for the kinds of interrelationships between humans-in-the-loop, and the objects in their environment, and kinds of networked communication, between humans-humans, humans-agents, and agent-agent organizations.

In summary, this section has proposed a taxonomy and presented a small selection of the recent state of the art to determine what is being done with current IoT, context and MR-related projects, frameworks, and taxonomies – a snapshot of this rapidly unfolding XRI landscape. The authors agree with other state-of-the-art studies, such as [24], wherein the key themes for XRI systems involve data management, object tracking, object control, and interface directions. Additions to these directions include those identified in the above sections, particularly the importance of context-aware frameworks and human factor-aware designs.

33.3 Contextual Reality (CoRe) Frameworks for XRI System Design

Contextual reality (CoRe) is an architectural IoT framework, proposed by the author in [41], that leverages multiple IoT services and devices to be applied within an environment's contextual setting. It considers privacy and security needs within adaptive context-aware environments and is designed to incorporate both users and devices to be aware of the information they are generating and receiving – this relationship is described as "humans-in-the-loop for IoT". It applies environmental context to IoT controllers and appropriate system visuals and interfaces for humans-in-the-loop [41]. CoRe is mainly responsible for harmonizing both humans-in-the-loop and things-in-the-loop, as it considers agent-agent communication within an environment's context, as well as the mixture of human-human, and human-agent communications.

The CoRe research platform, shown in Fig. 33.3 in its earliest conceptual configurations, extends the IoT environment by adding an IoT visual interface, particularly for head-mounted displays, via 3D engines, like Unity3D. This allows for 3D scene objects, 3D interface objects, audiovisual notifications, animations, text messages and labels, AR/VR scene property changes, 3D object overlaid on the environment, and even 3D actor agent behaviors to be applied to the user's visual field. Behind the scenes, however, this can be supported by IoT real-time processors, such as those which classify input data streams, such as visual or audio input (which could involve object recognition, inference processing, or audio recognition processing). This may also engage with IoT object interaction services across cloud platforms for cloud data analytics, or local analytics such as through trained IoT inference devices. Further, to enable a decentralized convergence of this information, a multi-agent platform is considered, as a means toward dynamic sense-making (such

33

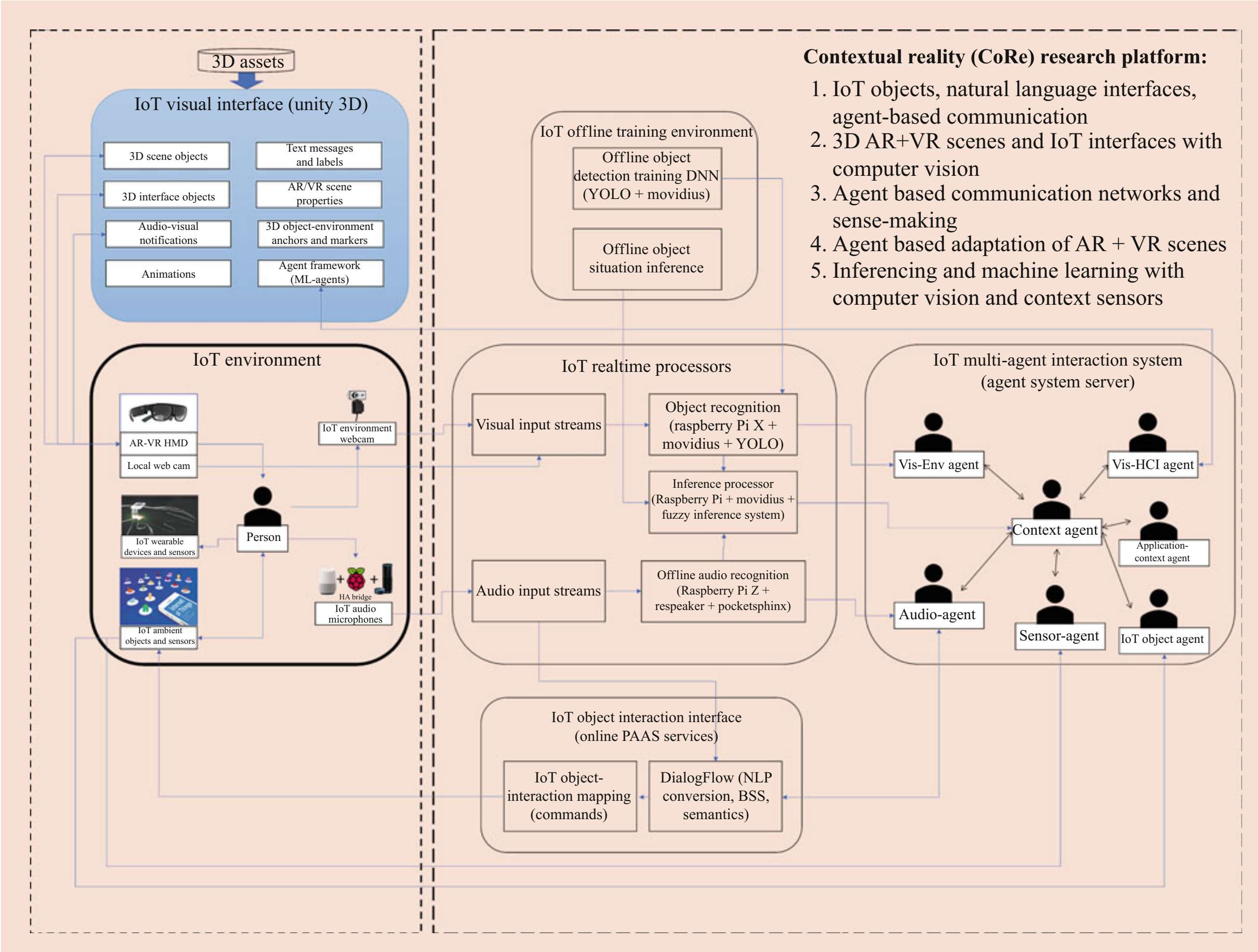


Fig. 33.3 Early concept of a CoRe XRI framework as a general research platform for XRI. This indicates some of the hardware and software components and possible configurations to achieve XRI

as having a vision agent, audio agent, context agent, visual interaction agent, application context agent, external sensor agent, or other IoT object agent; all conferring with a context manager agent). Together, this outlines a capacity toward environments that bring together IoT objects, interface agents, multi-agent communication, 3D mixed reality scenes with IoT interface components, and inferencing for sense-making.

CoRe comprises six main perspective types to target for this adaptive nature of information processing: *virtual, ambient, collaborative, informational context, inferential, and networking*. For XRI interfaces, the focus is on the virtual, ambient, collaborative, and informational context perspectives primarily, as those which are directly impactful to users. As discussed in [51], the virtual perspective refers to 3D object models and agent behavior components; the ambient perspective refers to the IoT objects and components physically available within the local environment and the interaction between these and any platform services or peripheral devices and conventional displays; the informational context perspective refers to the dynamics and contextual properties of data for IoT devices and presentation content to humans-in-the-loop. The virtual handles the visuals for the humans-in-the-loop, while the ambient deals with the collaborative use and application of IoT devices.

The CoRe framework, as in Fig. 33.4, considers the overlapping concepts that an XRI system would need in practical terms, which includes not just mixed reality toolsets, like 3D frameworks and object models, scenes, and IoT frontend devices and controls, but also backend services, like model servers for learning and data analysis, platform-as-a-service (PAAS) cloud frameworks for informational context, and the overall networking middleware necessary. Together, such a broad, holistic approach allows for the creation of smart XRI environment prototypes.

Systems that aim to apply these notions must consider the extension to the conventional IoT stack, as seen in Fig. 33.5, where in addition to low-level hardware, networking, and software application designs for the IoT, there is a need for

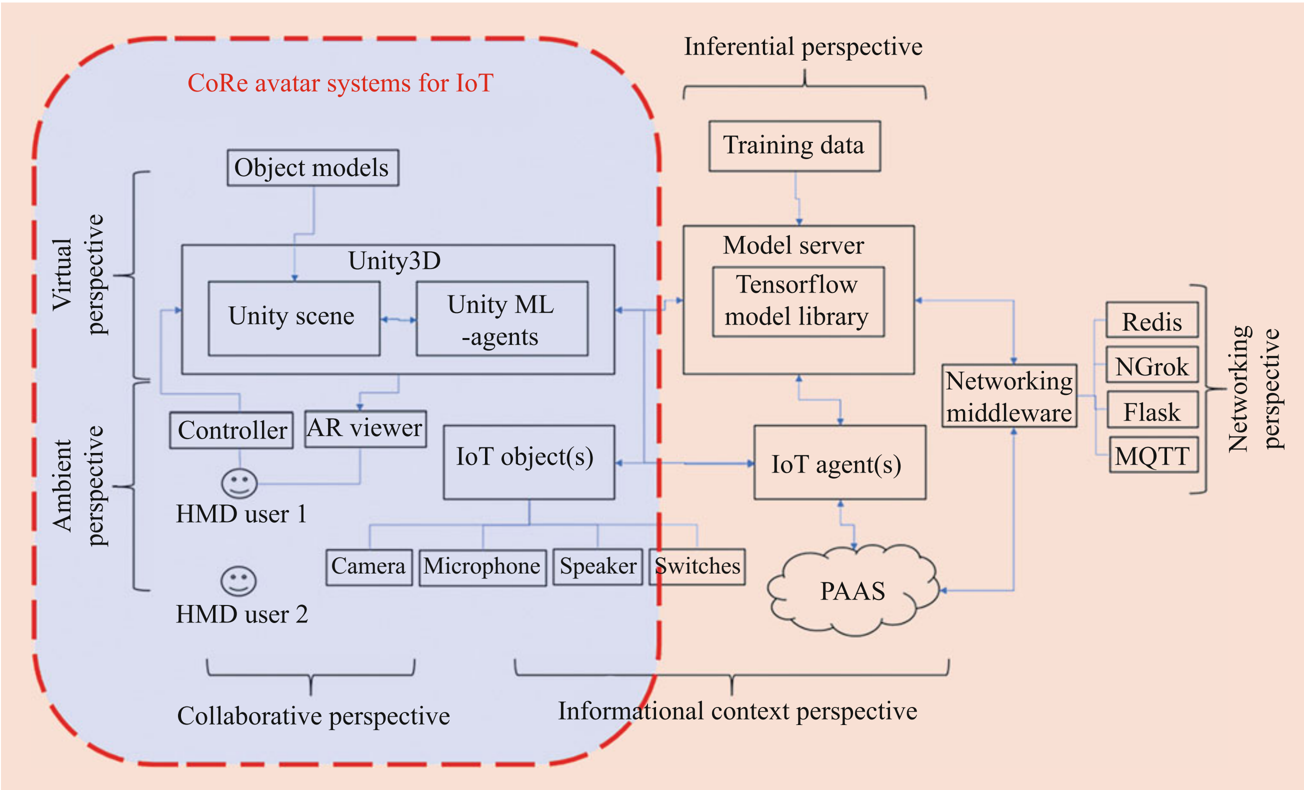


Fig. 33.4 CoRe IoT Framework for adaptive context aware IoT and mixed reality visualization [41, 51]

both agent system logic and agent avatar presentation. These two higher layers define the behaviors of XRI projected content and interfaces, as well as the design of engaging and immersive representations of IoT data objects – designed especially to interface with users of the IoT with both classical 2D and 3D interface widgets – and design metaphors that are similar to those of conventional game design dashboards and character controllers. These IoT avatars [42, 51] speak to the potential to extend the IoT interaction in new directions at the foreground of visible interface, aimed at improving the relationship between humans-in-the-loop and hybrid IoT and mixed reality objects, as in the simple scenario in Fig. 33.6.

33.3.1 Designing Proof-of-Concept XRI Systems with CoRe

This section highlights existing proof-of-concept designs in [41], highlighting the author's approaches to extend the traditional pillars of IoT architecture designtoward expanding the relationship between users within smart IoT environments and the IoT physical edge devices around them, in addition to hybrid representations, real- and mixed reality-enabled IoT objects, and even the physical environment space. This concept has been presented in [42].

In this direction, two prototype IoT avatars have been considered, for the plant scenario, and represent the opportunity to transform IoT interfaces from being in traditional 2D dashboard form factors, or raw graph data, into 3D object character visualizations, with potential to incorporate designs for IoT object control through the control and interaction with a virtual character as the primary form of the interface. This is a direction where a mixed reality overlay can provide both conventional contents anchored to a physical object while also adding a new interactive dimension to the IoT-enabled object. When this anchoring is in place appropriately, allowing for engaging with the physical, informational, and virtual representation, the hybrid object can be considered as an XRI instance. This requires the design of the system to account for both IoT device control and interface and virtual avatar level control and interface design and runtime behaviors. Further, as seen in Figs. 33.7 and 33.8, avatar expressiveness can be used to enhance the level of immersion and engagement with the underlying IoT object, for both handheld, mobile phone-based experiences and head-mounted displays for more immersive presentations.

However, it is noted that the degree to which this benefits the IoT interaction experience is a promising avenue for further study, among others. Figure 33.9 highlights several of these areas for exploration and enhancement of IoT avatar interfaces for hybrid XRI objects. This includes the need to

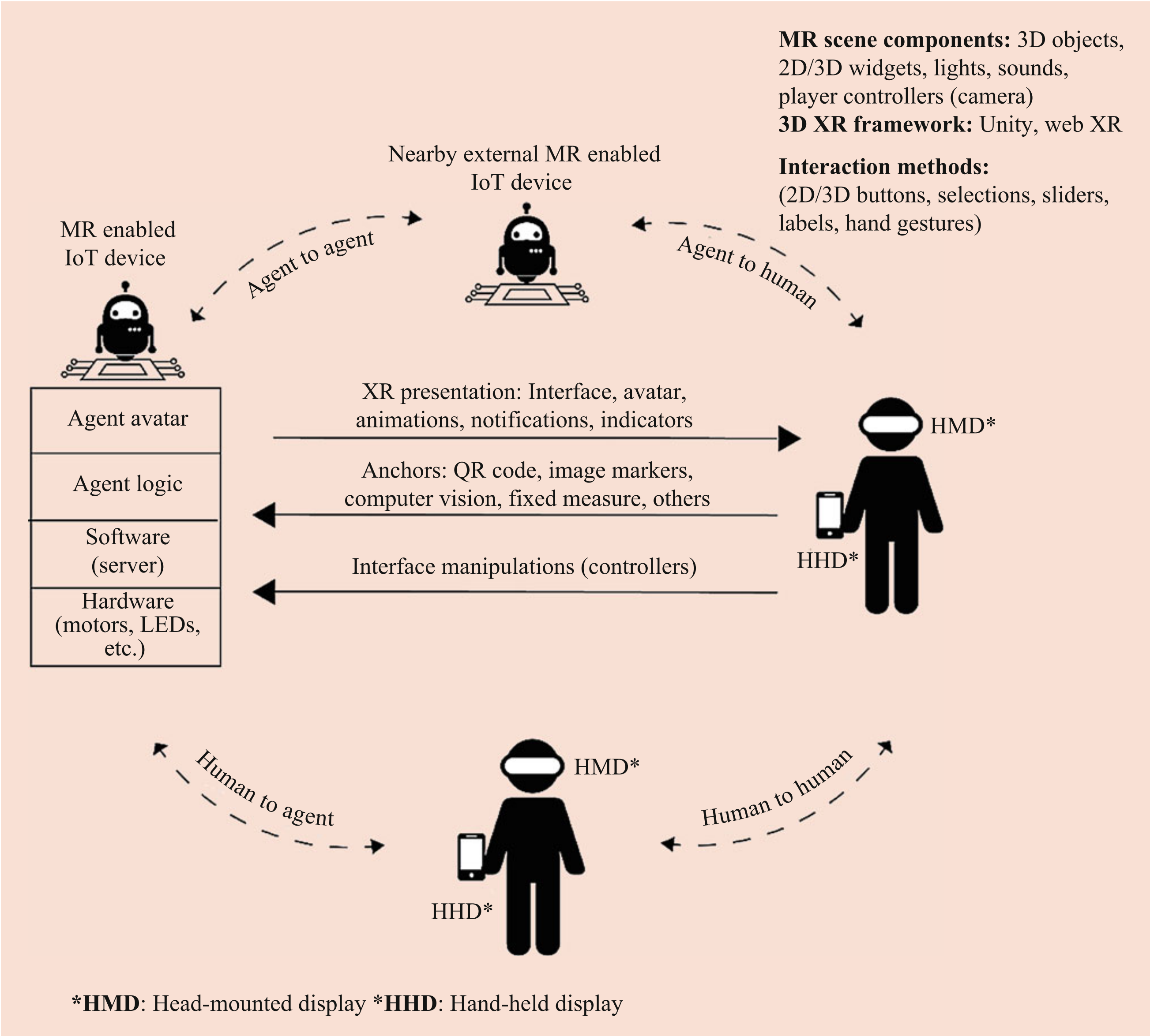


Fig. 33.5 Extending the IoT stack with XR presentation and agent avatar layers and themes for a multi-agent XRI system [51]. Note this introduces multiple interactions to consider between humans-in-the-loop, the IoT system, and IoT avatar agents

extend (i) the conversational chat interaction features; (ii) the visualization properties related to the design of the avatar, the contextual intelligence driving the avatar behavior, and interface presentation; (iii) experience design of tasks and interactions that fit the avatar character; and (iv) domain-specific scenario designs. Together, these outline the themes for considering new XRI system designs while maintaining a focus on the themes identified in the overall taxonomy highlighted above, namely, the need for smart space considerations, interactions and behavior design, context-driven artificial intelligence, and human-centered considerations.

33.3.2 CoRe Framework Insights in Relation to the XRI Taxonomy Perspectives

The above CoRe Framework architectures have been used to derive insights related to the XRI Taxonomy and the perspectives identified. Some of these insights are highlighted below, as a means toward reflecting how the multidisciplinary needs of CoRe XRI system are related to the XRI taxonomy:

In Terms of the Reality-Virtuality Continuum *Virtual Perspective* – CoRe XRI systems may be implemented across

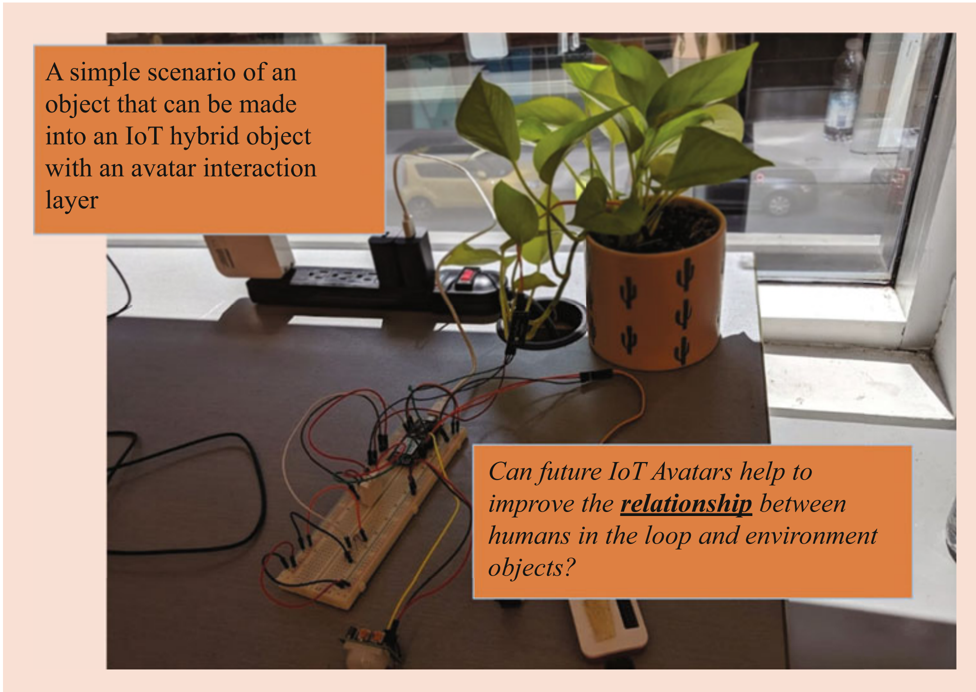


Fig. 33.6 A plant avatar scenario for exploring the CoRe architecture designs

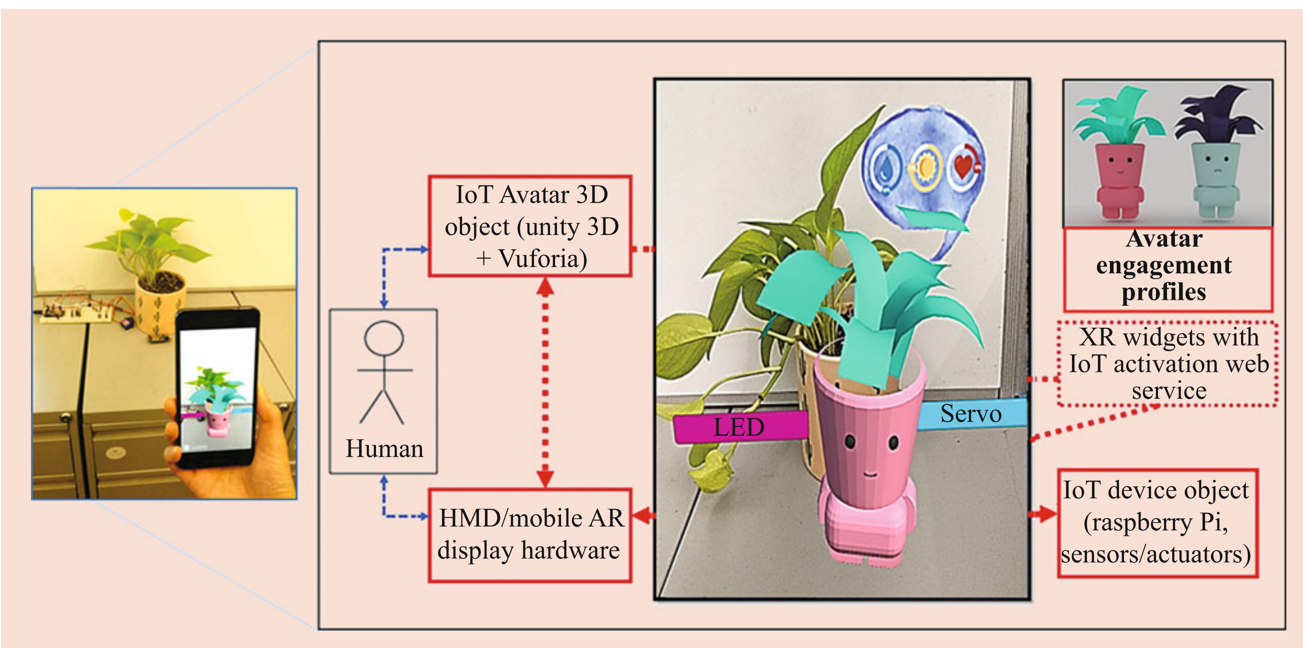


Fig. 33.7 An early handheld IoT avatar for the plant scenario, from [51]. This allowed for considering how to enhance an IoT implementation with an engaging representation and the functions of mobile AR in this domain

33

the RV spectrum, depending on the selected display device, and can be both head-mounted and handheld designs. Smart glass displays are considered. *Ambient Perspective* – CoRe XRI system designs are composed of hybrid virtual and physical objects in the selected runtime environment and require consideration of the application's target displays across the RV spectrum in order to bridge virtual and physical objects and edge devices. *Collaborative Perspective* – CoRe XRI system designs may consider the potential for multiuser to device interactions and collaborative use cases, where

Fig. 33.8 A more recent plant avatar with head mounted display, from [42]. This allowed for extending the immersion and presence of the IoT avatar while also considering expressiveness and emotion as a form of communication

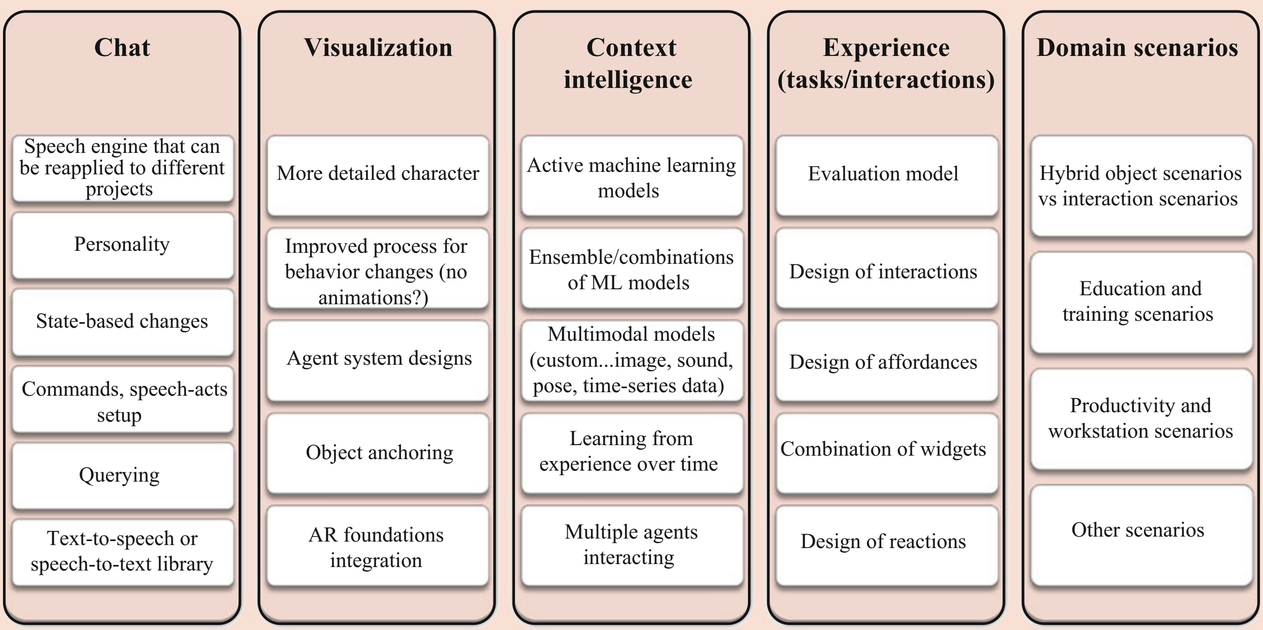


Fig. 33.9 Extending the early CoRe prototypes will require examining the above dimensions, including chat, visualization, intelligence, experience design, and domain requirements

appropriate visualizations are needed that fit the application user's displays across the RV spectrum. *Informational Context Perspective* - CoRe XRI systems require components to acquire, monitor, update, and respond to changes in user situational context information, as well as application or object context information, where hybrid virtual and physical IoT objects are deployed in the environment; this must also account for the appropriate representation of these context changes at the level of visualization across the RV spectrum. *Inferential Perspective* – CoRe XRI system designs may apply learning and inference in order to adapt system visualizations appropriately within the RV spectrum, according to the user's display; additionally, inference may depend on understanding the user's relationship to virtual and physical objects and other users. *Networking Perspective* – CoRe XRI system designs must have conventional IoT networking supports to address the variety of communication within a hybrid virtual and physical environment; this communication may be appropriately visualized also.

In Terms of Levels of Presence *Virtual Perspective* – CoRe XRI systems involve 3D object visualizations and rendering engines fit for the display; however, these should focus on maintaining high levels of presence in order to fit within and respond effectively to the environment. *Ambient Perspective* – CoRe XRI systems must be designed to have appropriately situated visualizations and control interactions that fit the deployed environment, increasing presence, and alignment with edge devices, objects, and users within the environment. *Collaborative Perspective* – CoRe XRI system designs should consider multiuser collaborative scenarios within the environment and the needs to maintain presence when hybrid objects are shared across user visualizations and interactions. *Informational Context Perspective* – CoRe XRI systems should consider contextual information from the perspective of multiple users when collaborative scenarios are applied, in order to maintain presence through appropriate visualizations and interactions that depend on each user's source of context information or devices. *Inferential Perspective* – CoRe XRI systems may apply learning and inference techniques based on acquired information, and responses based on this inference require appropriate visualizations that maintain a user's presence according to the hybrid object. *Networking Perspective* – CoRe XRI systems require backend networking components, which may not require visualization; however, forms of communication between devices may require appropriate representations and response behaviors of the hybrid object, in order to maintain presence.

In Terms of Human-Computer Interaction and User Experience *Virtual Perspective* – CoRe XRI system components must apply HCI design and UX guidelines appropriate for virtual as well as physical IoT objects and edge devices, which match the virtuality paradigm. *Ambient Perspective* – CoRe XRI system designs must address HCI and UX requirements that fit the user's environment and which address interactions across IoT edge devices, as well as virtual objects. *Collaborative Perspective* – CoRe XRI systems may require HCI and UX components that are multiuser focused, highlighting potential shared interactions and display widgets. *Informational Context Perspective* – CoRe XRI system interactions and interfaces should be responsive to user situational context and object device situational context, reflecting these appropriately within the interface or hybrid object. *Inferential Perspective* – CoRe XRI system interactions and interfaces need to reflect changes in user information and appropriately adapt to and infer future interactions, following appropriate design guidelines. *Networking Perspective* – CoRe XRI system interactions and interfaces must incorporate networking into its design, allowing for appropriate communication types to be addressed, such as human-human, human-agent, or agent-agent communication, within the interface.

In Terms of Human Factors *Virtual Perspective* – CoRe XRI system components must consider appropriate human factor needs for visualizations and interfaces, such as physical, psychological, social, organizational, and political needs where virtual objects are concerned. *Ambient Perspective* – CoRe XRI systems must consider human factor designs related to the environment where the system is deployed, for both objects and other humans-in-the-loop. *Collaborative Perspective* – CoRe XRI systems must account for human factor designs related to the interrelationships between users within the system and other system objects. *Informational Context Perspective* – CoRe XRI systems must address human factor needs related to information context and the life cycle of contextual information, including acquisition, preservation, sensitivity, security, and access of information, among others. *Inferential Perspective* – CoRe XRI systems must allow for inference that fit the user application while balancing the human factor needs related to both the input information and the presentation and representation and effective communication of inferred knowledge regarding users within the system, or other system objects. *Networking Perspective* – CoRe XRI systems must account for human factor needs related to communication and networking across user devices, voice communications, gestures, and conventional message passing, within and across subsystems, or middleware services.

Together, the above highlights how a multidisciplinary lens within the XRI taxonomy relates to system design requirements and objectives toward rich hybrid virtual and physical IoT environments.

33.4 Discussion

There remains a gap for deeper exploration into techniques for the use of mixed reality to enrich human-computer interaction within the new wave of Internet of Things-enabled environments. Possible avenues of research to address these themes consider questions like the following: (i) How best to build such XRI systems well, and how to evaluate them? (ii) How to make such systems relatable and practical for consumers and humans-in-the-loop dynamically? (iii) How to ensure that such system designs can have a positive impact on society? (iv) How to standardize XRI avatar designs for hybrid virtual and physical IoT objects and environments using wearable and wireless devices? Lastly, how to extend these designs across indoor, outdoor, and vehicular usage scenarios? Such questions demand a variety of integration solutions, prototypes, user testing, and real-world use case exploration.

This indicates an opportunity to explore contextually relevant and adaptive visual interfaces for IoT applications. In doing so, there are critical interface and assistance challenges in IoT where frameworks for immersive adaptive environments are needed. This includes design and creation of (i) mixed reality interactive content, widgets, and processes for information visualization and understanding of underlying IoT data; (ii) deep computer vision models and situational inferencing for environmental understanding of IoT situations; (iii) context-aware agent techniques that adapt responses to users in XRI-enabled environments; (iv) approaches that foster and consider human factors, like privacy and trust, in the proposed framework; and (v) evaluation methods for such XRI frameworks in practical application domains.

Further challenges involve data management, wherein XRI data is often sensitive and constantly collected. Similarly, adaptive interfaces are essential to reflect dynamic situations of the user and the effective relay of such contextual systems to users, in order to bring effective insight of the IoT environment's common operating picture to users. As such, the mixed reality designs to address these user interaction needs within an age of the IoT are also a challenge [14, 24, 50], in that users within such systems must engage consistently with a more fluid information eco-system. Lastly, there is also a need for frameworks that encourage and support robust information exchange and trust within the IoT [20].

Effective designs for the IoT paradigm that address these challenges have the potential to enhance and enable adaptive capabilities of the IoT, and research in this direction would ideally (a) increase the interface between users and the IoT environment, (b) keep users informed regarding IoT information, (c) increase the interaction between users within the IoT environment and allow users to more fluidly interact with each other within the system, (d) increase the trust that users have in their interactions with the IoT and with other users, and (e) ensure that user interactions are private by design and remain within the domain of ownership of information owners/stakeholders.

33.4.1 Directions for Future Work

Directions for future research in XRI extending these themes is considered in Fig. 33.10 below, including research explorations that examine the design of frontend interaction tools using 3D XR components, as well as new user interface

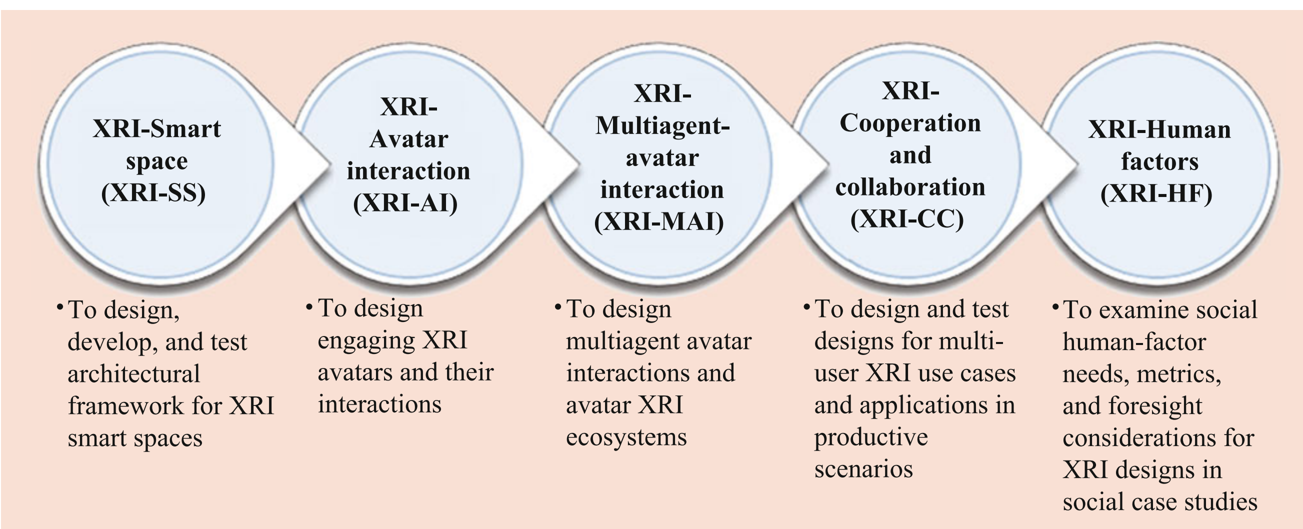


Fig. 33.10 There are new opportunities for XRI research to address smart spaces, avatar interaction, multi-agent interaction, cooperation and collaboration, and human factors, among others

metaphors that allow humans-in-the-loop to identify and understand IoT system components and their behaviors. These research directions will reveal how best to meaningfully communicate and control IoT components through commands and gestures and how best to allow IoT-embedded systems to present visualizations that are content-rich, immersive, context-sensitive, and engaging. These are broken into the following research subthemes:

- *Hybrid Mixed Reality IoT Smart Spaces (XRI-SS):* To understand how humans can best navigate and orient themselves in an immersive and visually rich smart environment, a new XR + IoT fully active smart space needs exploration.
- *Mixed Reality IoT Avatar Interaction (XRI-AI):* To qualify how well humans interact within such spaces, an expanded set of user experiences/interactions needs exploration, to merge work in XR + human-computer interaction + IoT interfaces alongside context-aware control and computer vision.
- *Multi-agent Mixed Reality IoT Avatar Interactions (XRI-MAI):* To develop coherent visualizations of IoT information and the levels of information that can be provided to humans-in-the-loop using 3D avatars, a range of interactive XR + IoT multi-agent avatars needs exploration.
- *Cooperation and Collaboration in Hybrid Mixed Reality IoT (XRI-CC):* To determine how multiple persons can apply immersive IoT environments collaboratively, XR + IoT multi-person interaction designs needs exploration.
- *Human-Factors for Hybrid Mixed Reality IoT (XRI-HF):* To identify the optimal use cases of this expanded and immersive IoT, the creation of a foresight inspired, and human factors extended, IoT model for experimentation, for studying use cases, and for determining positive and negative effects and proper and improper deployments of these hybrid IoT systems needs exploration.

In future work , the authors aim to address these directions toward the vision that the reach of the IoT will be expanded from the passive background of devices embedded in the environment into active, visible, and engaging foreground information infrastructures that have potential to drive everyday information interaction – tailored to the dynamics of users, their smart spaces, and their evolving situations. This would give rise to new forms of (i) mixed reality hybrid interfaces for IoT, to develop decentralized, adaptive XR interfaces, architectural paradigms, and frameworks for adaptive IoT systems; (ii) multi-agent interaction for IoT, to integrate agent-based context awareness and context management support for users of IoT systems and devices, allowing for streamlining the information presentation so as to be noninvasive to the user while remaining effective; (iii) context-driven IoT adaptive environments, to evaluate contextually relevant information presentation and interaction with adaptive hybrid IoT devices through deployments of adaptive and immersive interface systems in indoor, outdoor, and vehicular domain applications; and (iv) sociotechnical human factors for extending the IoT, to develop decentralized techniques to design IoT XRI systems while encouraging privacy and security, usability, and user experience, alongside socially responsible human factor designs and tools for IoT system development.

33.5 Conclusion

In this chapter, the broad and multidisciplinary domains required to bring mixed reality IoT (XRI) devices to the forefront have been examined. A hybrid taxonomy for this has been presented, alongside background literature, and existing proof-of-concept architectural frameworks and resulting XRI prototype instantiations, alongside related literature and opportunities for further research. It is hoped that such a holistic perspective of the convergence between IoT and mixed reality will encourage researchers to explore this interwoven relationship of context-driven IoT and MR for adaptive technologies and environments while maintaining focus on the human factor needs as this new landscape unfolds.

Supplementary Video Material

Examples of the applications described in this chapter can be seen in the following video(s): https://www.youtube.com/watch?v=o80bpAzf24E

Acknowledgments The authors gratefully acknowledge support from the Tri-Council of Canada under the Canada Research Chairs program.

References

1. Ahmed, E., Yaqoob, I., Gani, A., Imran, M., Guizani, M.: Internet-of-things-based smart environments: State of the art, taxonomy, and open research challenges. IEEE Wirel. Commun. **23**, 10–16 (2016). https://doi.org/10.1109/MWC.2016.7721736
2. Bernal, G., Colombo, S., Al Ai Baky, M., Casalegno, F.: Safety++ designing IoT and wearable systems for industrial safety through a user centered design approach. In: Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments, pp. 163–170 (2017)
3. Bilda, Z., Edmonds, E., Candy, L.: Designing for creative engagement. Design Stud. **29**, 525–540 (2007). https://doi.org/10.1016/j.destud.2008.07.009
4. Billinghurst, M., Kato, H., Poupyrev, I., et al.: Tangible augmented reality. ACM SIGGRAPH Asia **7**(2), 1–10 (2008)
5. Billinghurst, M., Clark, A., Lee, G., et al.: A survey of augmented reality. Found. Trends® Hum.-Comput. Interact. **8**(2-3), 73–272 (2015)

6. Blanco-Novoa, Ó., Fraga-Lamas, P., Vilar-Montesinos, M.A., Fernández-Caramés, T.M.: Towards the internet of augmented things: an open-source framework to interconnect IoT devices and augmented reality systems. In: Multidisciplinary Digital Publishing Institute Proceedings, vol. 42, p. 50 (2019)
7. Bogost, I.: Alien Phenomenology, or What It's Like to Be a Thing. University of Minnesota (2012)
8. Bridle, J.: New dark age: Technology and the end of the future. Verso Books (2018)
9. Brudy, F., Holz, C., Rädle, R., Wu, C.J., Houben, S., Klokmose, C.N., Marquardt, N.: Cross-device taxonomy: survey, opportunities and challenges of interactions spanning across multiple devices. In: Conference on Human Factors in Computing Systems - Proceedings pp. 1–28 (2019). https://doi.org/10.1145/3290605.3300792
10. Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., Ebel, P.: The future of human-ai collaboration: A taxonomy of design knowledge for hybrid intelligence systems. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, vol. 6, pp. 274–283 (2019). https://doi.org/10.24251/hicss.2019.034
11. Dey, A.K., Abowd, G.D.: The what, who, where, when, why and how of context-awareness (2000)
12. Di Geronimo, L., Braz, L., Fregnan, E., Palomba, F., Bacchelli, A.: Ui dark patterns and where to find them: a study on mobile applications and user perception. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020)
13. Fisher, J.A.: Utilizing the mixed reality cube taxonomy for interactive documentary research. AltMM 2016 - Proceedings of the 1st International Workshop on Multimedia Alternate Realities, co-located with ACM Multimedia 2016 pp. 9–14 (2016). https://doi.org/10.1145/2983298.2983299
14. Flavián, C., Ibáñez-Sánchez, S., Orús, C.: The impact of virtual, augmented and mixed reality technologies on the customer experience. J. Bus. Res. **100**, 547–560 (2019)
15. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): A vision, architectural elements, and future directions. Fut. Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
16. Gushima, K., Nakajima, T.: A design space for virtuality-introduced Internet of Things. Fut. Internet **9**(4), 60 (2017)
17. Harman, G.: Heidegger Explained, vol. 4. Open Court Publishing (2007)
18. Iatsyshyn, V.A., Kovach, O.V., Romanenko, O.Y., Deinega, I.I., Iatsyshyn, V.A., Popov, O.O., Kutsan, G.Y., Artemchuck, O.V., Burov, Y.O., Lytvynova, H.S.: Application of augmented reality technologies for preparation of specialists of new technological era. English Article **2018**, 181–200 (2020)
19. Jeschke, S., Brecher, C., Meisen, T., Özdemir, D., Eschert, T.: Industrial internet of things and cyber manufacturing systems. In: Industrial Internet of Things, pp. 3–19. Springer, Berlin (2017)
20. Jing, Q., Vasilakos, A.V., Wan, J., Lu, J., Qiu, D.: Security of the Internet of Things: perspectives and challenges. Wirel. Netw. **20**(8), 2481–2501 (2014)
21. Jo, D., Kim, G.J.: Ariot: scalable augmented reality framework for interacting with internet of things appliances everywhere. IEEE Trans. Consum. Electron. **62**(3), 334–340 (2016)
22. Jo, D., Kim, G.J.: In-situ ar manuals for IoT appliances. In: 2016 IEEE International Conference on Consumer Electronics (ICCE), pp. 409–410. IEEE, Piscataway (2016)
23. Jo, D., Kim, G.J.: Ar enabled IoT for a smart and interactive environment: a survey and future directions. Sensors **19**(19), 4330 (2019)
24. Jo, D., Kim, G.J.: IoT+ AR: pervasive and augmented environments for "digi-log" shopping experience. Hum.-Centric Comput. Inform. Sci. **9**(1), 1–17 (2019)
25. Jo, D., Kim, K.H., Kim, G.J.: Effects of avatar and background representation forms to co-presence in mixed reality (MR) teleconference systems. In: SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments, pp. 1–4 (2016)
26. Karthick, M., Kumar, C.S., Robert, T.P.: Investigation of Human Factors Using HFACS Framework—A Case Study for Unintended Reactor Trip Events in NPP. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-9008-1
27. Kayes, A.S., Kalaria, R., Sarker, I.H., Islam, M.S., Watters, P.A., Ng, A., Hammoudeh, M., Badsha, S., Kumara, I.: A survey of context-aware access control mechanisms for cloud and fog networks: Taxonomy and open research issues. Sensors (Switzerland) **20** (2020). https://doi.org/10.3390/s20092464
28. Hoon Kim, T., Ramos, C., Mohammed, S.: Smart city and IoT. Fut. Gener. Comput. Syst. **76**, 159–162 (2017). https://doi.org/10.1016/j.future.2017.03.034
29. Kim, Y.M., Rhiu, I., Yun, M.H.: A systematic review of a virtual reality system from the perspective of user experience. Int. J. Hum.-Comput. Interact. **36**, 893–910 (2020). https://doi.org/10.1080/10447318.2019.1699746
30. Kranz, M., Holleis, P., Schmidt, A.: Embedded interaction: interacting with the Internet of Things. IEEE Internet Comput. **14**(2), 46–53 (2009)
31. Kress, B., Saeedi, E., Brac-de-la Perriere, V.: The segmentation of the HMD market: optics for smart glasses, smart eyewear, AR and VR headsets. In: Photonics Applications for Aviation, Aerospace, Commercial, and Harsh Environments V, vol. 9202, p. 92020D. International Society for Optics and Photonics (2014)
32. Kress, B.C., Cummings, W.J.: 11-1: Invited paper: Towards the ultimate mixed reality experience: Hololens display architecture choices. In: SID Symposium Digest of Technical Papers, vol. 48, pp. 127–131. Wiley Online Library (2017)
33. Lee, L.H., Hui, P.: Interaction methods for smart glasses: a survey. IEEE Access **6**, 28712–28732 (2018)
34. Leotta, F., Mecella, M., Sora, D., Catarci, T.: Surveying human habit modeling and mining techniques in smart spaces. Fut. Internet **11** (2019). https://doi.org/10.3390/fi11010023
35. Lupiana, D., Driscoll, C.O., Mtenzi, F.: Defining smart space in the context of ubiquitous computing. In: Special Issue onICIT 2009 Conference -Web and Agent Systems, vol. 4, pp. 516–524 (2009)
36. Mann, S., Furness, T., Yuan, Y., Iorio, J., Wang, Z.: All reality: Values, taxonomy, and continuum, for virtual, augmented, extended/mixed (x), mediated (x,y), and multimediated reality/intelligence. Awe (2018)
37. Mignonneau, L., Sommerer, C.: Designing emotional, metaphoric, natural and intuitive interfaces for interactive art, edutainment and mobile communications. Comput. Graph. (Pergamon) **29**, 837–851 (2005). https://doi.org/10.1016/j.cag.2005.09.001
38. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inform. Syst. **77**(12), 1321–1329 (1994)
39. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: a class of displays on the reality-virtuality continuum. In: Telemanipulator and Telepresence Technologies, vol. 2351, pp. 282–292. International Society for Optics and Photonics (1995)
40. Milhorat, P., Schlögl, S., Chollet, G., Boudy, J., Esposito, A., Pelosi, G.: Building the next generation of personal digital assistants. In: 2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 458–463. IEEE, Piscataway (2014)
41. Morris, A., Lessio, N.: Deriving privacy and security considerations for core: an indoor IoT adaptive context environment. In: Proceedings of the ACM Conference on Computer and Communications Security pp. 2–11 (2018). https://doi.org/10.1145/3267357.3267363

42. Morris, A., Guan, J., Lessio, N., Shao, Y.: Toward mixed reality hybrid objects with IoT avatar agents. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 766–773. IEEE, Piscataway (2020)
43. Muhanna, M.A.: Virtual reality and the cave: Taxonomy, interaction challenges and research directions. J. King Saud Univ. Comput. Inform. Sci. **27**, 344–361 (2015). https://doi.org/10.1016/j.jksuci.2014.03.023
44. Nielsen, J.: Iterative user-interface design. Computer **26**(11), 32–41 (1993)
45. Norman, D.: The Design of Everyday Things, doubleday/edn. Basic Books (1990)
46. Nuamah, J., Seong, Y.: Human machine interface in the internet of things (IoT). In: 2017 12th System of Systems Engineering Conference (SoSE), pp. 1–6. IEEE, Piscataway (2017)
47. Paradiso, J.A., Landay, J.A.: Guest editors' introduction: cross-reality environments. IEEE Pervasive Comput. **8**(3), 14–15 (2009)
48. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Context aware computing for the internet of things: a survey. IEEE Commun. Surv. Tutorials **16**, 414–454 (2014). https://doi.org/10.1109/SURV.2013.042313.00197
49. Phupattanasilp, P., Tong, S.R.: Augmented reality in the integrative internet of things (AR-IoT): Application for precision farming. Sustainability **11**(9), 2658 (2019)
50. Rokhsaritalemi, S., Sadeghi-Niaraki, A., Choi, S.M.: A review on mixed reality: current trends, challenges and prospects. Appl. Sci. (Switzerland) **10** (2020). https://doi.org/10.3390/app10020636
51. Shao, Y., Lessio, N., Morris, A.: IoT avatars: mixed reality hybrid objects for core ambient intelligent environments. Proc. Comput. Sci. **155**, 433–440 (2019). https://doi.org/10.1016/j.procs.2019.08.060
52. Shin, D.: A socio-technical framework for internet-of-things design: a human-centered design for the Internet of Things. Telematics Inform. **31**(4), 519–531 (2014)
53. Shinde, G.R., Dhotre, P.S., Mahalle, P.N., Dey, N.: Internet of Things Integrated Augmented Reality. Springer, Berlin (2020)
54. Shirehjini, A.A.N., Semsar, A.: Human interaction with IoT-based smart environments. Multimedia Tools Appl. **76**(11), 13343–13365 (2017)
55. Thierer, A., Castillo, A.: Projecting the growth and economic impact of the Internet of Things. George Mason University, Mercatus Center, vol. 15 (2015)
56. Truelove, A., Chowdhury, F.N., Gnawali, O., Alipour, M.A.: Topics of concern: identifying user issues in reviews of IoT apps and devices. In: 2019 IEEE/ACM 1st International Workshop on Software Engineering Research & Practices for the Internet of Things (SERP4IoT), pp. 33–40. IEEE, Piscataway (2019)
57. Tuladhar, S., AlSallami, N., Alsadoon, A., Prasad, P.W., Alsadoon, O.H., Haddad, S., Alrubaie, A.: A recent review and a taxonomy for hard and soft tissue visualization-based mixed reality. Int. J. Med. Robot. **16**(5), 1–22 (2020). https://doi.org/10.1002/rcs.2120
58. Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Duan, X., Meekhof, C., Stühmer, J., Cashman, T.J., Tekin, B., Schönberger, J.L., et al.: Hololens 2 research mode as a tool for computer vision research (2020). arXiv preprint arXiv:2008.11239
59. Vicente, K.: The human factor: revolutionizing the way we live with technology. Vintage Canada (2004)
60. West, D.M.: How 5g technology enables the health internet of things. Brookings Center for Technology Innovation, vol. 3, pp. 1–20 (2016)
61. White, G., Cabrera, C., Palade, A., Clarke, S.: Augmented reality in IoT. In: International Conference on Service-Oriented Computing, pp. 149–160. Springer, Berlin (2018)
62. Zemmoudj, S., Bermad, N., Omar, M.: Context-aware pseudonymization and authorization model for IoT-based smart hospitals. J. Ambient Intell. Hum. Comput. **10**, 4473–4490 (2019). https://doi.org/10.1007/s12652-018-1129-0
63. Zollmann, S., Grasset, R., Langlotz, T., Lo, W.H., Mori, S., Regenbrecht, H.: Visualization techniques in augmented reality: a taxonomy, methods and patterns. IEEE Trans. Vis. Comput. Graph. **2626**, 1–20 (2020). https://doi.org/10.1109/TVCG.2020.2986247

Georgina Yeboah specializes in Digital Illustration, Interaction Design, and Human-Computer Interaction (HCI) research. She holds a BFA in New Media from Ryerson University and an MDes in Digital Futures from OCAD University. Her main research focuses on the nature of interactivity within digital and physical spaces. She is interested in smart space design, context-aware applications, and the Internet of Things.

Alexis Morris is the Canada Research Chair (Tier II) in the Internet of Things, Director of the Adaptive Context Environments Lab, and an Assistant Professor in the Digital Futures program at OCAD University. He is also an Interdisciplinary Computer Scientist with a PhD from the University of New Brunswick, specializing in mixed reality adaptive interfaces and context awareness in smart environments.

33

34 Digital Twin and Extended Reality: Strategic Approach and Practical Implementation

Roberto Rocca, Roman Felipe Bastidas Santacruz, Claudio Sassanelli, Paolo Rosa, Luca Fumagalli, and Elisa Negri

Contents

34.1 Introduction ... 853

34.2 DTs Enabling Technologies ... 854
34.2.1 Enabling Technologies of DT ... 854
34.2.2 Digital Twins ... 856

34.3 Research Methodology ... 857
34.3.1 Literature Review ... 857

34.4 Results ... 860
34.4.1 Findings from the Literature Review ... 860
34.4.2 DT Strategic Tool ... 864

34.5 Practical Implementation of a Combined DT-ER Solution ... 868

34.6 Discussion ... 874

34.7 Conclusion ... 874

References ... 875

Abstract

Nowadays, the digital transformation (driven by the Industry 4.0 (I4.0) paradigm) is becoming one of the most promising and valuable strategies to address business needs of manufacturing players in terms of agility, efficiency, and real-time reactivity. Among available digital (I4.0-based) technologies, Digital Twins (DTs) are described by researchers and practitioners as a key element in terms of smart manufacturing because of their potential to enable the shift from automation to autonomy. To this aim, the present chapter highlights characteristics and benefits of DTs, by proposing a strategic tool to support and guide Small- and Medium-sized Enterprises (SMEs) toward their adoption and exploitation. In addition, potentialities coming from the integration of both DTs and Extended Reality (ER) tools are shown through a laboratory application case.

Keywords

Digital twin · Virtual reality · Industry 4.0 · Strategic tool · Laboratory application case

R. Rocca (✉) · R. F. B. Santacruz · C. Sassanelli · P. Rosa · L. Fumagalli · E. Negri
Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy
e-mail: roberto.rocca@polimi.it; romanfelipe.bastidas@polimi.it; claudio.sassanelli@polimi.it; paolo1.rosa@polimi.it; luca1.fumagalli@polimi.it; elisa.negri@polimi.it

34.1 Introduction

Within modern industrial contexts, digital transformation represents a precondition toward both production flexibility and fast reaction to current changing trends, such as volatile market demand, increasing frequency of new products introduction and new technological developments [1, 2]. Digital transformation, globally recognized as the transition enabled by Industry 4.0 (I4.0) technologies, allows new automation architectures for production processes: (i) enhancing flexibility and scalability, (ii) enabling the integration of modern Information and Communication Technologies (ICTs), and (iii) increasing efficiency and production performances [3, 4]. Together, technologies like Machine-to-Machine (M2M) communications, Cyber-Physical Systems (CPSs), Internet of Things (IoT), and Cloud Computing (CC), allow companies to manage manufacturing processes in a more intelligent and agile way [5, 6]. These smart technologies are changing classic industrial processes (e.g., storage, transportation, and transformation of raw materials into useful products) architectures [7], basing on the well-known International Standards of Automation (ISA-95) [8]. This change affects both horizontally and vertically industrial supply chains. From one side, data availability and data sharing are transforming the way companies interact within the same supply chain. From another side, ICT structures of manufacturing companies are

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_34

becoming an integrated (complex) system where dataflows are relevant for every level of the company. These transformations that the manufacturing environment is experiencing depend on new digital technologies and strategies, which allow companies to take advantage from potential benefits and new/enhanced business models brought by I4.0 technologies.

This emerging revolution does not refer to a single disruptive invention, but a set of different technologies that can be combined in several ways. Some of these technologies already existed in the past, others are still evolving at a rapid pace. The technological domain is extremely vast and heterogeneous, with more than 30 different fields of technology fields [9]. To better understand I4.0 drivers, Gulot et al. [10], identified four clusters describing the nature of different technologies:

- Physical-digital interface technologies: They connect the cyber and physical worlds (machines, product and people). These technologies refer to CPSs, IoT and Extended Reality (ER).
- Network technologies. They improve connectivity through online functionalities: These technologies include CC, blockchain technology, and cybersecurity solutions [11].
- Data-processing technologies: They support decision-making, easing its transition from reactive to proactive processes and even real-time. Simulation and modelling, Machine Learning (ML), Artificial Intelligence (AI) and Big Data Analytics (BDA) are the most valuable examples.
- Physical-digital process technologies: They comprise technologies like 3D printing and advanced robotics.

In this heterogeneous technological landscape, Digital Twins (DTs) have gained significant impetus, both in academia and industry, as one of the most promising technologies to achieve factory digitalization [12]. DTs close the loop between physical and virtual worlds, through a real-time data interaction between virtual models and physical resources, by generating information that can improve manufacturing [13]. In this chapter, a review of the current state of the art on DTs is presented, by showing their characteristics and benefits and proposing a strategic tool for guiding their implementation in manufacturing companies. Moreover, potentialities coming from the integration of DT with ER are presented and discussed through a laboratory application case, where a Waste from Electrical and Electronic Equipment (WEEE) disassembly plant configuration has been virtually and practically tested through a set of dedicated simulation tools and a fully automatized manufacturing line. The chapter is organized as follows. Section 34.2 provides a brief description of the main digital technologies to be exploited for a DT solution development and a brief overview on DTs. Section 34.3 reports the systematic literature review conducted, and Sect. 34.4 shows the main findings deriving from it. Section 34.5 highlights the benefits coming from a practical implementation of a combined DT-ER solution. Finally, Sects. 34.6 and 34.7 present some discussions and concluding remarks.

34.2 DTs Enabling Technologies

34.2.1 Enabling Technologies of DT

According to the National Institute of Standards and Technology (NIST), Smart Manufacturing (SM) is defined as "fully-integrated, collaborative manufacturing systems responding in real time to meet changing demands and conditions in the factory, in the supply network and in customer needs" [14]. Therefore, SM is constituted by a network of technologies allowing decentralized machines to collaborate and coordinate their tasks to increment the productivity and quality, by reducing the overall production cost. However, the evolution of manufacturing into SM through I4.0 technologies can maintain the old automation pyramid intact [15], by just acting on the different way layers can interact between them, gather information from the field, and store data [12, 16, 17] (see Fig. 34.1 below).

Hereafter, a brief description of the main I4.0 technologies on which a DT (and the lab-scaled application case discussed in Sect. 34.5) is built on is provided.

Cyber-Physical Systems

CPSs link all I4.0 technologies together to make them useful for the enhancement of industrial processes. CPSs integrate computation, networking, and physical processes [18–20]. They represent a new generation of digital systems, consisting of two main functional components: (i) an advanced connectivity to ensure real-time data acquisition from the physical world and information feedbacks from the cyber space and (ii) an intelligent data management, analytics and computational capability that constructs the cyber space [18]. On one hand, CPSs enhance communication between field entities (sensors and actuators) and cyber computational resources to real-time monitor and control physical entities in a reliable, safe, collaborative, robust, and efficient way. This way, resources (products, material, and energy) can be allocated efficiently, basing on intelligent cross-linked value creation modules [21]. On the other hand, DTs create high-fidelity digital counterparts of physical objects in a virtual space, by integrating historical and real-time data of physical systems with physics-based models and advanced analytics,

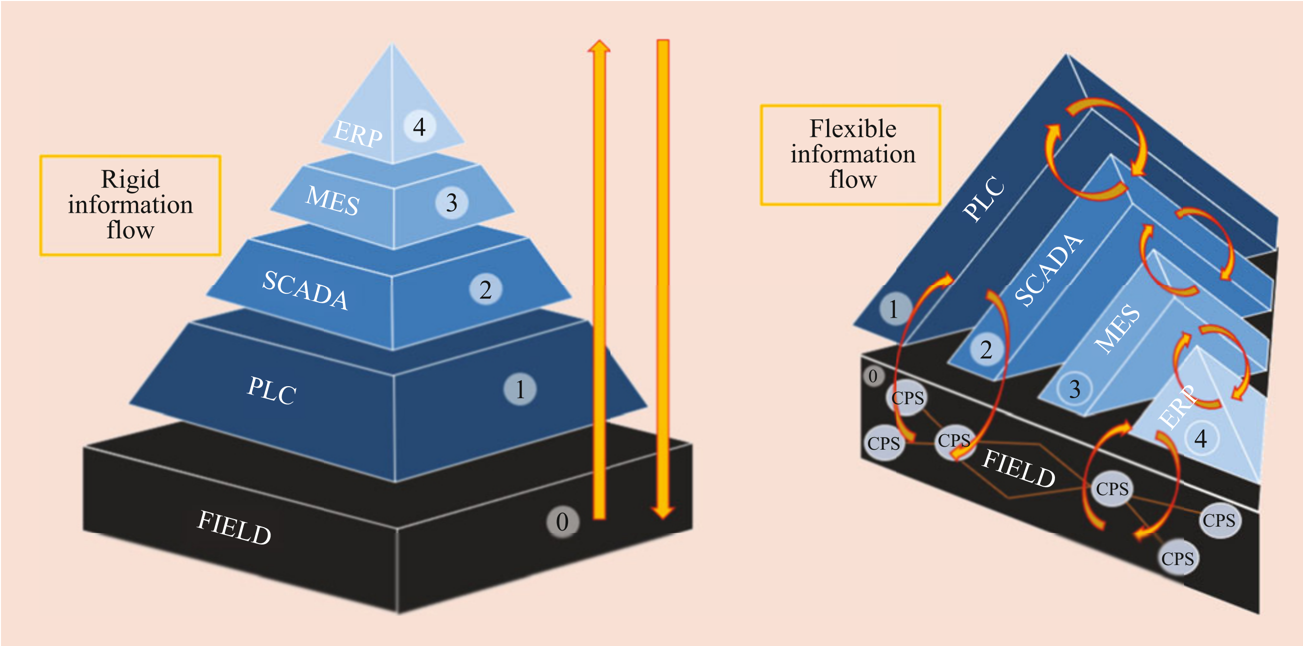


Fig. 34.1 ISA-95 automation pyramid transition into flexible automation pyramid. Adapted from [7]

by guiding the decision-making process through simulated and predicted behaviours of manufacturing entities [22].

Internet of Things

IoT refers to the interconnectivity between things (e.g., embedded sensors, electronic devices, machines, transportation modes and the internet), through unique identification codes [23]. IoT allows "things," that refer to any element or device that can be connected to internet and transmit or receive data, to communicate with one another. IoT role lies in collecting data from the field and transfer them (e.g., through CPSs [24]) to other I4.0 technologies (e.g., BDA and CC) able to analyze them and extrapolate information for a better decision-making either at product (e.g., [25–27]) or process level (e.g., [28]). IoT intrinsically increases resource efficiency [27, 29–31], and provides process-related data, optimizing manufacturing practices and enabling better production planning and control [28, 29]. Moreover, the information exchange between things generates a large quantity of data, which, if analyzed, can be valuable for organizations. Technologies such as BDA and CC support systems that require to read, transmit, store, and analyse big sets of data from the manufacturing field. IoT allows to gather sensors and actuators values directly from the field, and BDA helps to gain in-depth understanding of the hidden values of gathered data. CC provides solutions for the storage and processing of BDA; it uses huge computing and storage resources under concentrated management, so as to provide BDA applications with fine-grained computing capacity [32].

Simulation

The term simulation refers to a collection of technics to mimic a specific behaviour of a real or ideal system, using resources (time and knowledge) to simulate experiments and then support the decision-making process [33]. Indeed, simulating means recreating the function of a system over an extended timeframe, in a shorter (simulated) time, to predict its dynamic behavior given a particular scenario. Computers became powerful enough to mimic and predict the physical world, allowing them to simulate real world situations in a digital environment [34]. Computer simulation encourages experimentation and creative thinking since it is relatively quick and cheap to deploy, and the cost of failure is minimal. As it is a powerful tool for analysing complex stochastic systems, computer simulation has been adopted in various sectors, such as manufacturing, services, defence, healthcare, and public services [35]. Particularly, simulation plays a significant role in manufacturing, as it supports the design of new solutions, the improvement of actual systems/solutions, the performance analysis, what-if analysis (how output change according to variation of some process parameters), and training [36]. Within the I4.0 paradigm, simulation is used to replicate real world behaviors in virtual environments, where physical and virtual dimensions coexist and are synchronized in real-time [18]. However, synchronization requires full data models of alternative scenarios to be simulated [37–39]. This issue led to the concept of DTs, or a virtual representation of physical objects copying their behavior through a real-time data acquisition from the field

[12, 40]. DTs not only allow a prognostic assessment at design stage (static perspective), but also a real-time synchronization and optimization of the virtual object (dynamic perspective) [41].

Extended Reality

Extended Reality (ER) is a term that embraces all technologies that allow users to interact with real and virtual worlds [42]. Different technological variations can be found in this group such as: (i) Virtual Reality (VR), (ii) Augmented Reality (AR), and (iii) Mixed Reality (MR). The first, VR is a computer-generated simulation where users can interact in an artificial and virtual environment using VR peripherals (gloves, googles, helmets, displays, etc.) [43, 44]. The second, AR enhances the real world, by adding graphical digital elements to a live view of the real world [42, 45], and then, MR combines characteristics from both AR and VR. Here, real and virtual environments interact with the user. ER technologies are becoming mature in line with the development of related I4.0 technologies (e.g., IoT, computers graphics, BDAs, CPSs and DTs). Within the SM context, DTs and ER are considered two of the most relevant technologies to enhance performance of physical entities [46] in any phase of product lifecycle, from design to disposal [43]. The integration of DTs and any form of ER technologies potentiates the benefits that each of them can offer individually and opens a new set of applications. To this aim, its incorporation has been considered in many manufacturing processes since it can create an interactive environment that allows a seamless integration between the digital and physical worlds [47]. Nowadays, the main applications involving these two technologies are focused on visualization, physical-virtual integration, user experience enhancement, and remote collaboration [47]. ER/DTs can also: (i) support training activities and enable an interaction with DTs of a factory floor or assets [48], (ii) allow a virtual exploration of a factory and/or a machine by new workers, (iii) support the learning process of new machinery by experienced workers [49], and (iv) assess the operators' safety conditions and study the layout and ergonomics of the workstations.

Even when the integration of DTs and ER promises immense benefits for enterprises, these technologies have important barriers that limit its adoption:

(i) Computing power: Computers advance each year in terms of processing capacity, nevertheless, the capacity needed for some DTs is too high in order to get accurate results [50]. This ends up translating in the limitation of models for the analysis of physical processes.
(ii) Communication: DTs and ER are based on the capacity to transmit, receive, and process information. Besides computing power, it is also vital to count on a solid communication infrastructure. High data transmission speed must be achieved to count on reliable real time applications [51].
(iii) Lack of general frameworks that allows companies to follow stablished procedures to implement these two technologies in their processes.

Augmented Reality

From the ER technological group, AR has gained attention in the manufacturing environment as it is one of the most relevant technologies to be implemented in conjunction with DTs. AR allows a real-time direct or indirect view of a physical environment [52]. The main technology's objective is to simplify the user's interaction with the real world; this is done by bringing information to the user's surroundings. By superposing virtual objects and visual indicators on the real-world environment, AR increases data visibility and improves its analysis process [52].

AR is not limited only to visual enhancement, but the possibility to support processes with the improvement of touch and hearing senses. In this context, AR is one of the most suitable technologies to support DT applications, mainly by showing information to the user and allowing a seamless interaction with the industrial environment. One of the possible applications of AR and DTs is the 3D visual assembly of complex products (e.g., engines, turbines, aircrafts, etc.), given its critical relevance over the products' end quality and performance. A 3D visual assembly is the integration of digital models of an asset in the physical assembly environment. This way, operators can interact in real time with the assembly process of the product [43]. From one side, the integration of DTs and AR can provide technical support in the process, for example by showing the geometrical characteristics products (e.g., color, size, tolerances, etc.) and assembly information (e.g., part numbers and components) [53]. From another side, it can bring direct benefits as higher quality, increased assembly efficiency and reduced costs. [54] investigated an assembly processes adopting AR and DTs. Here, these two I4.0 technologies allowed the optimization and improvement of the planning, guidance, and training processes. Another example of AR/DTs application is represented by Asset Lifecycle Management (ALM), where DTs and AR thanks to their capacity to offer high quality predictions and data analysis [51, 55] allowed to achieve (i) higher savings due to increased efficiency; (ii) optimization and increase of machinery lifecycles; and (iii) reduced machinery downtimes [55].

34.2.2 Digital Twins

DTs are described as a key drivers in the vision of SM [13, 51, 56]. DT is meant as the virtual and computerized counterpart of a physical system [47] enabled through data and simulators for real-time prediction, optimization, monitoring, control, and improved decision-making. To realize a real-time synchronization with the physical system, DTs must be supported by a proper data model containing all the necessary information of the system operations (e.g., history, behaviour and current state). A virtual-physical interaction allows the real-time monitoring and adjustment of manufacturing processes during the current manufacturing execution stage. This way, the virtual workshop updates itself in real-time with the sensed data coming from the field and (through simulation) the virtual factory evaluates various manufacturing strategies to achieve optimal manufacturing [12, 56]. DTs are, therefore, able to continuously collect and process information, creating a comprehensive picture of a given product/production process, thus creating a digital factory environment. In this context, decision makers (or autonomous systems) can take the right decisions to optimize actual/future production resulting in higher efficiency, accuracy in the production, and economic benefits for the company [12, 57, 58].

According to Kritzinger et al. [59], it is possible to classify DTs in three subcategories: (i) Digital Models (DMs), (ii) Digital Shadows (DSs), and (iii) DTs. Despite the latter are often used synonymously, they differ in the level of data integration between physical and digital counterparts. A DM is a digital representation of an existing or planned physical object, where data exchange between the physical and digital side is done manually. Since DMs do not use any form of automated data exchange, a change in state of the physical object has no direct effect on the digital object and vice versa [13]. DSs, instead, is a DM with an automated one-way data flow between the state of an existing physical object and a digital object. This way, a change in state of the physical object leads to a change of state in the digital object, but not vice versa. Finally, DTs refers to a combination where data flows between an existing physical object and a digital object are fully integrated in both directions. The action of physical object may induce a change in state in the digital object, and the latter can act as controlling instance of the physical object. In this sense, a change in state of the physical object directly leads to a change in state of the digital object and vice versa.

34.3 Research Methodology

This section presents the systematic literature review conducted to address the state of the art of DTs and answer to the following research questions: (i) how widespread is DT in manufacturing contexts? (ii) which are the different types of DTs and their application in manufacturing? (iii) which are the current benefits of DTs? (iv) how are DTs supported by AR technologies? The review allowed to assess the main characteristics of DTs and led to the creation of a novel qualitative strategic tool to support and help Small and Medium-sized Enterprises (SMEs) to implement DT solutions within the factory. In addition, the objective (iv) is performed with the intention to explore the convergence between DTs and AR, and clarify how these two technologies together can support different processes within factories taking into account the actual state of the art of their applications. Moreover, the benefits coming from a digital solution that combines DT and ER through a laboratory application scale are highlighted and linked to the literature review findings in Sect. 34.5.

34.3.1 Literature Review

Inspired by the guidelines proposed by Kitchenham [60], the literature review has been structured following three main phases: (i) planning the review; (ii) conducting the review; (iii) reporting the review.

Planning the Review

According to the guidelines proposed by [60], the first and critical step in a systematic review is the definition of one or a set of research questions. The PICOC model [61] (Population, Intervention, Comparison, Outcome, Context) has been firstly defined, as explained below:

- Population: the practitioners affected by the intervention. In this review, the research is centred on the manufacturing field, without any specific sector. For this reason, the main population is the manufacturing industry.
- Intervention: the software methodology/tool/technology/procedure that addresses a specific issue. In this work, the technology involved is the DT and from the ER technologies, AR is further explored.
- Comparison: is the software engineering methodology/tool/technology/procedure with which the intervention is being compared. DT and AR are considered new smart technologies, still not very used in the manufacturing field. For this reason, the intervention is compared to manufacturing supported by traditional technologies.
- Outcomes: factors of importance to practitioners. All relevant outcomes should be specified. All relevant outcomes like increased productivity, flexibility, reduced production cost, and reduced time to market will be taken into consideration.
- Context: the context in which the comparison takes place. The context of analysis in this chapter is centred on manufacturing industry and academia research. Academia experiments may not be representative of what might occur with practitioners working in the industry but represent a valuable source of information.

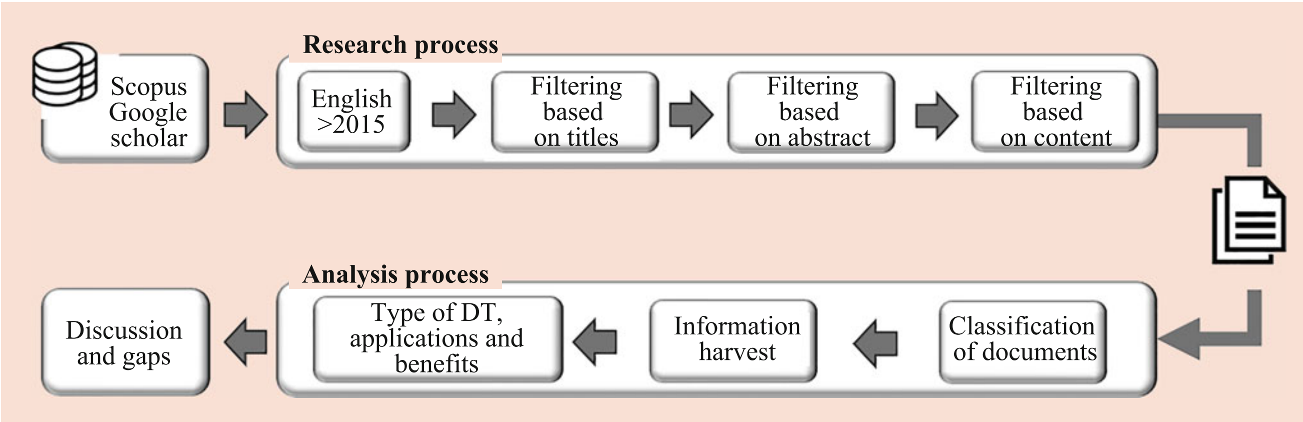


Fig. 34.2 Research and analysis process for literature review

Then, the experimental design has been defined. Since there are many studies on DTs available in literature, the experimental design in this work consists in gathering the most relevant studies and industrial applications and try to collect the relevant information for manufacturing industry. The adopted experimental design consists in a combination and aggregation of primary studies, like application case in laboratories or industry, with secondary studies to get broad information on the topic.

Conducting the Review

Figure 34.2 reported below represents the process followed in conducting the systematic literature review. In particular, the process consists in two sub-processes (phases): (i) research process and (ii) analysis process. The aim of the research process is to find all the scientific studies relating to the agreed research questions using an unbiased searching strategy. In this step, a funnel-like method has been followed. The purpose of this narrowing process is to find a limited set of papers or citations that might be useful under the review process perspective. As a starting point, the search engines Scopus and Google Scholar were used to gather publications on DT in Manufacturing. "Digital Twin" and "Manufacturing" were used as separate keywords for searching articles in Scopus database. Results counted 752 papers. The same searching strategy was followed on Google Scholar, resulting in more than 10.000 articles, which is an enormous amount to effectively filter for the relevant ones. To lower initial number of papers, "Digital Twin Manufacturing" was chosen as combined keyword to conduct the searching process in Google Scholar, and the results were reduced to 567 articles. From this broad search a total number of almost 1319 potential articles were identified. It can be noticed that many articles were repeated across the engines, thus the number of non-repeated publications remained almost 800.

After the identification of the potentially relevant primary sources, their actual relevance has been assessed by defined criteria that will be described in this section. This way, the inclusion or exclusion of the articles in the investigation of the work was decided. Papers were initially filtered by considering the English language and a publication date after 2015. In fact, one of the first review extracted from the search [12] highlighted that the publications linked to concept of DT in production systems in the I4.0 started to increase in 2015. The number of papers decreased to 480. In order to perform a further filtering, the keywords and titles were assessed. The documents not related with manufacturing applications or centred in other industry outside manufacturing scope were discarded, thus remaining in 362 articles. The ones chosen to be relevant then were subjected to a third filter based on its abstract, to assess if they could be useful or not to the research questions. In this filtering, it was considered if the concept of DT was a central topic of the paper and if the link with manufacturing was clear. Only 168 articles remained. After that, the filtering process continued with full-text assessment of the papers, evaluating again the consistency in terms of field of application with respect to the manufacturing and industrial environment. After this last step, some other articles were discarded, because a deeper analysis found that the content of the paper was not related with the study. Finally, the papers selected to be relevant for the research study were 72 (see Fig. 34.3).

Then, the analysis process has been started with a classification of the documents selected in the previous phase. In the next tables, a first classification is reported. Each table reports documents centred on specific types of DTs, i.e., the scope and level of virtualization and synchronization (such as asset, process, factory, product, people, or general). Each table also involves several variables, as summarized below:

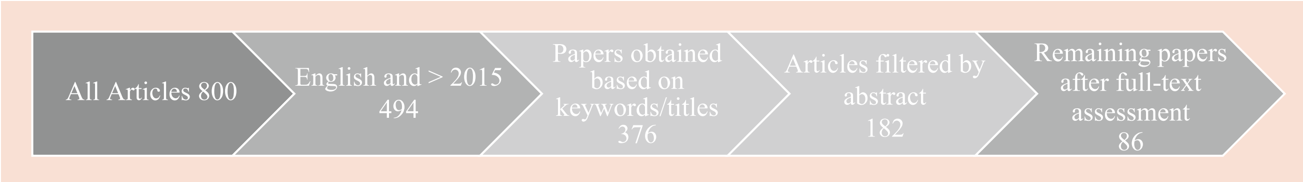


Fig. 34.3 Literature review flow diagram

- Typology of publications: Articles were analysed concerning their type, i.e., if they were reviews, case studies or concept-papers. In case a publication includes more than one category, then it was classified into the one that is most discussed in the document.
- Focused area: DTs encourage different applications in manufacturing, such as asset management, product lifecycle, and production planning and control (PPC), maintenance, factory design, process improvement and optimization, etc. In case that the document analysed is not focused in a particular area, but it discusses DT in a broader sense, then the document is categorised as "manufacturing in general".
- Lifecycle perspective: For each article analysed, was noted if the lifecycle point of view is considered or not (in the next tables, an "x" corresponds to the presence of the lifecycle perspective).
- Benefits: The contribution and perceived benefits that DT brings to manufacturing are reported.

The above tables are classified by the types of DTs existent in literature. In each table, the specific DTs types are evaluated. It is identified which is the kind of application they were used for, and the respective benefits that it can bring.

Then, with the goal to analyze the state of the art of the current applications of DT that involve AR, the documents found to discuss these two technologies and its relation, were clustered with the intention to study in detail this relationship. The documents were classified in depending on the following characteristics as it can be seen in Table 34.5:

- The focused area of application of the technology in the manufacturing context
- Type of ER technology implemented or studied in the document
- The hardware, software and ER peripherals implemented in the interaction between DT and AR.
- Characteristic of AR utilized to enhance the DT application
- The benefits that the technology adopter company would get from the implementation of a DT-AR application

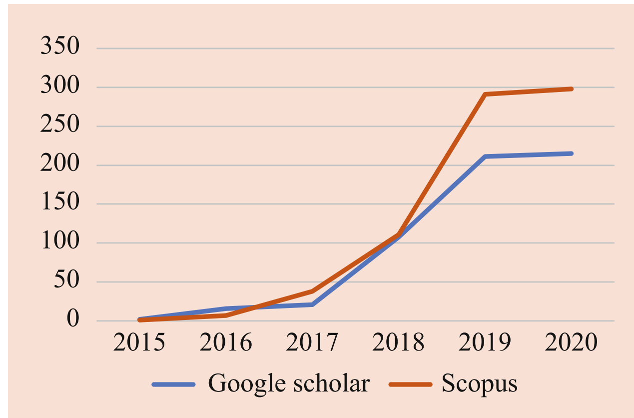


Fig. 34.4 Number of publications on DT in manufacturing

The table above shows the analysis performed in the documents of the literature review that were strictly focused on DT-AR applications. The next section explains in detail the results obtained from the analysis process

Reporting the Review

The first finding coming from the analysis is that in the last few years, DT has gained a lot of attention in manufacturing [13, 37, 131] (see Fig. 34.4). DT has been adopted in several areas of different industries and the increase number of publications highlight the interest in this technology. The reasons of the rapid growth of applications arose from a combination of technology push and market pull. On the one hand, the recent high-pace development in technologies such as sensors, ICT, AI and CC creates the perfect environment to develop DT application. On the other hand, some important changes driven by customer demand and market uncertainty like the increasingly urgent flexibility in operation, the need of agility and the personalization of products are some of the most evident market pull [58]. In addition, the DT is considered one important concept to achieve SM. The digitalization of manufacturing systems has the potential to transform the landscape of manufacturing, enabling the shift from analyzing the historical data to predicting the future [81]. These are the reasons why this technology is becoming part of digital strategy in many manufacturing companies.

34.4 Results

34.4.1 Findings from the Literature Review

As previously mentioned, DTs have a pivotal role in mirroring the vision of SM. From the classification "focused areas" reported in each one of the previous tables, it emerges a number of specific aspects in which DT can influence future manufacturing: (i) asset management, (ii) process improvement, (iii) factories, (iv) people, and (v) product optimization. Apart from "manufacturing in general" category, that includes publications without a focus on a particular area within manufacturing, the other areas suggest that there are different types of DTs. Those areas will not singularly be described, while the main DT types within its main applications are described below.

DTs for Asset Management in Manufacturing

In the last years, the concept of Asset Management is gaining more attention [62]. It can be defined as "the coordinated activities of an organization to realize value from asset". A central principle of the asset management is the Asset Lifecycle for which any asset is characterized by three phases: Beginning of Life (BOL), Middle of Life (MOL), and End of Life (EOL) [132]. The Lifecycle of a physical asset begins when its acquisition is first considered and ends when it is finally taken out of service for disposal or redeployment. Moreover, ALM is founded in the following principles [62]:

- Lifecycle orientation: A decision should be taken considering the whole asset lifecycle. Therefore, the decision-making is driven by long-term objectives and perform measurements.
- System orientation: Asset systems should be considered entirety. Every time a local decision is taken on an individual component the global impact at system level can also be considered.
- Risk orientation: Relevance of a risk culture to manage the failure of critical assets. ALM empowers the utilization of analytical tools, which offer anticipated predictions on possible threats in operations.
- Asset-centric orientation: ALM requires a clear awareness and knowledge of the key assets to make sound business decisions.

A manufacturing asset can be connected and abstracted to the cyberspace via its DT [57]. This way, manufacturing companies are able to exploit a model capable to recreate the state and behavior of the asset in real time. Thanks to this digital counterpart of the asset, it is possible to capture data from the physical asset, evaluate its real-performances, and operating conditions with the aim of making proactive optimal operation decisions [133]. Indeed, having a clearer picture of real-world status of the manufacturing asset, the manufacturers can improve their situational awareness and enhance operation resilience and flexibility, especially in the context of mass personalization [46]. In the perspective of Asset Management, the DT can support the asset-related decision-making process under several perspectives:

- Asset configuration: DTs are used for modelling and simulating a production line, aiming at assessing the best design solution. The selection of the best solution can be done through the evaluation of systemic Reliability, Availability, and Maintainability performance (RAM performance) that allows the computation of Total Cost of Ownership (sum of all significant costs associated with an asset over the defined lifecycle) [62, 134].
- Asset reconfiguration: DTs allow to quickly evaluate the best reconfiguration alternative of a complex production plant. Homogenously to asset configuration, the assessment of RAM performance is necessary for the prediction of Total Cost of Ownership of the asset.
- Asset commissioning: DTs are used to make the virtual commissioning of the manufacturing system. DT, conceived as semantic data model, data analytics, and advanced simulation, allows a quick commission.
- Asset condition monitoring and health assessment: DTs are also used for the asset diagnosis, helping to assess its health status based on the monitored condition. DTs provide the data analytics in order to estimate asset reliability, monitor anomalies, and thus limit unreliability situations.

Above, the variety of decisions supported by DTs has been described in different asset lifecycle phases and at different asset control levels. In a lifecycle perspective, DTs can be used as a digital reflection of an asset life cycle, to maintain an uninterrupted data flow on various stages of an asset lifecycle.

DTs for Process Improvement

The main scope of a DT for production process improvement is to evaluate, optimize and predict the production process based on real-time simulation data, real-time production data, and historical production data. According to the analysis of this data and the machine states simulation, the real-time controlling instructions are fed back to the physical space to optimize the production process [104]. On one hand, DTs can be used to identify process faults and reconfigure control parameters. Based on fault information in the process, controllers can make decisions and optimizations to guarantee the safe operation of process systems. On the other hand, data collected from normal processes can be used to perform process monitoring and diagnosis through DTs [51, 100]. It is worth mentioning that DTs are primarily developed for timely monitoring, diagnosis, and tolerance control in process factories. This way, it is possible to extend the operation up-time of any faulty plant with an acceptable performance until the next maintenance time is reached. In other words, DT systems shorten the downtime of factories and extend the operation time for process systems [58].

According to the literature review has been conducted, the main efforts have mainly focused on process improvement or process optimization. In addition, DTs are proposed as a process knowledge reuse method. The process knowledge reuse is an important factor to consider in order to manufacture high value-added products with the best possible quality and short lead time at a competitive cost [104].

DTs for Factories

DTs can also work for factories, making a replica of a live factory environment. The DT-based method builds a continuous interactive process between the physical manufacturing factory and a virtual digital factory [37, 104]. The virtual digital factory will continuously collect real-time data from the physical production line. Then, real-time and historic data are utilized for model training, model verifying, and model updating, and ultimately providing feedback to the real factory for production control purposes [135]. The data tracking offers capabilities of tracing product fault sources, analyzing production efficient bottlenecks and predicting future resource requirements. DTs and data-driven production operations can allow the establishment of a self-organizing factory environment with complete operational visibility and flexibility. This way, manufacturing factories can achieve accurate and agile production control in response to the changes in market demand [81]. From the analysis of the publications, it has been emerged that DTs can be beneficial mostly for factory design. Factory design includes detail works such as factory layout, capacity calculation, machine utilization, number of machines, designing of the logistics, and calculation of its efficiency. Therefore, the dynamic behavior of the manufacturing system is difficult to be presented and predicted by traditional algorithms. DTs can effectively solve this problem because of the fidelity to physical factory. Indeed, DTs can increase the feasibility of the design project by supporting designer to escape design flaws, and quickly evaluate different design alternatives. The application approach of DTs in factory design is different with the other applications, since the twin is mirroring the designed factory that will be physical in the future. So far, it has been presented the DT concept to make designed factory healthy and strong [73]. The future vision is that DTs will be applied through the whole life circle of smart factory, helping the decision-making in both factory design and operation.

DTs for People

DTs can be used also to understand the human state at work and guide designing human-centred and human-machine collaboration strategies in production [118]. In fact, through DTs, workers can be connected at shop floor and it is possible to establish models to understand personal wellbeing and working conditions of human in a factory. Improving the physical and psychological health of workers is crucial to achieve best production performances [81]. However, only few papers of the literature review focused on the human factor. DTs can be used also to set up personalized virtual training programs of workers and factories. Through ultra-realistic training programs workers can upskill themselves, and this can lead to tremendous resource optimization and operational efficiency improvement [42]. The main evidence that emerges from the literature is that the implementation of the DT approach enables the optimization of the planning and commissioning of human-based production processes using simulation-based approaches.

DTs for Product

DTs support the Product Lifecycle Management (PLM) as an information vehicle along the whole lifecycle of the product [51, 136]. This means that the information created in each stage of the product lifecycle are made available to the subsequent phase. In the product design stage, usually designers must carry out various tests to constantly prove the validity and usability of the design.

Traditional product design processes imply the utilization of professional knowledge and experience at the centre. Contrary to it, the modern product design processes search to enhance the participation of customers. Based on DTs, the product design process can be divided into:

- Conceptual design: Here, designers need to determine the future designing direction of the entire product defining the concept, aesthetics, and the main functions of the new product. Usually, in this phase, the designer encounters difficulties in dealing with various kinds of data needed, such as customer satisfaction, product sales, product competitiveness, investment plans, and many other information. DTs can be a solution for this problem, since they are able to integrate all these kinds of data and information in the product's physical space [137]. Having a single source of information can help designers to make a quick understand on where should be improved. Moreover, being DTs a faithful representation of the physical product, the communication between clients and designers can become more transparent and faster by using the real-time transmission data.
- Detailed design: this stage mainly involves the design and construction of the product prototype. Furthermore, designers need to refine the product design scheme which

includes product functions and appearance, product configuration, design parameters, and test data based on the former stage. The detailed design step requires repeated simulation tests to ensure that product prototype can achieve the desired performance. However, to make this simulation tests reliable, real-time data, and environmental-impacted data are necessary. DTs can provide that information, thus recording all data of the product and the influence of environment. This is possible by the fact that DTs exist and evolve in the whole lifecycle of the product.

- Virtual verification: DT-driven virtual verification greatly improves the design efficiency by avoiding tedious verification and testing. Traditionally, the design scheme is validated when carrying out small batches production after finishing product design. This extends the production cycle and increase the cost of time and money. Using DTs, it is possible to debug and predict accessories' quality directly in the model of DTs, before they are produced. Through DTs, it is possible to perform virtual verifications, permitting the identification of design defects, and its root causes, and then the redesigning can be done faster and convenient. This is done by taking full use of the data of equipment, environment, material, customers' physical characteristics, and history data of the last generation. This method can test whether there is a design defect and find the cause of it, and then the redesigning will be fast and convenient. Moreover, DTs can provide operation and service to optimize the auxiliary system and predict the physical objects based on virtual models. In this way, simulation tests can be effectively applied on prototypes and the actual performance of physical products can be predicted as far as possible [51].

DTs Benefits

DTs applications introduced by the literature review are potential improvements for processes in terms of efficiency and added value in manufacturing systems. However, the existent work had barely scratched the surface of the potentiality of DTs. It could be a marked resistance to invest in this technology, especially among SMEs, explained by the complexity in understanding the real contribution a DT can bring in their reality. Although, during the last years, main manufacturers and developers of DTs made an effort in explaining and communicating the benefits of this technology, its potentialities are not yet clear. Even in academic publications, perceived benefits are often listed, but rarely their impact is quantified. Some works available in literature [70, 98, 138] highlight the uncertainties of economic benefit, while others [46, 85] show that exists very few examples of quantification of scale and nature of benefits against existing processes and validation of tangible improvements. From the papers analyzed, the 74% of them showed at least one perceived benefit coming from the implementation of DTs in the manufacturing field. In most of the cases, those benefits are superficially explained, describing general improvement of the production like an increase in productivity and efficiency, but without accurately detailing the way DTs can contribute to these improvements. The literature review herein conducted clusters – and argues better in this section – the improvements into productivity, efficiency, knowledge, improve decision making, flexibility, monitoring, improve performances, cost reduction, quality, adaptability, reliability, sustainability, safety, personalization, time reduction, security, risk, availability, reconfigurability, new revenues streams/business model, improve product, decentralization, agility, connectivity. Those benefits were not singularly analysed but enclosed in few dimensions of performance where the company must perform at its best to be successful: (i) productivity, (ii) flexibility, (iii) knowledge, (iv) sustainability, (v) risk reduction, (vi) quality improvement, and (vii) business model innovation.

Productivity It is envisioned that DTs will improve productivity and efficiency of the production system [59]. Through the continuous monitoring of resources, it enables a more efficient utilization and consequently a reduction of costs. Its application allows to reach better performances with the same number of resources, by optimizing and continuously improving processes and reducing wastes. Moreover, DTs improve automation and autonomy of machines, shifting the most dangerous and dull activities from operators to machine. In this way, human will be able to focus on more creative and innovative jobs [76]. The literature analysis showed that 38 papers stress out advantages in productivity, making these benefit the most common connected to DTs.

Flexibility In the context of factories of the future, DTs are considered a good driver to improve flexibility, that is the capability of a process to react in a rapid way and with little costs to external changes [37, 49]. Another important aspect is the decentralization of the production [90]. DTs flatten the automation pyramid, leading to large scale distributed automation solutions and provides the production system with the ability to respond in autonomous way to unexpected changes, as for example the ability of self-recovery after a failure [73]. DTs do not provide only improvements in production flexibility, but also an increase of adaptability of systems [109]. If a new product is introduced, DTs facilitate the reconfiguration of processes by accelerating the time and reducing the costs of system's adjustments needed to meet the new requirements. Those benefits are discussed in 23 papers of the literature review.

Knowledge The third benefit in the context of DT is knowledge management. On an organizational level, DTs can be considered a technology that has the potential to improve the administration of knowledge for the benefit of companies. First, DTs allow real-time monitoring and control, generating a big amount of information about its physical counterpart. That information can be accessible anywhere and used to control the performance of the system even remotely. The availability of real-time information will improve the ability of companies to make decisions, leading to a faster and more informed decision-making [76, 118]. Then, this technology enables the visibility of operation processes and machines, thereby improving the transparency for stakeholders [93]. In the context of factory of the future, machines will make decision autonomously. So, they will both generate and apply knowledge, changing their role from enablers to active participants in value creation [139]. Twenty-four papers discuss the value of DTs in knowledge creation.

Sustainability In four papers [7, 83, 105, 117] found in literature, the benefits of DTs implementation were connected to the improvement of sustainability. As a matter of fact, DTs can be useful to monitor fuel consumption and conduct energy optimization strategy [7]. The continuous monitoring of resources enables a more efficient utilization and consequently a reduction of wastes. Moreover, this technology enables a higher visibility of operation processes and machines, thereby improving the transparency for stakeholder about how products are manufactured.

Risk Reduction The use of what-if analysis and risk assessments of potential scenarios, through smart analysis of real-time data, enables DTs to reduce the occurrences of problematic events like machines failures [76, 89]. Predicting issues before breakdown occur is a powerful way to reduce downtime and improve availability [62]. A better maintenance scheduling lower maintenance cost and time and consequently boosts performances [122]. In addition, in order to reduce costs, DTs can also protect the safety of workers by reducing asset and process related incidents [118]. Eleven papers emphasize the importance of DTs in risk mitigating, focusing mostly on predictive maintenance.

Quality Improvement A few papers insert quality as one of the benefits deriving from DTs. This technology can improve the quality of products in two-way. On the one hand, through an improved monitoring of activities performed by machineries, it is possible to better control the process and ensure better quality [85]. On the other hand, the information provided to the DT by the physical product is enhancing insight into the performances of products, their operating conditions, the preferences of various stakeholders, making it easier to improve the customization and the design [51].

Business Model Innovation In the last decades, companies started to realize that in response to changes in their environment, business model themselves became subject of innovation. A couple of recent papers [64, 80] highlight that DTs are an opportunity to discover new revenues streams and improve the current business model, shifting to product-centric to service-centric business model. Using opportunities of DTs to extend products across their products' lifecycles, providing after sales services, it is possible to gain in value creation [85]. In addition, this technology is a way to explore new business opportunities like mass customization and small-batch manufacturing.

AR in DT Applications

In the manufacturing context, DTs are used to have a constant up-to-date and accurate representation of a physical asset or process. These systems show values such as states, properties, or any relevant information that can be utilized for various ends, i.e., improvement of performance, security, quality, KPIs, etc. Usually in manufacturing, the access to the information and to the DT itself is done through terminals in machines or computers, which are not handy, easy to navigate or even may not be in the same location of the physical part of the DT. AR can solve this issue by allowing a seamless connection between the digital and the physical parts of the DT system.

Some works had been focused on the possibilities of the technology in different manufacturing environments. These works were gathered in the literature review process taking into account its inclusion of AR and DT in factories. More specifically, in some documents the utilization of AR is considered vital to improve machinery performances, improve assembly processes and reduce downtime and maintenance time through this technology [49, 124]. In [124], the application of AR displays relevant parameters to take into account while the maintenance process is being performed in manufacturing machinery. While other studies focus their efforts on different stages of the product lifecycle, such as assembly processes [49, 54, 128]. The implementation of AR is not only utilized to show relevant information on site, this with the aim to create a virtual DT capable to be displayed in any place and show accurate information of its physical counterpart in real time, and in case that the user is located in the physical location of the asset, it is also possible to position it on-site and align it on top of the real factory assets [123]. The paper argues that using ER with DT will help users (managers, supervisors, and technicians) to have a broader view of important data and hence, increase productivity and prevent downtime. Other works, propose different innovative models to implement DT in manufacturing environments, taking into consideration the different technologies of the I4.0 paradigm [47, 57, 129]. In these papers, frameworks for the implementation of DT-AR architectures are defined. Specifically, in [115, 129] the model focuses on the implementation of DT-AR in the machining process of materials. It searches to reduce quality issues in the end product. In both applications, the capacity of AR to display the data form the DT model is highlighted.

The result of this analysis shows how DTs applications can be supported by the integration of AR technologies. It was found that in the applications where these two technologies are integrated, AR has the objective to support the DT that usually lack of realism for human-centered design in manufacturing environments [49, 115]. The applications start from a virtual DT model that receive data from real world and simulates the actual state of the physical DT and/or forecasts states of it. AR, being a human-centered tool, acts as a key tool that can represent information and the system itself, allowing seamless interactions between users and the DT. In a nutshell, AR is utilized as a backbone to close the gap between digital and physical environments [140]. Nevertheless, some additional applications were found. In [126], AR is implemented as a complement of the DT model of robotic arms utilized for additive manufacturing. In this application, AR provides spatial location information (created by the DT) to the user by overlapping the objects in the location simulated, making easier for the user to locate the robots physically or to get their location in the digital world.

The final relevant aspect to consider is that the software that are most utilized in DT-AR applications are 3D modeling software with which the physical asset data is integrated. In this context, the different documents emphasize in the limitations in terms of data transfer speed that can limit the application. The hardware most utilized are mainly sensors that gather data from the field and the physical entity that varies in each application. Finally, the AR peripherals are heterogeneous and vary in quality, from phones to wearable holographic lenses.

Lifecycle Perspective

Conducting the analysis of articles, it has been noted that often DTs with a lifecycle perspective was proposed by authors. From the full-text assessment of papers emerged that the 40% of them had or proposed DTs under the lifecycle point of view. Those authors see the DTs as more valuable if employed during all the lifecycle of a system, as a production system or a product or an asset. Almost the half of the analysed papers suggest this perspective, however only few of them present the DTs lifecycle as the principal argument of the paper. Among this little cluster, the lifecycle perspective was found mainly in articles on DTs of product or asset. In fact, in previous section, in which the different type of DTs where described, the lifecycle point of view was assessed for these two types of DTs. These considerations suggested to conduct an exploratory analysis on case studies found in literature, with the aim to understand if a DTs lifecycle has ever been realized. DTs lifecycle is intended as DTs that are developed in the various phases of system lifecycle. Therefore, the explorative analysis has been performed by selecting the stage or stages of system lifecycle in which DTs were implemented in the specific case study. This way, each case study has been associated to one of three macro-phases of system lifecycle (i.e., BOL, MOL, and EOL).

The latter refer to the general steps in which a system evolve in its lifecycle. According to the type of system under consideration, each of these steps involve various sub-phases. In the Fig. 34.5, the possible subphases that characterize respectively the lifecycle of a product and the lifecycle of an asset, are represented.

The table below (Table 34.7) shows the classification of case studies in the three categories, respectively BOL, MOL, and EOL.

From the analysis of the case studies on DTs, it can be noted that in most of cases DTs application refers to just a phase of the lifecycle. Only three case studies out of thirty-one present a digital application in more than one phase. In particular, most of the case studies presented a DT implementation in the MOL phase, while few proposed a DT for the last stage of Lifecyle. Most of the implementation and use of DTs are specific for a lifecycle phase. However, the realization of ad-hoc DTs for a specific phase is difficult to be extended to other phases of the life cycle in the future. This orientation of DTs is in contrast with the concept of DT lifecycle.

34.4.2 DT Strategic Tool

The literature review revealed several applications and implementations of DTs that are currently being developed in manufacturing. From the literature review, it was possible to identify that in several documents, it is declared that DTs implementations can bring potential improvements to manufacturing systems. However, since DTs are still a blurry concept, it is difficult to have a precise understanding of its potentials and possible economic and operational impacts that its implementation can have. Therefore, the improvements and the benefits of the technology are still unclear, and the way DTs can contribute to these improvements is not accurately explained. The complexity in understanding the real DT contribution causes a lack of trust and a marked resistance to invest in this technology, especially among SMEs. Due to the lack of knowledge and competences, SMEs are struggling in realizing the benefits of the application of this technology. Moreover, considering the uncertainty of the return on investment, it is difficult for them to justify a substantial change by implementing DTs.

The literature has also shown that DTs have so far mainly been applied to one phase of the lifecycle. The negative point of this ad-hoc realization is that rarely a DT is scalable to other purposes. Considering all the financial and not-financial resources necessary for the realization and imple-

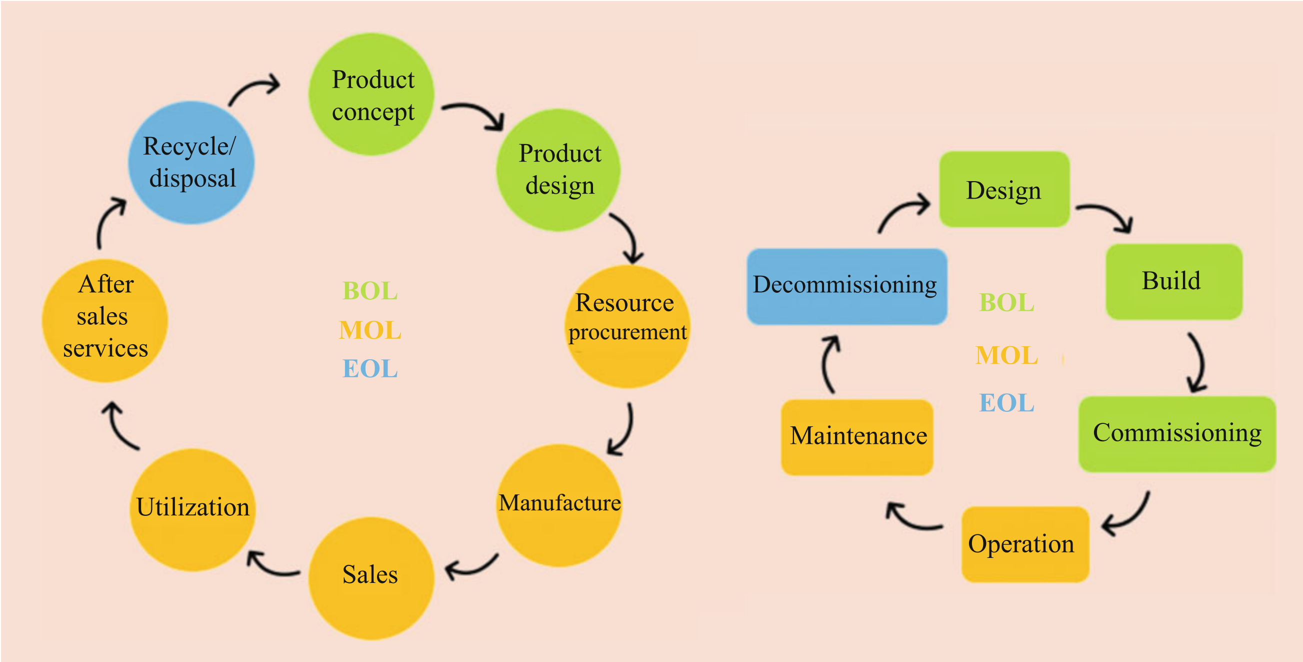


Fig. 34.5 Product and asset lifecycle

mentation of a DT, investing in application-purpose DT could not be an optimal strategy. Instead, a more valuable investment option could be the implementation of a DT that can be used for multi-purposes. DTs lifecycle seems to go in this direction. Despite in literature very few applications of DTs lifecycle were found, many authors highlight a possibility of improvement of the technology through its use in the entire lifecycle. However, since the concept is still little studied in literature, it remains obscure to companies who miss out on this investment opportunity.

From the above considerations, the following gaps emerged:

- The complexity in understanding the real contribution of DTs and its strategic impact for companies.
- The difficulty in identifying the kind of DTs application to employ in order to realize the objectives each company context requires.
- Unexplored opportunities of the implementation of scalable DTs multi-purposes; for this reason, its added value is not understood.
- The lack of a clear explanation on DTs lifecycle concept makes loses the comprehension.

These gaps are highly relevant as they represent key points to guarantee the dissemination of DTs concepts in the manufacturing industry. From one hand, clarifying the expectation and potentials of the technology, its acceptance level increases significantly. On the other hand, the level of comprehension of the technology plays a key role in its adoption level. Starting from the gaps highlighted, the objective of this work is to explain DTs contribution to improve company's performances by showing the link between DTs applications and benefits. In this way, the main scope is to provide a strategy tool that helps SMEs' managers to understand the strategic impact of DT and guide them in the identification of the proper DTs application to employ to reach their company's goals.

34

As already stated, the lack of transparency of benefits represents a relevant common barrier that prevent companies on adopting a I4.0 strategy and investing in technologies, as the DTs [141]. Indeed, since the DTs are still a blurry concept, it is difficult to have a clear and precise understanding of its potentiality and the economic impact of its improvement. SMEs innovation capabilities are often limited by lack of know-how, limited resources, and financial constraints. This is particularly true in the adoption of innovative technologies like I4.0 technologies and DT. However, technological innovation is essential to ensure the development of their business performances and to improve their sustainability, thus ensuring the success in the long term. Starting from this first result emerged from the literature review, it is important to show the link between DT application and benefits to explain the contribution of DTs in improving company's performances, with the final aim to bring companies closer to the technology and overcome their conservativism. Regarding DT adoption, SMEs are struggling in realizing the benefits of this technology and understanding the strategic impact of its implementation. The authors propose a strategic tool that explains which are the strategic impact and the benefits of

DTs, linking DTs applications with company's operational goal.

The use of tools is broadly deployed in management. However, there are no consistent definitions of it and the concept is often used interchangeably with other terms like techniques, methods and frameworks. For this work, the definition proposed by T. Brady and H. Rush [142] has been adopted: "a tool is a document, a framework, procedure, system or method which enables a company to achieve or clarify an objective. The objective may be a make or buy decision, a forecast, an analysis or one of many other tasks. A tool may be used by specific individuals or groups within a company or may be applicable across the entire company."

Starting from this definition, the aim of the proposed tool is to guide the managers of SMEs in the decision of investing or not in this technology. Therefore, the tool is addressed to practitioners that are considering investments in this technology and want to deepen the topic but have little knowledge on DT. The methodology implements a top-down approach starting from the identification of the market requirements, the identification of operational goals, and the analysis of the current operational status; progressing toward the analysis of DTs benefits and applications that leads to the selection of a tailored application to achieve company's goals.

Structure of the Tool

Company management can think to DTs implementation as a part of the company business strategy, that led to the creation of a competitive advantage. In this perspective, the aim is to develop an IT strategy that perfectly mirrors the company's business strategy. The adoption of DTs and, therefore, the digital transformation of manufacturing systems, should not be considered as a stand-alone solution, but rather incorporated with the strategic processes of the enterprise [141]. The implementation of DTs is a competitive lever that helps to reach operational goals. Its implementation should not be evaluated in monetary terms, showing its impact on the creation of value for the company and maximizing that business value, but looking at the operational performances that the company wants to improve. To this aim, SMEs should focus on one or few operational goals that they think are critical to their businesses, as well, SMEs should ask themselves if and how the DTs could facilitate the realization of those goals.

The tool is designed to help decision makers in identifying critical market characteristics and, after translating them into desired operational performances, to understand the degree to which DTs could support the operational objective of the firm. The tool begins with clear competitive market requirements and eventually leads to the suitable DTs application to be identified of interest for the company. The goal is to understand if the company needs to improve some operational performances and to evaluate if DTs can be a lever to this aim. In case the tool shows an alignment between the DTs benefits and the company's operational goals, the investment in this technology should be considered. Next, the tool supports the identification of a specific DT application to address those goals. On contrary, if the alignment between DTs benefits and company's objectives does not exist, considering the tool's perspective, an investment in DTs is not recommended.

To address this scope, the tool is composed by two levels, each one of which is then subdivided into two sections:

(i) Strategic level: This level is composed by two sections, the market requirements section and desired performances section.
(ii) Technological level: This level is also composed by two sections, the achievable performances section and applications section.

The structure of the tool is graphically shown in the Fig. 34.6. Each level is further explained in the next sections.

Strategic Level

For the strategic level, a practical methodology is suggested to translate market requirements into operational attributes that the company desire to achieve. This part is built around the concept of market-operation alignment. Numerous researchers have suggested the importance of achieving a strong fit between manufacturing and marketing in support of the overall corporate strategy [143]. This concept was first introduced by Skinner [144], who described the plant as a "competitive weapon because its entire apparatus is focused to accomplish the particular manufacturing task demanded by the company's overall strategy and marketing perspective." The description of the two sections of the strategic level are described below:

Market Requirements Today competitive markets are characterized by high turbulence, that is the rate of change in the composition of customers and their preferences. In most of the industries there are emerging trends and changing customer's expectations. SMEs managers should start from a deep analysis of market evolution and trends that characterize their industry. This analysis should be supported by a clear depiction of the customer's need and requirements. In addition, company's positioning and competitor's moves should be contemplated. The grouping of this set of information will help the company to identify the objective it needs to reach.

Desired Performances Once the market requirements are identified, managers should translate the desires of the customers into the performances their companies need to achieve. As mentioned before, the performances analysis should not be economic measures but operational ones,

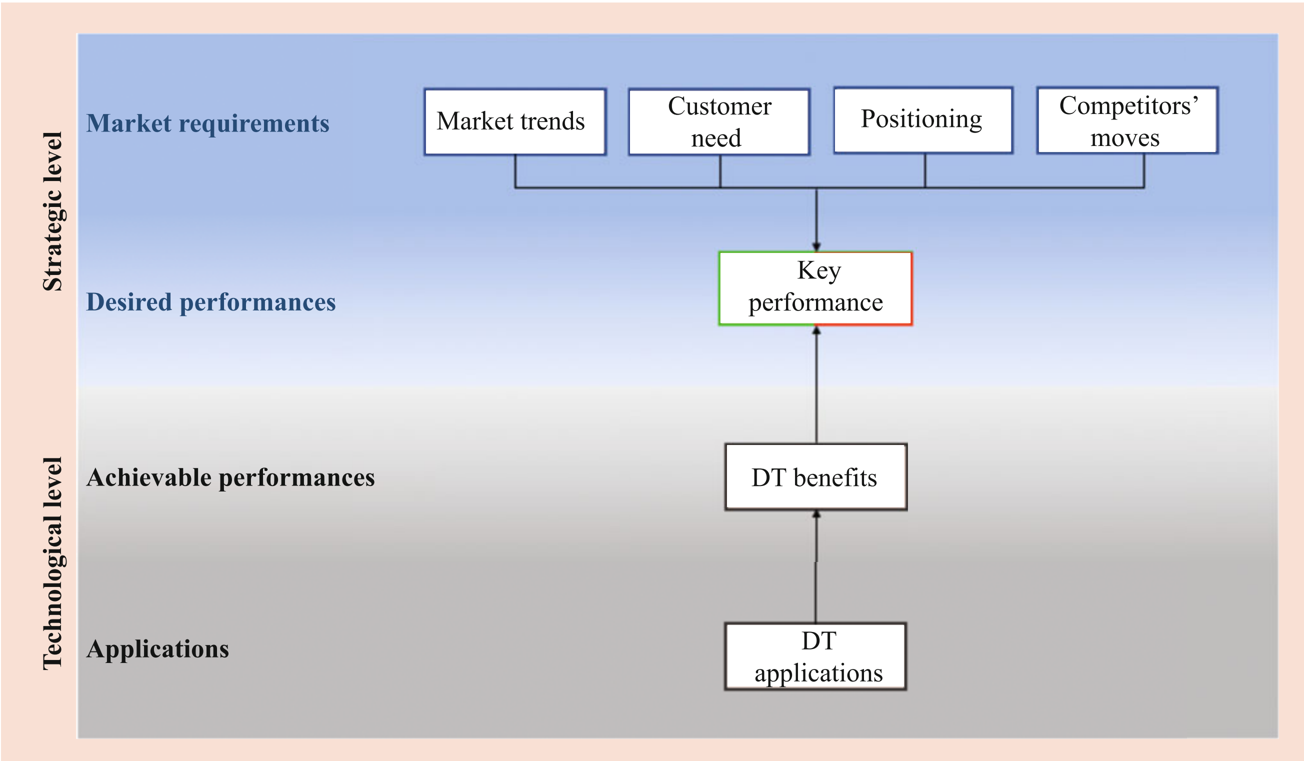


Fig. 34.6 DT strategic tool

deeply linked with processes of the company. To better assess the congruence between market requirements and operations, operational performances has been evaluated in light of customers' performance criteria.

There are eight competitive priorities, which fall into four group, that are used in the tool, all of which to a greater or lesser extent will affect customer satisfaction and business competitiveness:

- Cost
- Time: Time-to-market; Delivery speed; Delivery reliability
- Quality: Quality of design; Quality of conformance
- Flexibility: Customization; Volume flexibility

According to the idea of focus plant proposed by Skinner [144], operational infrastructure must be designed to serve a limited manufacturing task. Therefore, it is impossible and incorrect to try to maximize all those dimensions. Mangers should focus on the most important set of performances, checking the link with market requirements. This set of performances represent the desired performances that the company need to reach to succeed in the market. However, the company may not be able to meet these objectives already. Therefore, it must be defined the AS-IS situation, quantifying the actual state of the operational performances through an analysis of the processes. In this way, the tool is used to measure the readiness of the enterprise regarding the target state and, if some criticalities emerged, in order to quantify the gap between the current state and the desired performances. The performances that are required to improve are those that can potentially benefit from the introduction of DTs.

Technological Level

Within the technological level, DTs perspective is introduced. The development of this level refers to some results found in the research work done in literature review. The aim is to improve mangers' knowledge about DTs, showing what are the main operational advantages achievable through DTs and suggesting the appropriate application of DTs for their specific situation among the possible applications already implemented in manufacturing. Following, a brief explanation of the two sublevel included in the technological level:

Achievable Performances In these category the perceived and potential benefits of DTs are listed. The clusters identified in previous section are here reported:

- Productivity
- Flexibility
- Knowledge
- Quality improvement

34

Table 34.1 Document analysis for asset focused DT documents

Ref.	Year	Typology	Type DT	Focused area	Lifecyle perspective	Benefits
[62]	2018	Case study	Asset	Asset Lifecycle Management	X	Reliability; Availability; Maintainability
[63]	2019	Case study	Asset	Design	X	Knowledge; Reliability
[64]	2018	Concept	Asset	Maintenance		Availability; Efficiency
[65]	2017	Case study	Asset	Maintenance		
[66]	2019	Concept	Asset	Manufacturing in general		
[67]	2019	Concept	Asset	Asset Management		Improve performances
[68]	2020	Case study	Asset	Maintenance	X	Flexibility; Time reduction; Efficiency
[69]	2019	Concept	Asset	Maintenance		

Table 34.2 Document analysis for factory focused DT documents

Ref.	Year	Typology	Type DT	Focused area	Lifecyle perspective	Benefits
[70]	2017	Case study	Factory	Factory Design	x	Uncertainties of economic benefits; Knowledge
[71]	2018	Case study	Factory	Design		
[72]	2017	Case study	Factory	Design		Quality
[73]	2018	Case study	Factory	Design		

- Risk reduction
- Sustainability

It has been decided to not include the last cluster (business model innovation), since it is not directly linked with operative performances. These listed benefits were denoted as possible to achieve by companies through the application of DTs.

Applications With the analysis performed and explained in previous sections it was possible to identify the most common applications of DT in the manufacturing area for each type of DT. The first column of the Table 34.8 reports all the type of DT applications found in literature. The aim of the technological level is not only to list the main benefits and DT applications, but to explain and show the relationship between them. In order to do this, each application has been analysed and connected to one or more benefits. The aim of this interrelationship it is to inform managers of the operative opportunities of each DT application.

The methodology used to create this connection is the following. Firstly, starting from the tables created during the systematic literature review (from Tables 34.1, 34.2, 34.3, 34.4, and 34.5), each case study was analysed to spot the peculiar application implemented and, if possible, to connect the achieved benefits. When this was possible, the connection between application and benefits was solid, since there was practical evidence of its relationship. Secondly, in order to connect the remaining applications, the information contained in the concept-type papers and reviews were used in combination with the analysis and evaluation of the authors.

After the identification of the DT benefits that the final user would gain, the application that mostly cover those benefits can be selected.

34.5 Practical Implementation of a Combined DT-ER Solution

In the previous sections, an overview of the main technologies to develop a DT have been presented, together with a systematic literature review aiming to understand how widespread this technology in manufacturing contexts is and which are its different types, applications, and benefits. In this section, an evaluation of the benefits deriving from a practical DT implementation that is present in literature is showed, trying to apply the classification and the results coming from the literature review. The evaluation analysed is only one of the possible applications of the technology; nevertheless, its implementation shows great benefits that companies can get from DTs/ER technologies. This way, the main objective is to show managers and professionals one practical project in which these technologies can be exploited. The final goal is to motivate SMEs to understand DTs-ER technologies and additionally, to identify and address their specific challenges taking advantage of them. The implementation considered is a digital solution that combines DT and ER through a laboratory application scale [7]. In this laboratory case, VR is utilized a first step in the adoption of further ER technologies. In the literature, review was found that the implementation of technology variations such as AR start with the construction of a virtual model of the physical system [115]. The reference application case has been developed, implemented and

Table 34.3 Document analysis for documents centred in general DT analysis

Ref.	Year	Typology	Type DT	Focused area	Lifecyle perspective	Benefits
[12]	2017	Review	General	Manufacturing in general	X	
[59]	2018	Review	General	Manufacturing in general	X	Efficiency
[74]	2016	Concept	General	Manufacturing in general	X	
[75]	2019	Concept	General	Manufacturing in general	X	
[76]	2020	Review	General	Manufacturing in general		Improve decision making; Efficiency; Safety; Risk reduction; Personalization
[77]	2019	Concept	General	Manufacturing in general		Knowledge
[78]	2018	Review	General	Manufacturing in general	X	Personalization; Efficiency
[47]	2017	Concept	General	Manufacturing in general		Reliability; Availability; Maintainability
[79]	2018	Concept	General	Manufacturing in general	X	Time reduction
[80]	2019	Review	General	Manufacturing in general	X	Cost reduction; Reliability; Efficiency
[81]	2019	Review	General	Manufacturing in general		Monitoring; Flexibility
[82]	2017	Concept	General	Manufacturing in general	X	
[83]	2019	Review	General	Manufacturing in general		Sustainability
[84]	2018	Review	General	Manufacturing in general	X	Productivity
[46]	2020	Review	General	Manufacturing in general		Cost reduction; Security; Risk reduction; Time reduction; Reconfigurability; Safety; Reliability; Efficiency; Flexibility
[85]	2019	Concept	General	Manufacturing in general	X	Agility; Improve product; New revenues streams/business model
[86]	2017	Concept	General	Real-time monitoring		Improve decision making
[87]	2019	Concept	General	Manufacturing in general		Cost reduction; Efficiency; Time reduction
[88]	2020	Concept	General	Manufacturing in general	X	Knowledge
[89]	2018	Concept	General	Configuration	X	Improve decision making; New revenues streams/business model; Improve quality; Improve performances; Personalization
[90]	2018	Concept	General	Manufacturing in general	X	Knowledge; Agility
[83]	2019	Review	General	Manufacturing in general		Monitoring; Improve performances; Connectivity
[91]	2018	Concept	General	Manufacturing in general		Productivity; Efficiency
[92]	2019	Concept	General	Manufacturing in general		Productivity; Efficiency; Time reduction; Cost reduction
[93]	2019	Review	General	Manufacturing in general	X	Efficiency; Productivity
[94]	2019	Review	General	Manufacturing in general		Efficiency; Time reduction; Improve performances; Adaptability; Quality improvement; Reconfiguration; Sustainability

Table 34.4 Document analysis for process focused DT documents

Ref.	Year	Typology	Type DT	Focused area	Lifecyle perspective	Benefits
[95]	2016	Case study	Process	Simulation Optimization		Productivity
[96]	2017	Concept	Process	System Optimization		Uncertainties of economic benefits
[97]	2018	Case Study	Process	Production Planning & Control		Knowledge
[98]	2017	Concept	Process	Process Optimization	x	Uncertainties of economic benefits
[99]	2018	Case study	Process	Production Planning & Control		Knowledge
[100]	2019	Case study	Process	Process monitoring and diagnosis		Security
[101]	2019	Case study	Process	PPC		
[102]	2017	Case study	Process	Maintenance		
[103]	2017	Case study	Process	Design	x	Efficiency; Adaptability
[56]	2015	Concept	Process	Manufacturing in general	x	Improve decision making; Flexibility
[104]	2018	Case study	Process	Process Improvement		Productivity
[105]	2019	Case study	Process	Reconfiguration	x	Knowledge; Efficiency; Sustainability
[7]	2020	Case study	Process	Energy monitoring		Sustainability
[106]	2018	Concept	Process	Process Optimization		
[107]	2020	Concept	Process	Root-cause analysis		Improve performances
[108]	2018	Concept	Process	Real-time monitoring	x	Flexibility, Efficiency
[109]	2018	Case study	Process	System optimization		Adaptability; Improve performances; Flexibility; Improve quality
[110]	2018	Concept	Process	Process improvement		Efficiency; Improve quality
[111]	2020	Case study	Process	Process planning		Efficiency
[112]	2019	Case study	Process	Scheduling		Efficiency
[113]	2019	Case study	Process	Process Control		
[13]	2019	Case study	Process	Process Control		Monitoring; Decentralization; Flexibility
[114]	2018	Case study	Process	Process improvement		
[115]	2019	Case study	Process	System Control		Efficiency; Improve decision making
[116]	2019	Case study	Process	System Control	x	
[117]	2019	Case study	Process	Process improvement		Knowledge
[58]	2019	Review	Process	Manufacturing in general		
[118]	2018	Case study	Process, People	Process improvement		Improve product; Safety; Cost reduction; Accuracy

Table 34.5 Document analysis for product focused DT documents

Ref.	Year	Typology	Type DT	Focused area	Lifecyle perspective	Benefits
[51]	2017	Case study	Product	Manufacturing in general	X	Efficiency
[119]	2018	Concept	Product	Manufacturing in general		Connectivity
[120]	2019	Case study	Product	Manufacturing in general		Efficiency; Flexibility; Knowledge
[121]	2019	Review	Product	Product lifecycle	X	Monitoring
[121]	2019	Review	Product	Product lifecycle	X	Monitoring
[122]	2017	Concept	Product	Design; Real-time control	X	Safety; Reliability

tested at the Industry 4.0 Lab of the School of Management of Politecnico di Milano [145], pertaining to the research activities funded by the H2020 FENIX project [7]. Within the H2020 FENIX project, the Industry 4.0 Lab has been used for demonstrating how several smart technologies can be exploited together for testing, managing, and optimizing a WEEE disassembly process for the recovery of valuable resources under a Circular Economy (CE) perspective.

According to the categorization of DT type presented in Sects. 34.3 and 34.4, the application case selected belongs to the "DT Process" type, while the main focused area on which it refers are:

Table 34.6 Analysis of documents centered specifically in applications of DT and AR

Reference	Year	Industrial field	Focused area	MR Type	AR characteristic for DT	Benefits	Research type	Hardware	Software	ER peripherals
[123]	2017	MN	P	AR	DV	Reduced downtime, improved productivity	E	Sensors, Actuators	–	HoloLens
[124]	2009	MN	MRO	AR	IV	Reduced break-up and maintenance downtime	E	Sensors	–	Wrist-worn controller, HL
[47]	2017	MN	P	AR	IE	Enhanced data visualization, Improved resource allocation, improved productivity	MP	Sensors, Actuators	CAD	–
[78]	2018	MN	MRO	AR	DV	Enhanced data visualization and agility for MRO	MP	-	–	–
[125]	2019	G	IS	AR	DV	Enhanced data visualization, safety and security, alarm functionalities	E	Strain Gauges, Wheat-stone bridge, Arduino Card	ThingSpeak	F4 Glasses
[57]	2019	G	P	AR	DV	Enhanced data visualization and monitoring	SC	Tablets, mobile devices	SDK Vuforia	Tablets, mobile devices
[49]	2019	G	ASB	AR, VR	IV	Enhanced data visualization and agility for MRO	SC	–	Catia, Modelica	Tablets
[126]	2020	G	AM	AR	PCV	Improved productivity, security, avoidance of collision possibilities	E	Camera, Robotic arms	SolidWorks	Display
[127]	2020	G	DTaaS	AR	DV	Reduced break-up and maintenance downtime, Didactic capability enhanced	E	Sensors, actuators	Rhinoceros, Vuforia	Mobile devices
[128]	2019	MN	ASB	AR	DV	Enhanced data visualization and agility for MRO	LR	–	–	–
[129]	2021	MN	AM	AR	DV	Increased quality, reduced waste	E	Sensors, strain gauges	–	Holographic Lens
[54]	2019	MN	ASB	AR, VR	IDV	Optimized assembly process	SC	Sensors, projec-tors, laser projec-tors, laser trackers	CAD	Holographic Lens
[130]	2018	MN	MRO	AR	IDV	Enhanced data visualization and agility for MRO and configuration support	SC	Terminal regulator	NX-MCD by Siemens	Tablets
[115]	2019	MN	P	AR	DV	Improved productivity, security, increased quality	E	Sensors, CNC machine	Unity 3D	HoloLens

MN Manufacturing, *G* General, *P* Production, *MRO* Maintenance repairment and Overhaul, *IS* Industrial Safety, *DTaaS* Digital Twin as a Service, *ASB* Assembly, *AM* Additive Manufacturing, *DV* Data visualization, *IV* Instruction visualization, *IE* Immersive environment *PCV* Positioning coordinates visualization, *IDV* Instructions and Data visualization, *MP* Model Proposal, *E* Experiment, *SC* Study case, *LR* Literature review

Table 34.7 Lifecycle stages classification

	BOL	MOL		EOL
Ref.	Design	Operation	Maintenance	Disposal/Recycling
[70]	x			
[95]	x			
[97]		x		
[99]		x		
[100]		x		
[101]		x		
[51]	x			
		x		
				x
[102]			x	
[71]	x			
[103]	x			
[104]		x		
[120]		x		
[105]	x	x	x	x
[62]	x		x	
[72]	x			
[63]	x	x		
[7]				x
[118]		x		
[109]		x		
[65]			x	
[68]			x	
[111]	x			
[112]		x		
[113]		x		
[13]		x		
[114]		x		
[73]	x			
[115]	x			
[116]	x			
[117]		x		

Table 34.8 Link between manufacturing applications and benefits

	P	F	K	R	Q	S
DT for manufacturing asset						
Configuration			✓✓		✓✓	
Reconfiguration		✓✓	✓✓			
Simulation			✓			
Machine state monitoring		✓	✓✓	✓		
Operative conditions monitoring			✓	✓✓		
Proactive maintenance	✓			✓✓		
Support fault diagnosis and life prediction				✓✓		
Fuel and energy monitoring						✓✓
Asset Lifecycle Management	✓			✓✓		
DT for process						
Simulation			✓			
Visualization			✓			
Status monitoring			✓	✓		
Identify root causes			✓✓			
Diagnosis				✓✓		
PPC	✓✓		✓✓			
Optimization	✓✓				✓✓	
Identify causes and reconfigure parameters		✓✓	✓			
Modify parameters to deal with contingencies		✓✓				
Tolerance control					✓✓	
Design	✓✓	✓✓				
Real-time monitoring			✓	✓		
Improvement	✓✓				✓✓	
Process Control		✓✓	✓✓			
DT for factory						
Factory layout			✓✓			
Factory design	✓	✓				
DT for people						
Training tool	✓✓					
Modelling human tasks				✓✓		
DT for product						
Conceptual design, detailed design and Virtual verification	✓✓	✓✓				
Real-time monitoring			✓✓	✓✓		
Optimization	✓✓				✓✓	
Simulation			✓			
Faults and life prediction				✓✓		
Identify root causes			✓✓			

P productivity, *F* flexibility, *K* knowledge
R risk reduction, *Q* quality improvement, *S* sustainability
✓✓ means that the link is based on case study analysis (solid connection).
✓ means that the link is based on concept-papers, reviews and author considerations.

Simulation optimization: The development of the combined DT-ER simulation models for a disassembly process optimization presented by Rocca et al. [7] are: i) one manufacturing system design simulation model (i.e., ER-based disassembly process configuration model); and (ii) one manufacturing system operation simulation model (i.e., ET-based real-time process optimization tool) [146].

Process optimization: The main scope of the combined solution was to optimize the production process (i.e., disassembly process) based on real-time simulation data, real-time production data, and historical production data. According to the analysis of this data and to the machine states simulation, the real-time controlling instructions are fed back to the physical space to optimize the disassembly process.

- Reconfiguration: With the aim to practically demonstrate the benefits coming from ER solutions integrated to DT, the ER-based configuration tool has been developed for supporting disassembly processes reconfigurability and implementation. In particular, a manufacturing line originally designed for assembly processes has been reconfigured as disassembly line. With the modification of the manufacturing line, it was possible to exploit a virtual environment where simulating and optimizing a disassembly process before efficiently replicating it on the real world [7]. In order to do this, the authors have selected CIROS® Studio 6.0 as reference software [147]. The virtual disassembly process configuration tool has been implemented by following five steps: (i) disassembly line modelling; (ii) disassembly process design; (iii) robotic disassembly program coding; (iv) disassembly workplan creation within the Manufacturing Execution System (MES) software and process simulation; (v) disassembly configuration uploading on the real system.

- Energy monitoring: one of the test presented in the work [7] is related to the energy consumption optimization of the disassembly process. The authors develop a new energy-based KPI (e-OEE) able to evaluate the energetic performance of the system. This indicator has been introduced in the DT-ER solution, which is able to extract all the values about the real-time energy consumption using an accumulator function. At the end, thanks to the Graphical User Interface (GUI) created, it has been possible to have a clearer and readable way to find useful information for the energy management. Monitoring of parameters such as energy are intended to be integrated with AR the future developments of the project.
- Real-time monitoring: the DT-ER solution developed at the Industry 4.0 Lab has been built by exploiting a particular Machine-to-Machine (M2M) communication protocol compliant with the IEC 62541 standard: the OLE for Process Control (OPC) Unified Architecture (OPC-UA) communication protocol [7]. This protocol allows a real-time information exchange with sensors and actuators of the real system, between machines and an external simulation environment [148]. Thanks to its client-server architecture, OPC-UA protocol enables the communication between industrial equipment and systems for a continuous data collection and control, by creating a virtual mirror of the system. The main capabilities that have been add to the existent manufacturing line are the possibility to: (i) identify machine states (e.g., errors, failures, and downtimes); (ii) identify sensors and actuators states/values (e.g., presence of products, temperature, or air pressure); (iii) monitor and control operations performances; (iv) real-time analysis of signals; (v) define maintenance plans of machines; (vi) store all the gathered data; and (vii) execute data analytics on operational and energy parameters.

Among the DT benefits analyzed and presented in Sect. 34.5, it is possible to state that the experiment presented in [7], allow to optimize the process under these perspectives:

Flexibility and Risk Reduction DTs are considered a good driver to improve flexibility, that is the capability of a process to react in a rapid way and with little costs to external changes [37, 49]. In order to exploit the line to disassemble the obsolete products, DTs facilitate the reconfiguration of processes by accelerating the time and reducing the costs of system's adjustments needed in order to meet the new requirements. The main function related to flexibility and embedded in the combined DT-VR solution is the ability to identify the conditions in which the reactive scheduling of a disassembly process is needed. In fact, in order to avoid the scrapping of a component that is not compliant with some standards, a disassembly work plan can be reactively scheduled [7]. This greatly reduces the risk of discarding non-compliant productions exploiting the reactive reconfigurability of the system.

Sustainability From the literature review, it has been highlighted that DTs can be useful to monitor fuel consumption and conduct energy optimization strategy. The continuous monitoring of resources enables a more efficient utilization and consequently a reduction of wastes. In this sense, the practical implementation has been analyzed proposes an introduction to two of the main benefits that I4.0 allows to reach for boosting CE, i.e.: (i) process effectiveness, minimizing resource flows from the BOL to the EOL by generating a cyclic metabolism, allowing the materials to maintain their original state; and (ii) process efficiency, by minimizing volume and consumption of both energy and material resources. This is implemented in this work by energy data collection and KPI systems creation for decision-making process [7].

Business Model Innovation Regarding this benefit, the laboratory application case showed in [7] represents a clear evidence on how companies can implement circular business models optimizing the disassembly process of obsolete product. This allows to recover valuable materials and to activate new business revenues. In this sense, the application case aimed at demonstrating how the adoption of CE principles can enable more sustainable supply chains, by increasing quality, market value, and alternative exploitation of secondary materials.

It is important to highlight the future development of the project described previously. In Sect. 34.3 it was found that ER can bring numerous benefits to manufacturing applications when it is combined with DT technologies. As it is defined by [115], it is possible to define a DT-ER application by considering five main components, i.e., the physical part, virtual part, control process, augmented process, and the AR calibration phase. In the laboratory case analyzed, the physical part includes the physical disassembly machinery of the laboratory including all sensors and data acquisition devices. The virtual part includes the digital model of the disassembly process and the data gathered from it. The control process in the described application is composed by the manual interactions of the users with the machinery and the intelligent control from the DT, i.e., the capacity to reconfigure the line from assembly to disassembly. Based on this framework, the application described in this section counts with all the characteristic to further develop the project. The reference application are as future objective the implementation of the augmented process component and the calibration phase of the model, which will integrate in a seamless way the physical and virtual parts. The main aim of the new developments inside the project are to show the data of the DT-ER application in real time through the utilization of AR, with the final objective to enhance variables control, Human to Machine (H2M) interaction, and the decision-making process.

34.6 Discussion

Thanks to the analysis performed with the literature review, it is possible to state that the decision to implement DTs/ER technologies is not only based on an economic perspective. In fact, the enterprises' competitiveness in manufacturing environments is based on many additional factors such as time, flexibility, and quality, between others, that depend on each company specialization. The tool proposed in this chapter can be implemented by any company, regardless their operational field. With the aim of contributing to fill the gaps identified in literature, the tool contributes to the identification of competitive companies' weaknesses that can be addressed with these technologies. Moreover, it gives a valuable contribution on the decision-making process for SMEs. This is achieved through the identification of the strategic impacts of DTs technology within the company, exposing the different types of DTs depending on the companies' strategic impacts required and finally, by presenting and clarifying the concept of DTs lifecycle.

The tool presented in this chapter was designed as a practical and intuitive guide to show the potentiality of DT to mangers for their specific company situation. Although the tool accurately addresses the gaps found in literature, it has some limitations that must be mentioned:

- The described approach provides only qualitative and subjective decisions, that are influenced by the ability of the decision makers involved in the evaluation and may not offer an optimal decision.
- The tool does not pretend to be inclusive of everything. The benefits and applications individuated through the literature review can be incomplete. Considering that DT is a new technology and its use in manufacturing is quite recent; many benefits and applications are not yet discovered. The ability of early adopters is to find new opportunities not yet explored.
- DT is not only useful as a competitive lever to reach operational goals that are previously identified by a plan. However, in high turbulent environment, it is not possible to plan everything in advance. The new technology has the potentiality to spot new emergent strategies, that are "patters evolved in the absence of intentions" [149]. In the literature review, it is showed that the DT is changing business model and the way value is created by companies. Often, these new possibilities are difficult to foresee, but emerge with the application of the technology. This perspective is missing in the proposed tool.
- The tool guides in the decision of investment of DT. However, in case of the recognition of a need of investing in the technology, the tool does not show a migration strategy to follow in the implementation.

Even though the limitations presented are relevant, this chapter also highly contributes to the theoretical DTs research. After gathering highly dispersed information centred on DTs and performing a detailed analysis on it, the main applications and benefits of DTs were identified, together with the lifecycle perspective of the technology. From this, the tool derived has an important impact in terms of technology understanding, but still, it is open for further academic development for specific manufacturing areas. Finally, it has been also analyzed a practical implementation of the technology, with the scope to show the different benefits that can be obtained from one of its possible applications.

From a managerial perspective, the current chapter develops an analysis and a tool that has a great potential of impact for SMEs. Usually, SMEs struggle in the process of identification of technologies that can support their digital transformation. Taking this into consideration, the need to exhibit the applications and benefits of technologies such as DTs is critical to promote competitiveness and development of SMEs.

34.7 Conclusion

This chapter introduces the DT concept, also presenting an overivew of the technologies enabling its adoption (i.e., CPS, IoT, simulation, ER).

To further ground these results from a theoretical perspective, a literature review focusing on DT has also been conducted. First, it detected five main possible categories on which DTs could be applied in manufacturing (asset, process, factories, people, product). Second, it theoretically identified the possible benefits provided by the adoption of DTs technologies (i.e., (i) productivity; (ii) flexibility; (iii) knowledge;

(iv) sustainability; (v) risk reduction; (vi) quality improvement; (vii) business model innovation). Finally, with the literature review, it was possible to analyze the impact of this kind of technologies on the entire asset lifecycle, and, based on the main gaps found in literature, it provided the main directions to the development of a DT strategic tool. Indeed, although DT draws big attentions in the research field, its real contribution is unclear for companies, especially for SMEs, causing a resistance to invest in the technology, and hence, a delay in the digitalization process of these type of companies. The strategic tool proposed in this chapter works as a valuable instrument to help managers in understanding the potentials of the DT in their peculiar situation, and even make it clear when the technology could be implemented. This is done by showing a clear link between DT application and the performances that the company is interested to achieve.

Supporting this intertwined relationship among these different technologies belonging to the I4.0 paradigm, a real application case related to a combined DT-ER development has been analyzed. The application described provides practical evidence of what has to be considered to develop and implement in a manufacturing context a combined DT-ER solution (from the models and process design up to the program coding and workplan creation through the MES up to the final uploading of the process configuration on the real system). The case is a lab-scale application exploiting all I4.0 technologies together for managing and optimizing a WEEE disassembly process through a DT powered by ER. Through its evaluation, it was possible to unveil the practical benefits that could be achieved through the adoption of a combined DT-ER solution in the manufacturing domain. Aditionally, the need to further implement complementing ER technologies such as AR was identified. The goal is that the laboratory case analyzed can take complete advantage of the new posibilities that ER brings to manufacturing.

Wrapping up, thanks to a new integrated data management along the automation pyramid and thanks to the industrial automation improvements introduced by the fourth industrial revolution, the experiment evaluate in Sect. 34.5 proposes an introduction to some of the main benefits that DTs allow to reach among those presented in Sect. 34.4. The exploitation of these tools and techniques is made possible thanks to the introduction of CPSs and I4.0 technologies making more flexible the automation pyramid and representing the way to create powerful simulation models and a better resource monitoring tool.

The chapter is aimed to improve the dissemination of DT concept in the manufacturing industry, supporting the recognition of opportunities and the level of understanding. DTs is a high relevant concept for companies that walk the path in the I4.0 paradigm, and its implementation can bring great opportunities in companies' environment. Addressing a series of gaps emerged from the literature review, the main objective of the chapter has been to spread the knowledge and the value of the technology to manufacturing industry, as well to single companies.

References

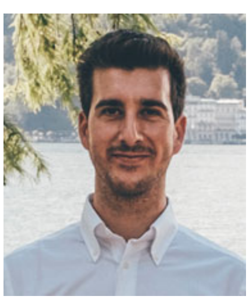
1. Ambra Cala, A. L., Boschi, F., Tavola, G., Taisch, M.: Migration towards digital manufacturing automation – An assessment approach. In: Proceedings – 2018 IEEE Industrial Cyber-Physical Systems, ICPS 2018 IEEE, St. Petersburg, 2018, pp. 714–719.
2. ElMaraghy, H.A., Wiendahl, H.-P.: *Changeable and Reconfigurable Manufacturing Systems*. Springer, London (2009)
3. Hofmann, E., Rüsch, M.: Industry 4.0 and the current status as well as future prospects on logistics. *Comput. Ind.***89**, 23–34 (2017)
4. Calà, A., Boschi, F., Fantini, P., Lüder, A., Taisch, M.: Migration strategies towards the digital manufacturing automation. In: Soldatos, J., Lazaro, O., Cavadini, F. (eds.) The Digital Shopfloor: Industrial Automation in the Industry 4.0 Era, pp. 365–391. River Publisher (2019)
5. Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M.: How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective. *Int. J. Sci. Eng. Technol.***8**(1), 37–44 (2014)
6. Burke, R., Mussomeli, A., Laaper, S., Hartigan, M., Sniderman, B.: The smart factory. (2017). https://www2.deloitte.com/us/en/insights/focus/industry-4-0/smart-factory-connected-manufacturing.html .
7. Rocca, R., Rosa, P., Sassanelli, C., Fumagalli, L., Terzi, S.: Integrating virtual reality and digital twin in circular economy practices: a laboratory application case. *Sustainability*. **12**(6), 2286 (2020)
8. ANSI/ISA: Enterprise-Control System Integration – Part 1: Models and Terminology (2010).
9. Chiarello, F., Trivelli, L., Bonaccorsi, A., Fantoni, G.: Extracting and mapping industry 4.0 technologies using wikipedia. *Comput. Ind.***100**(September 2018), 244–257 (2018)
10. Culot, G., Nassimbeni, G., Orzes, G., Sartor, M.: Behind the definition of industry 4.0: analysis and open questions. *Int. J. Prod. Econ.***226**(January), 107617 (2020)
11. Bastidas, R.F., Rocca, S.R., Negri, E., Fumagalli, L.: A review of features and applications of distributed ledger technologies for smart manufacturing. *Int. J. Ind. Syst. Eng.* (2021)
12. Negri, E., Fumagalli, L., Macchi, M.: A review of the roles of digital twin in CPS-based production systems. *Procedia Manuf.***11**, 939–948 (2017)
13. Cimino, C., Negri, E., Fumagalli, L.: Review of digital twin applications in manufacturing. *Comput. Ind.***113**, 103130 (2019)
14. National Institute of Standards and Technology (NIST): Product definitions for smart manufacturing. (2018).
15. ANSI/ISA: Enterprise-Control System Integration – Part 1: Models and Terminology. (2010).
16. Pathfinder project: Pointing Out Research Priorities Leading To The Next Generation of Simulation and Forecasting Technologies (2014).
17. Fumagalli, L., Pala, S., Garetti, M., Negri, E.: Ontology-based modeling of manufacturing and logistics systems for a new MES architecture. *IFIP Adv. Inf. Commun. Technol.***438**(PART 1), 192–200 (2014)
18. Lee, J., Bagheri, B., Kao, H.A.: A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf. Lett.***3**(January 2015), 18–23 (2015)

19. Jazdi, N.: Cyber physical systems in the context of Industry 4.0. In: Proceedings – 2014 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR) 2014 IEEE, Cluj-Napoca, 2014, pp. 1–4.
20. Baheti, R., Gill, H.: Cyber-physical systems. *Impact Control Technol*. **12**(1), 161–166 (2011)
21. Stock, T., Seliger, G.: Opportunities of sustainable manufacturing in industry 4.0. *Procedia CIRP*. **40**, 536–541 (2016)
22. de Sousa Jabbour, A.B.L., Jabbour, C.J.C., Filho, M.G., Roubaud, D.: Industry 4.0 and the circular economy: a proposed research agenda and original roadmap for sustainable operations. *Ann. Oper. Res.***270**, 273–286 (2018)
23. Zorzo, A. F., Nunes, H. C., Lunardi, R. C., Michelin, R. A., Kanhere, S. S.: Dependable IoT Using Blockchain-Based Technology. In: 2018 Eighth Latin-American Symposium on Dependable Computing. IEEE, Foz do Iguaçu2018, pp. 1–19.
24. Roblek, V., Meško, M., Krapež, A.: A complex view of industry 4.0. *SAGE Open*. **6**(2), 1–11 (2016)
25. Davis, J., Edgar, T., Porter, J., Bernaden, J., Sarli, M.: Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Comput. Chem. Eng.***47**(December 2012), 145–156 (2012)
26. Lee, J., Kao, H., Yang, S.: Service innovation and smart analytics for Industry 4.0 and big data environment. *Procedia CIRP*. **16**, 3–8 (2014)
27. Yang, S., Raghavendra, M.R.A., Kaminski, J., Pepin, H.: Opportunities for industry 4.0 to support remanufacturing. *Appl. Sci.***8**(7), 1177 (2018)
28. De Man, J.C., Strandhagen, J.O.: An industry 4.0 research agenda for sustainable business models. *Procedia CIRP*. **63**, 721–726 (2017)
29. Bressanelli, G., Adrodegari, F., Perona, M., Saccani, N.: Exploring how usage-focused business models enable circular economy through digital technologies. *Sustain.***10**(3) (2018)
30. Pagoropoulos, A., Pigosso, D.C.A., McAloone, T.C.: The emergent role of digital technologies in the circular economy: a review. *Procedia CIRP*. **64**, 19–24 (2017)
31. Chang, M.M.L., Ong, S.K., Nee, A.Y.C.: Approaches and challenges in product disassembly planning for sustainability. *Procedia CIRP*. **60**, 506–511 (2017)
32. Chen, M., Mao, S., Liu, Y.: Big data: a survey. *Mob. networks Appl.***19**(2), 171–209 (2014)
33. Fishman, G.S.: *Principles of Discrete Event Simulation*. John Wiley & Sons, Inc., New York, NY (1978)
34. Banks, J.: *Handbook of simulation – Principles, Methodology, Advances, Applications, and Practice*. Engineer & Management Press (1998)
35. Goodall, P.A., Sharpe, R.G., West, A.A.: A data-driven simulation to support remanufacturing operations. *Comput. Ind.***105**, 48–60 (2019)
36. Negahban, A., Smith, J.S.: Simulation for manufacturing system design and operation: literature review and analysis. *J. Manuf. Syst.***33**(2), 241–261 (2014)
37. Weyer, S., Meyer, T., Ohmer, M., Gorecky, D., Zühlke, D.: Future modeling and simulation of CPS-based factories: an example from the automotive industry. *IFAC-PapersOnLine*. **49**(31), 97–102 (2016)
38. Legat, C., Seitz, C., Lamparter, S., Feldmann, S.: Semantics to the shop floor: Towards ontology modularization and reuse in the automation domain. *IFAC Proc.***19**, 3444–3449 (2014)
39. Negri, E., Fumagalli, L., Garetti, M., Tanca, L.: Requirements and languages for the semantic representation of manufacturing systems. *Comput. Ind.***81**, 55–66 (2016)
40. Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., De Amicis, R., Pinto, E.B., Eisert, P., Döllner, J., I.: Vallarino: visual computing as a key enabling technology for industrie 4.0 and industrial internet. *IEEE Comput. Graph. Appl.***35**(2), 26–40 (2015)
41. Wang, X.V., Wang, L.: Digital twin-based WEEE recycling, recovery and remanufacturing in the background of industry 4.0. *Int. J. Prod. Res.***57**(12), 3892–3902 (2019)
42. Boud, A. C., Haniff, D. J., Baber, C., Steiner, S. J.: Virtual reality and augmented reality as a training tool for assembly tasks. In: Proceedings of the international conference on information vision. IEEE, London, 1999, pp. 32–36.
43. Ke, S., Xiang, F., Zhang, Z., Zuo, Y.: A enhanced interaction framework based on VR, AR and MR in digital twin. In: 11th CIRP conference on industrial product-service systems. Elsevier B.V., 2019, pp. 753–758.
44. Yoh, M. S.: The reality of virtual reality. In: Proceedings – 7th international conference on virtual systems and multimedia, VSMM 2001. IEEE, Berkeley, 2001, pp. 666–674.
45. Latta, J.N., Oberg, D.J.: A conceptual virtual reality model. *IEEE Comput. Graph. Appl.***14**(1), 23–29 (1994)
46. Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B.: Characterising the digital twin: a systematic literature review. *CIRP J. Manuf. Sci. Technol.***29**, 36–52 (2020)
47. Tao, F., Zhang, M.: Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. *IEEE Access*. **5**, 20418–20427 (2017)
48. Roldán, J.J., Crespo, E., Martín-Barrio, A., Peña-Tapia, E., Barrientos, A.: A training system for Industry 4.0 operators in complex assemblies based on virtual reality and process mining. *Robot. Comput. Integr. Manuf.***59**, 305–316 (2019)
49. Havard, V., Jeanne, B., Lacomblez, M., Baudry, D.: Digital twin and virtual reality: a co-simulation environment for design and assessment of industrial workstations. *Prod. Manuf. Res.***7**(1), 472–489 (2019)
50. West, T. D. Blackburn, M.: Is digital thread/digital twin affordable? a systemic assessment of the cost of DoD's latest Manhattan Project. Procedia Computer Sciences. (Elsevier B.V., Chicago, 2017), pp. 47–56.
51. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F.: Digital twin-driven product design, manufacturing and service with big data. *Int. J. Adv. Manuf. Technol.***94**(9–12), 3563–3576 (2018)
52. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. *Multimed. Tools Appl.***51**(1), 341–377 (2011)
53. Wang, P., Zhang, S., Billinghurst, M., Bai, X., He, W., Wang, S., Sun, M., Zhang, X.: A comprehensive survey of AR/MR-based co-design in manufacturing. *Eng. Comput.***36**(4), 1715–1738 (2020)
54. Tao, F., Zhang, M., Nee, A. Y. C. Y. C.: Digital twin and virtual reality and augmented reality/mixed reality. In: Digital Twin Driven Smart Manufacturing. Elsevier, 2019.
55. Rabah, S., Assila, A., Khouri, E., Maier, F., Ababsa, F., Bourny, V., Maier, P., Mérienne, F.: Towards improving the future of manufacturing through digital twin and augmented reality technologies. *Procedia Manuf.***17**, 460–467 (2018)
56. Rosen, R., Von Wichert, G., Lo, G., Bettenhausen, K.D., Von Wichert, G., Lo, G., Bettenhausen, K.D., Von Wichert, G., Lo, G., Bettenhausen, K.D.: About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine*. **28**(3), 567–572 (2015)
57. Schroeder, G., Steinmetz, C., Pereira, C. E., Muller, I., Garcia, N., Espindola, D., Rodrigues, R.: Visualising the digital twin using web services and augmented reality. In: 2016 IEEE 14th international conference on industrial informatics. IEEE, Poitiers, 2016, pp. 522–527.
58. Shao, G., Helu, M.: Framework for a digital twin in manufacturing: Scope and requirements. *Manuf. Lett.***24**, 105–107 (2020)

59. Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W.: Digital Twin in manufacturing: a categorical literature review and classification. *IFAC-PapersOnLine*. **51**(11), 1016–1022 (2018)
60. Kitchenham, B. Charters, S.: Guidelines for performing Systematic Literature Reviews in SE (2007).
61. Pollock, A., Berge, E.: How to do a systematic review. *Int. J. Stroke*. **13**(2), 138–156 (2018)
62. Macchi, M., Roda, I., Negri, E., Fumagalli, L.: Exploring the role of Digital Twin for asset lifecycle management. *IFAC-PapersOnLine*. **51**(11), 790–795 (2018)
63. Shangguan, D., Chen, L., Ding, J.: A hierarchical digital twin model framework for dynamic cyber-physical system design. In: ICMRE'19: the 5th international conference on mechatronics and robotics engineering. Association for Computing Machinery, Rome, 2019), pp. 123–129.
64. Liu, Z., Meyendorf, N., Mrad, N.: The role of data fusion in predictive maintenance using digital twin. In: AIP Conference Proceedings. American Institute of Physics Inc., 2018, pp. 1–6.
65. Erikstad, S. O.: Merging Physics, Big Data Analytics and Simulation for the Next-Generation Digital Twins. In: HIPER 2017, High-Performance Marine Vehicles. Zevenwacht, 2017, pp. 139–149.
66. He, X., Ai, Q., Qiu, R. C., Zhang, D.: Preliminary Exploration on Digital Twin for Power Systems: Challenges, Framework, and Applications. (2019).
67. Lu, Q., Xie, X., Heaton, J., Parlikad, A.K., Schooling, J.: From BIM towards digital twin: Strategy and future development for smart asset management. In: Borangiu, T., Trentesaux, D., Leitão, P., Boggino, A.G., Botti, V. (eds.) *SOHOMA 2019 Serv. Oriented, Holonic Multi-agent Manuf. Syst. Ind. Futur*, pp. 392–404. Springer (2020)
68. Lee, J., Azamfar, M., Singh, J., Siahpour, S.: Integration of digital twin and deep learning in cyber-physical systems: towards smart manufacturing. *IET Collab. Intell. Manuf.***2**(1), 34–36 (2020)
69. Aivaliotis, P., Georgoulias, K., Alexopoulos, K.: Using digital twin for maintenance applications in manufacturing: State of the Art and Gap analysis. In: 2019 IEEE International Conference on Engineering, Technology and Innovation. IEEE, Valbonne Sophia-Antipolis, 2019.
70. Rodič, B.: Industry 4.0 and the new simulation modelling paradigm. *Organizacija*. **50**(3), 193–207 (2017)
71. Liu, Q., Zhang, H., Leng, J., Chen, X.: Digital twin-driven rapid individualised designing of automated flow-shop manufacturing system. *Int. J. Prod. Res.***57**(12), 3903–3919 (2019)
72. Alam, K.M., El Saddik, A.: C2PS: a digital twin architecture reference model for the cloud-based cyber-physical systems. *IEEE Access*. **5**, 2050–2062 (2017)
73. Guo, J., Zhao, N., Sun, L., Zhang, S.: Modular based flexible digital twin for factory design. *J. Ambient Intell. Humaniz. Comput.***10**(3), 1189–1200 (2019)
74. Grieves, M.: *Origins of the Digital Twin Concept* (2016).
75. Biesinger, F., Weyrich, M.: The Facets of Digital Twins in Production and the Automotive Industry. In: 2019 23rd International Conference on Mechatronics Technology. IEEE, Salerno, 2019.
76. Rasheed, A., San, O., Kvamsdal, T.: Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*. **8**, 21980–22012 (2020)
77. Autiosalo, J., Vepsalainen, J., Viitala, R., Tammi, K.: A feature-based framework for structuring industrial digital twins. *IEEE Access*. **8**, 1193–1208 (2020)
78. Qi, Q., Tao, F.: Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *IEEE Access*. **6**, 3585–3593 (2018)
79. Rosen, R., Boschert, S., Sohr, A.: Next generation digital twin. *atp Mag.***60**(10), 86–96 (2018)
80. Tao, F., Zhang, H., Liu, A., Nee, A.Y.C.: Digital twin in industry: state-of-the-art. *IEEE Trans. Ind. Informatics*. **15**(4), 2405–2415 (2019)
81. Lu, Y., Liu, C., Wang, K.I.K., Huang, H., Xu, X.: Digital Twin-driven smart manufacturing: connotation, reference model, applications and research issues. *Robot. Comput. Integr. Manuf.***61**, 101837 (2020)
82. M. Grieves & J. Vickers: Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Springer International Publishing, 2016, pp. 85–113.
83. Fuller, A., Fan, Z., Day, C., Barlow, C.: Digital twin: enabling technologies, challenges and open research. *IEEE Access*. **8**, 108952–108971 (2020)
84. Tao, F., Qi, Q., Wang, L., Nee, A.Y.C.: Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: correlation and comparison. *Engineering*. **5**(4), 653–661 (2019)
85. Martinez, V., Ouyang, A., Neely, A., Burstall, C., Bisessar, D.: Service business model innovation : the digital twin technology. *EurOMA*. (2018)
86. Malek, Y. N., Kharbouch, A., Khoukhi, H. El, Bakhouya, M., De Florio, V., El Ouadghiri, D., Latre, S., Blondia, C.: On the use of IoT and Big Data Technologies for Real-time Monitoring and Data Processing. In: 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017). Elsevier B.V., Procedia Computer Science, 2017, pp. 429–434.
87. Qi, Q., Tao, F., Zuo, Y., Zhao, D.: Digital Twin Service towards Smart Manufacturing. In: 51st CIRP conference on manufacturing systems. Elsevier B.V., Procedia CIRP, 2018, pp. 237–242.
88. Wache, H. Dinter, B.: The Digital Twin – Birth of an Integrated System in the Digital Age. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. International Conference on System Sciences, Maui, Hawaii, 2020.
89. Madni, A., Madni, C., Lucero, S.: Leveraging digital twin technology in model-based systems engineering. *Systems*. **7**(1), 1–13 (2019)
90. Modoni, G. E., Caldarola, E. G., Sacco, M., Terkaj, W.: Synchronizing physical and digital factory: Benefits and technical challenges. In: Procedia CIRP. Elsevier B.V., Naples, 2019, pp. 472–477.
91. Bolton, R.N., McColl-Kennedy, J.R., Cheung, L., Gallan, A., Orsingher, C., Witell, L., Zaki, M.: Customer experience challenges: bringing together digital, physical and social realms. *J. Serv. Manag.***29**(5), 776–808 (2018)
92. Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., Wang, L., Nee, A.Y.C.: Enabling technologies and tools for digital twin. *J. Manuf. Syst.* (2019)
93. Melesse, T. Y., Di Pasquale, V., Riemma, S.: Digital twin models in industrial operations: a systematic literature review. In: Procedia Manufacturing. Elsevier B.V., 2020, pp. 267–272.
94. He, B., Bai, K.J.: Digital twin-based sustainable intelligent manufacturing: a review. *Adv. Manuf.***1–21** (2020)
95. Xu, J., Huang, E., Hsieh, L., Lee, L.H., Jia, Q.S., Chen, C.H.: Simulation optimization in the era of Industrial 4.0 and the Industrial Internet. *J. Simul.***10**(4), 310–320 (2016)
96. Uhlemann, T. H. J., Lehmann, C., Steinhilper, R.: The digital twin: realizing the cyber-physical production system for industry 4.0. In: The 24th CIRP conference on life cycle engineering. Elsevier B.V., 2017, pp. 335–340.
97. Liu, J., Zhou, H., Tian, G., Liu, X., Jing, X.: Digital twin-based process reuse and evaluation approach for smart process planning. *Int. J. Adv. Manuf. Technol.***100**(5–8), 1619–1634 (2019)
98. Uhlemann, T.H.J., Schock, C., Lehmann, C., Freiberger, S., Steinhilper, R.: The digital twin: demonstrating the potential of real

time data acquisition in production systems. *Procedia Manuf.***9**, 113–120 (2017)
99. Zhuang, C., Liu, J., Xiong, H.: Digital twin-based smart production management and control framework for the complex product assembly shop-floor. *Int. J. Adv. Manuf. Technol.***96**(1–4), 1149–1163 (2018)
100. He, R., Chen, G., Dong, C., Sun, S., Shen, X.: Data-driven digital twin technology for optimized control in process systems. *ISA Trans.***95**, 221–234 (2019)
101. Min, Q., Lu, Y., Liu, Z., Su, C., Wang, B.: Machine Learning based Digital Twin Framework for Production Optimization in Petrochemical Industry. *Int. J. Inf. Manage.***49**, 502–519 (2019)
102. Vachalek, J., Bartalsky, L., Rovny, O., Sismisova, D., Morhac, M., Loksik, M.: The digital twin of an industrial production line within the industry 4.0 concept. In: 2017 21st international conference on process control. IEEE, Strbske Pleso, 2017, pp. 258–262.
103. Zhang, H., Liu, Q., Chen, X., Zhang, D., Leng, J.: A digital twin-based approach for designing and multi-objective optimization of hollow glass production line. *IEEE Access*. **5**, 26901–26911 (2017)
104. Bao, J., Guo, D., Li, J., Zhang, J.: The modelling and operations for the digital twin in the context of manufacturing. *Enterp. Inf. Syst.***13**(4), 534–556 (2019)
105. Wang, Y., Wang, S., Yang, B., Zhu, L., Liu, F.: Big data driven hierarchical digital twin predictive remanufacturing paradigm: architecture, control mechanism, application scenario and benefits. *J. Clean. Prod.***248**, 119299 (2020)
106. Kostenko, D., Kudryashov, N., Maystrishin, M., Onufriev, V., Potekhin, V., Vasiliev, A.: Digital twin applications: diagnostics, optimisation and prediction. In: *Annual DAAAM Proceedings International DAAAM Symposim*, 574–0581 (2018).
107. Papacharalampopoulos, A., Giannoulis, C., Stavropoulos, P., Mourtzis, D.: A digital twin for automated root-cause search of production alarms based on KPIs aggregated from IoT. *Appl. Sci.***10**(7), 2377 (2020)
108. Ding, K., Chan, F.T.S., Zhang, X., Zhou, G., Zhang, F.: Defining a digital twin-based cyber-physical production system for autonomous manufacturing in smart shop floors. *Int. J. Prod. Res.***57**(20), 6315–6334 (2019)
109. Leng, J., Zhang, H., Yan, D., Liu, Q., Chen, X., Zhang, D.: Digital twin-driven manufacturing cyber-physical system for parallel controlling of smart workshop. *J. Ambient Intell. Humaniz. Comput.***10**(3), 1155–1166 (2019)
110. Stojanovic, N., Milenovic, D.: Data-driven Digital Twin approach for process optimization: an industry use case. In: 2018 IEEE international conference on Big Data (Big Data). IEEE, Seattle 2019, pp. 4202–4211.
111. Zhao, P., Liu, J., Jing, X., Tang, M., Sheng, S., Zhou, H., Liu, X.: The modeling and using strategy for the digital twin in process planning. *IEEE Access*. **8**, 41229–41245 (2020)
112. Fang, Y., Peng, C., Lou, P., Zhou, Z., Hu, J., Yan, J.: Digital-twin-based job shop scheduling toward smart manufacturing. *IEEE Trans. Ind. Informatics*. **15**(12), 6425–6435 (2019)
113. Negri, E., Fumagalli, L., Cimino, C., MacChi, M.: FMU-supported simulation for CPS digital twin. In: 7th International conference on Changeable, Agile, Reconfigurable and Virtual Production. (CARV2018), Procedia Manufacturing. Elsevier B.V., 2019, pp. 201–206.
114. Lu, Y., Xu, X.: Resource virtualization: a core technology for developing cyber-physical production systems. *J. Manuf. Syst.***47**, 128–140 (2018)
115. Zhu, Z., Liu, C., Xu, X.: Visualisation of the digital twin data in manufacturing by using augmented reality. *Procedia CIRP*. **81**, 898–903 (2019)
116. Liu, C., Jiang, P., Jiang, W.: Web-based digital twin modeling and remote control of cyber-physical production systems. *Robot. Comput. Integr. Manuf.***64**, 101956 (2020)
117. Zhou, G., Zhang, C., Li, Z., Ding, K., Wang, C.: Knowledge-driven digital twin manufacturing cell towards intelligent manufacturing. *Int. J. Prod. Res.***58**(4), 1034–1051 (2020)
118. Nikolakis, N., Alexopoulos, K., Xanthakis, E., Chryssolouris, G.: The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor. *Int. J. Comput. Integr. Manuf.***32**(1), 1–12 (2019)
119. Miller, A.M.D., Alvarez, R., Hartman, N.: Towards an extended model-based definition for the digital twin. *Comput. Aided. Des. Appl.***15**(6), 880–891 (2018)
120. Zheng, Y., Yang, S., Cheng, H.: An application framework of digital twin and its case study. *J. Ambient Intell. Humaniz. Comput.***10**(3), 1141–1153 (2019)
121. Zhang, H., Ma, L., Sun, J., Lin, H., Thürer, M.: Digital twin in services and industrial product service systems: Review and analysis. In: 11th CIRP Conference on Industrial Product-Service Systems Procedia CIRP. Elsevier B.V., 2019, pp. 57–60.
122. Söderberg, R., Wärmefjord, K., Carlson, J.S., Lindkvist, L.: Toward a Digital Twin for real-time geometry assurance in individualized production. *CIRP Ann. – Manuf. Technol.***66**(1), 137–140 (2017)
123. Kritzler, M., Funk, M., Michahelles, F., Rohde, W.: The virtual twin: controlling smart factories using a spatially-correct augmented reality representation. *ACM Int. Conf. Proceeding Ser.* (2017)
124. Henderson, S. J., Feiner, S.: Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: Science & Technology Proceedings,– IEEE 2009 International Symposium on Mixed and Augmented Reality. ISMAR 2009, 135–144 (2009).
125. Revetria, R., Tonelli, F., Damiani, L., Demartini, M., Bisio, F., Peruzzo, N.: A real-time mechanical structures monitoring system based on digital Twin, IOT and augmented reality. *Simul. Ser.***51**(1) (2019)
126. Cai, Y., Wang, Y., Burnett, M.: Using augmented reality to build digital twin for reconfigurable additive manufacturing system. *J. Manuf. Syst.***56**(April), 598–604 (2020)
127. Aheleroff, S., Xu, X., Zhong, R.Y., Lu, Y.: Digital Twin as a Service (DTaaS) in Industry 4.0: An Architecture Reference Model. *Adv. Eng. Informatics*. **47**(October 2020), 101225 (2021)
128. Qiu, C., Zhou, S., Liu, Z., Gao, Q., Tan, J.: Digital assembly technology based on augmented reality and digital twins: a review. *Virtual Real. Intell. Hardw.***1**(6), 597–610 (2019)
129. Liu, S., Lu, S., Li, J., Sun, X., Lu, Y., Bao, J.: Machining process-oriented monitoring method based on digital twin via augmented reality. *Int. J. Adv. Manuf. Technol.* (2021)
130. Rabah, S., Assila, A., Khouri, E., Maier, F., Ababsa, F., Bourny, V., Maier, P., Mérienne, F.: Towards improving the future of manufacturing through digital twin and augmented reality technologies. *Procedia Manuf.***17**, 460–467 (2018)
131. Mandolla, C., Petruzzelli, A.M., Percoco, G., Urbinati, A.: Building a digital twin for additive manufacturing through the exploitation of blockchain: a case analysis of the aircraft industry. *Comput. Ind.***109**, 134–152 (2019)
132. Roda, I., Macchi, M., Albanese, S.: Building a Total Cost of Ownership model to support manufacturing asset lifecycle management. *Prod. Plan. Control*. **31**(1), 19–37 (2020)
133. Yusupbekov, N., Abdurasulov, F., Adilov, F., Ivanyan, A.: Concepts and methods of "Digital Twins" models creation in industrial asset performance management systems. In: Kahraman, C., Cevik Onar, S., Oztaysi, B., Sari, I., Cebi, S. (eds.) International Conference on Intelligent and Fuzzy Systems INFUS 2020, pp. 1589–1595. Springer, Cham (2021)
134. Roda, I., Parlikad, A.K., Macchi, M., Garetti, M.: A framework for implementing value-based approach in asset management. In: Koskinen, K. (ed.) Proceedings of the 10th World Congress on

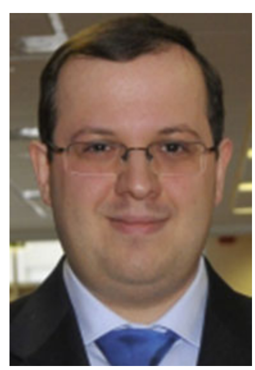
Engineering Asset Management. *WCEAM 2015*, pp. 487–495. Springer, Charm (2016)
135. Park, K.T., Nam, Y.W., Lee, H.S., Im, S.J., Do Noh, S., Son, J.Y., Kim, H.: Design and implementation of a digital twin application for a connected micro smart factory. *Int. J. Comput. Integr. Manuf.* **32**(6), 596–614 (2019)
136. Durão, L. F. C. S., Haag, S., Anderl, R., Schützer, K., Zancul, E.: Digital twin requirements in the context of industry 4.0. In: IFIP Advances in Information and Communication Technology. Springer New York LLC, 2018, pp. 204–214.
137. Schleich, B., Anwer, N., Mathieu, L., Wartzack, S.: Shaping the digital twin for design and production engineering. *CIRP Ann. – Manuf. Technol.* **66**(1), 141–144 (2017)
138. Farooq, M.U., Waseem, M., Mazhar, S., Khairi, A., Kamal, T.: A review on Internet of Things (IoT). *Int. J. Comput. Appl.* **113**(1), 1–7 (2015)
139. Lee, J., Azamfar, M., Singh, J.: A blockchain enabled cyber-physical system architecture for industry 4.0 manufacturing systems. *Manuf. Lett.* **20**, 34–39 (2019)
140. Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S.C.Y., Nee, A.Y.C.: Digital twin-driven product design framework. *Int. J. Prod. Res.* **57**(12), 3935–3953 (2019)
141. Rocca, R., Tavola, G., Boschi, F., Fantini, P., Taisch, M., Lambruschini, V. R.: Business Case Evaluation Methodology (BCEM) for Factories Digitalization. In: The Thirteenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies. UBICOMM 2019 Politecnico di Milano, Porto, 2019, pp. 134–141.
142. Brady, T., Rush, H., Hobday, M., Davies, A., Probert, D., Banerjee, S.: Tools for technology management: an academic perspective. *Technovation*. **17**(8), 417–426 (1997)
143. Bozarth, C., Edwards, S.: The impact of market requirements focus and manufacturing characteristics focus on plant performance. *J. Oper. Manag.* **15**(3), 161–180 (1997)
144. Skinner, W.: The focused factory. *Harv. Bus. Rev.*, 114–121 (1974). https://hbr.org/1974/05/the-focused-factory Accessed 10 Feb 2021.
145. Industry 4.0 Lab. Politecnico di Milano School of Management, (n.d.). https://www.industry40lab.org/ Accessed 10 Feb 2021).
146. Smith, J.S.: Survey on the use of simulation for manufacturing system design and operation. *J. Manuf. Syst.* **22**(2), 157–171 (2003)
147. Festo: Festo Didactics. CIROS software 2018. (2018).
148. Colombo, A.W., Bangemann, T., Karnouskos, S., Delsing, J., Stluka, P., Harrison, R., Jammes, F., Lastra, J.L.M.: *Industrial Cloud-Based Cyber-Physical Systems*. Springer, Cham (2014)
149. Mintzberg, H.: The strategy concept I: five Ps for strategy. *Calif. Manage. Rev.* **30**(1), 11–24 (1987)



Roberto Rocca is PhD candidate in the Manufacturing Group at the Department of Management, Economics and Industrial Engineering of the Politecnico di Milano, where he has been collaborating since 2016. His research interests are focused on Sustainable Manufacturing paradigm, Sustainability Metrics and Assessment, Circular Economy and I4.0 technologies supporting sustainable operations in manufacturing field.

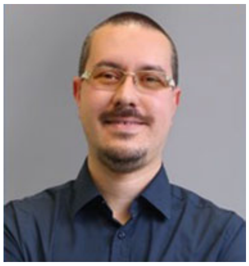
Roman Felipe Bastidas Santacruz is a Mechanical and Electronical Engineer from Univeridad de los Andes, Colombia. He received his MSc. In management engineering from Politecnico di Milano at Milan, Italy, in 2020. He is currently a researcher at Politecnico di Milano;s he works in European projects and his focus area is centred on industrial digital transformation in the I4.0 context.

Claudio Sassanelli received his PhD from Politecnico di Milano in 2017 (holding visiting positions at TMU and USP). His main research interest is PSS design, addressing to PLM, DfX, and Circular Economy and Industry 4.0. He is involved in research, industrial and European projects, carries out teaching activities on industrial plants, and disseminates his research in international journals and conferences.



Paolo Rosa is a Post-Doc researcher at the Department of Management, Economics, and Industrial Engineering of Politecnico di Milano. His research focuses on Product Lifecycle Management and Circular Economy. He is involved in several European research projects, plus teaching activities. Co-author of more than 40 journal and conference papers, he was awarded in 2015 with the Elsevier Atlas Award.

34



Luca Fumagalli is a Professor at Department of Management, Economics and Industrial Engineering of Politecnico di Milano. His research interests are currently focused on innovations in maintenance management and innovations in operations at manufacturing shop floor level, within the scope of Smart Manufacturing and Industry 4.0 framework. He is presently director of the Industry 4.0 Lab of Politecnico di Milano.

Elisa Negri is Assistant Professor at the Department of Management, Economics and Industrial Engineering of Politecnico di Milano. Her research focuses on smart manufacturing, from data modelling for manufacturing systems, to advanced simulation for production planning and control. She is involved in several research and industrial projects. She is Professor of the "Industrial Plant Management" Course at Politecnico di Milano.

35 Digital Twins as Foundation for Augmented Reality Applications in Aerospace

Georgia Albuquerque, Philipp M. Fischer, Syed Muhammad Azeem, Andreas-Christoph Bernstein, Sebastian Utzig, and Andreas Gerndt

Contents

35.1	**AR-Based System Modeling in Aerospace**	881
35.2	**Background**	882
35.2.1	Digital Twin	882
35.2.2	Digital Twins in Aerospace	883
35.2.3	Model-Based System Engineering	884
35.3	**AR Digital Twins in Aerospace Applications**	886
35.4	**Challenges in AR Digital Twin Implementations**	887
35.4.1	Overlay Precision	888
35.4.2	Interaction	889
35.5	**Dedicated AR Applications**	891
35.5.1	From Diagrams to AR in Spacecraft Design	891
35.5.2	On-Orbit Servicing	893
35.5.3	Remote Collaboration and Maintenance	896
35.6	**Closing Remarks**	897
References		898

Abstract

Augmented reality techniques can be used to support system modeling and industrial operations at different levels, enabling designers and engineers to augment their real environment with relevant virtual content. In aerospace, these techniques are tightly coupled to digital twins. Together, they can enhance scarce operational resources, facilitating skill transfer and knowledge retention. It is possible to define the term digital twin based on conceptual data models as used in model-based system engineering. In this definition, a conceptual data model is used to accompany a product as an unique evolving system model through the whole lifecycle, starting from virtual abstractions and progressing toward a virtual replication of the real-world entity. Throughout the whole lifecycle, digital twins help to analyze or predict system behavior for improving decision-making and avoiding cost-expensive prototyping. In this chapter, we discuss how digital twin representations can leverage on augmented reality approaches and provide an overview of how model-based system engineering can help to maintain information consistency through the different phases of the product lifecycle. In this context, we address different aspects related to the use of augmented reality approaches for digital twins in aerospace, such as overlay precision, interaction, data visualization, and remote collaboration. Our example applications take different phases of the product lifecycle into account, from creation to operation.

Keywords

Aerospace applications · Model-based system engineering (MBSE) · Augmented reality (AR) · Digital twin prototype (DTP) · Digital twin instances (DTI) · 3D interaction · Computer-supported collaborative work (CSCW) · Concurrent engineering · On-orbit servicing · System maintenance

Electronic Supplementary Material The online version of this chapter (https://doi.org/10.1007/978-3-030-67822-7_35) contains supplementary material, which is available to authorized users.

G. Albuquerque (✉) · P. M. Fischer · S. M. Azeem · A.-C. Bernstein · S. Utzig · A. Gerndt
Institute for Software Technology, DLR Braunschweig, Braunschweig, Germany
e-mail: georgia.albuquerque@dlr.de; Philipp.Fischer@dlr.de; Syed.Azeem@dlr.de; Andreas-Christoph.Bernstein@dlr.de; sebastian.utzig@dlr.de; andreas.gerndt@dlr.de

35.1 AR-Based System Modeling in Aerospace

Designing and operating space and aircraft are complex processes that involve huge amount of data. During design, a great deal of information needs to be consistently shared

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7_35

across the design team. In operation, recorded sensor data of an aircraft needs to be analyzed for maintenance tasks. Databases are filled with numbers and data series. These data contain important technical details, but they need to be linked to a context to become meaningful. All these requirements can be summarized under the terminology of a digital twin. It covers the basic idea of a one-to-one relation between a real-life system and its virtual counterpart, capturing and helping to process all relevant information.

Similar to digital twins, Model Based Systems Engineering (MBSE) is a paradigm that has been used to improve the whole product lifecycle of systems in recent years. Starting from the first idea, a spacecraft is accompanied by a model. This model is continuously used to capture, share, and verify system knowledge during design. These models grow over time and are also used during after the design for the whole product lifecycle. Together with operational data of the final products, these models gradually transform into powerful model-based digital twins .

Our research focuses on the development of such model-based digital twins and how they can leverage the use of Augmented Reality (AR). In the last years, AR has demonstrated the exceptional possibilities for improving the productivity of industrial and manufacturing workers. Combining digital twins and AR is a natural step to add contextual value to the stored data. In fact, AR is able to bridge both real and virtual worlds. Using AR, data from a digital twin can be directly used, for example, in maintenance applications. In such a use case, recently acquired sensor data is virtually overlaid on real physical parts which guides operators to manage critical tasks. From a remote location, engineers can be called in on the fly using AR mock-ups to support the operators. The digital twin data is eventually the single point of truth for synchronized collaboration.

Nevertheless, these applications raise new challenges to the use of AR. In this chapter we elaborate the combination between AR and digital twins and investigate different practical aspects related to this connection. These challenges include overlay of different kinds of data on real components, overlay precision for large virtual components, and user interaction aspects.

The remaining of the chapter is described as follows: Sect. 35.2 provides a definition to digital twin. Based on this definition, some applications in aerospace are highlighted. The section continues with the definition of MBSE. It explains its purpose in the spacecraft domain and makes the connection to digital twins. Section 35.3 answers the question why AR is an advantage in such digital twin applications. The associated challenges are covered in detail in Sect. 35.4, where specific challenges on remote collaboration, overlay calibration, or interaction are shown. Current and ongoing research providing answers to these challenges is displayed as well. The chapter is finalized with a closer look in three dedicated applications in Sect. 35.5. The first one covers the integration of AR in early spacecraft MBSE activities. The second one shows an application linked to spacecraft on-orbit servicing and training. The final one focuses on digital twins and AR in a remote collaboration maintenance task for aircraft components.

35.2 Background

The term *Digital Twin* has been extensively used in the last years. Nevertheless, it can be traced back to its origin in 2002. Back then it represented the idea of enhancing Product Lifecycle Management (PLM) and the Manufacturing Execution System (MES) processes with collected data. This idea is very similar to MBSE which intends to improve product design and production as well. This section discusses the related work to the origins of the digital twin. It continues with examples of digital twins in aerospace applications indicating essential advantages of digital twins and demonstrating the mapping to MBSE. MBSE itself is introduced in more detail in the last part of this section. It starts with definitions and visions of MBSE from the International Council on Systems Engineering (INCOSE) and continues explaining how MBSE can be applied to a spacecraft lifecycle management.

35.2.1 Digital Twin

Although already presented in 2002 as a conceptual sketch [1], the term *Digital Twin* was used for the first time by Michael Grieves in [2]. In his paper, the focus was mainly on the production process in factories and how to enhance the product lifecycle management (PLM). Later in 2014 [3], he summarized his concerns about the rise of MES , which on the one hand has in fact led to extensive data collection in all cells of the production but without considering the need to update the underlying digital design model as well. To overcome this gap, he has invented the digital twin which shows a strong relationship between the physical and the digital product. This relationship can be achieved by storing all measured data in the virtual space while information from the virtual space is depicted in the real production space. Thus, as one solution, Grieves proposed the implementation of a Unified Repository (UR) to tightly link real and virtual products. In the design phase, tags are introduced to specify the UR. Later on, the tags are incorporated into the MES and filled with information captured by all kinds of sensors used in the factory. Finally, the UR is used for the factory simulation again. Since the introduction of digital twins, many publications attempted to define digital twins even further for different purposes [4]. However, it is common sense that a digital twin is not just a Computer Aided Design (CAD) model. Rather all modifications and relations in the real world become part of the digital twin as well. Core concept is a bidirectional connectivity between the digital model and the physical counterpart. Actually, the digital twin can be beneficial in all four phases of the product lifecyle [1]:

- **Creation** : The system must be designed respecting all the required specifications and behaviors. A virtual 3D model should support the simulation of the system behavior. This helps to reduce costly production of physical prototypes. The system environment must also be considered in order to assess manufacturability, sustainability, supportability, and disposability.
- **Production**: While a system emerges to a physical artifact, one may find out that the manufacturing techniques are not sufficient or the design has to be changed to fit the reality. Evaluating and existing information and collecting new in-built information enhances understanding of components, processes, and tools.
- **Operation**: While the system is in operation, it has to be monitored and supported, the system behavior has to be assessed and predicted, and sub-systems have to be repaired or replaced. Digital twins support efficient maintenance. Front-run simulations fed with real-time information of the system can present consequences of decisions how to operate the system next. Front-run simulations fed with real-time system information as part of a digital twin can present consequences of decisions how to operate the system next.
- **Disposal**: The collected information of the previous phases has to be achieved as knowledge for future system developments. Additionally, used material of the system has to be disposed properly to protect the environment.

According to the different phases, the digital twin also comprises different tasks. Thus, Grieves and Vickers [1] distinguish between a digital twin in the creation phase, depicted as Digital Twin Prototype (DTP), where one has merely to model the virtual version, and the Digital Twin Instance (DTI) in the following phases, where the digital twin describes and collects corresponding aspects of the physical product. Additionally, they separate the environment from these digital twin models where they have to be executed and evaluated. They call it the Digital Twin Environment (DTE) where collected information of multiple instances may be combined to predict future behavior and states of the system.

A digital twin has to collect information throughout the entire lifecycle. As some kind of template, it needs to be developed with the evolution of the product and its environment. This leads to the requirement that the DTI has to be instantiated with each product instance [5]. This brings up the question how complex a digital twin has to be. Most authors assume that digital twins consist of all digital artifacts arising during development, production, and operation. All properties, conditions, and behaviors of real-life products are mirrored through models and data. However, this could create a high burden to prepare appropriate models. Boschert and Rosen [6] therefore recommend using only the information for the models that is essential for the use of digital twins.

This addresses also one of the main obstacles for digital twins to become efficient [1]: To mirror the physical world, the understanding of physical phenomena has always to be improved, which eventually results in an exploding state space of the system. Another obstacle, however, is a cultural aspect: Organizational siloing between designing, engineering, manufacturing, and support prevents an efficient embedding of digital twin approaches into a product lifecycle management. If the cultural aspect can be solved and a digital twin environment could be set up seamlessly through the entire lifecycle, it supports humans in multiple ways (Grieves, [3]):

- Conceptualization
- Comparison
- Collaboration

For Grieves, the most efficient support for humans in problem solving and decision-making is visualization. Although analyzing data in tables, reports, etc. can help already to conceptualize a situation, visualization can reduce the mental load to treat information complexity. Digital twins allow overlapping design parameters over the physical product or blending in factory information into the real environment. Physical objects are augmented with appropriate context information, which is a core aspect of AR. This overlapping visualization approach supports humans in their effort to compare continuously desired results with the real scenario to identify differences. But a shared common view enables also collaboration: It allows shared conceptualization. For collaboration sessions, the digital twin does not require being in the same physical space but enables discussion also in distributed environments.

35.2.2 Digital Twins in Aerospace

In the space domain, MBSE is currently on its way toward becoming the accepted paradigm for system engineering.

However, digital twin concepts are not being considered yet. Nevertheless, back in 2012, National Aeronautics and Space Administration (NASA)'s Technology Aera 11 (Software, Modeling, Simulation, and Information Processing) specified the utility of digital twins in an early draft of NASA's integrated technology roadmap [7]. Later, this technology roadmap was released as the 2020 NASA Technology Taxonomy [8].

In the aeronautics domain, digital twins are already seen as a dominant approach. In this case, digital transformation plays a much more important role, because production and market activities are organized in a completely different way from the spatial domain. Stronger competition between a large number of companies forces innovation in all fields of a system's lifecycle. An example is the idea outlined by Tuegel et al. [9]. In this work, the authors depicted a vision of how a digital twin might be used to assess the structure of an airplane in operation. According to detected damages, they proposed the use of a digital twin to predict the remaining lifetime of the related airplane instance. In their sketch, an airplane may first undergo a virtual flight before any real flight. Considering flight trajectory, weather conditions, current structure status, and other parameters, a physically based simulation may be performed to predict the aging of the structure. Exceptional situations may be additionally integrated to uncover unexpected system failures. If no risks are detected, the real flight is allowed. During the real flight, the linked digital twin instance records information from the structural sensing system and interpolates this generally sparse sensor data for the entire airframe. This yields a periodically updated material evolution and health monitoring system that can suggest the time for the next maintenance. The basis, however, of the structural life prediction approach is a complex multiphysics modeling, a precise damage model, and a highly coupled high-resolution structural analysis simulation. With the growing knowledge in material science, as well as with the advances in High Performance Computing (HPC), this use case vision is moving closer to a real implementation.

35.2.3 Model-Based System Engineering

System engineering can be considered as a trans-disciplinary domain that focuses on how to design and maintain complex systems along their lifecycle. In the space domain, NASA described it in their *System Engineering Handbook* [10] as an important approach, widely used throughout the lifecycle of a spacecraft. In Europe, the importance of system engineering was manifested in standards such as the *ECSS-E-ST-10C-Rev.1—System engineering general requirements* defined by the European Cooperation for Space Standardization (ECSS) [11].

In 1956, the term system engineering already existed but was not well defined yet [12]. The need for system engineering came from the feeling that satisfactory components could not always be combined as satisfactory complex systems. Following Schlager [12], Bell Telephone Laboratories was one of the first companies that contributed to the definition of system engineering. Back then, Schlager researched how different companies applied system engineering, asking them to define it. As a result, five keywords were identified: planning, analysis, optimization, integration, and evaluation. In his view, these words corresponded to the five stages of the system engineering process.

Nowadays, we can cite the INCOSE as an important authority in system engineering. INCOSE is a nonprofit organization that was created to develop and distribute interdisciplinary principles and practices of system engineering. In 2009, they redefined system engineering as *"a transdisciplinary and integrative approach to enable the successful realization, use, and retirement of engineered systems, using systems principles and concepts, and scientific, technological, and management methods"* [13], as part of an internal review to adapt the older definitions to current practice. However, such integrated, transdisciplinary approaches require extensive information and data exchange. This exchange is traditionally made through different kinds of documents, such as text files or spreadsheets. Keeping this information consistent during the system lifecycle is not a trivial task and remains as one of the main goals of MBSE.

In 2007, INCOSE published the System Engineering Vision 2020 [14], where they reported the state of the art in systems engineering and discussed a vision for the year 2020. In their work, they defined MBSE as *"the formalized application of modeling to support system requirements, design, analysis, verification and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases"*. According to INCOSE, MBSE can be seen as a long-term trend in system engineering activities toward the replacement of document-centric approaches by model-centric ones. They argued that the lack of interoperability of tools based on this central model is one of the biggest challenges for a widespread use of MBSE. From INCOSE's perspective, system engineering was going to be model-based in the future, enabling a seamless flow of product information at all stages of the system lifecycle. This is coupled with a reduced need for physical prototypes. Additionally, INCOSE claimed in this vision that multidimensional visualization capabilities should support the system engineer in what-if analysis.

Following the Vision 2020, INCOSE published the *Systems Engineering Vision 2025* [15] in 2014. It stated that future environments should incorporate modeling, simulation, and visualization. These elements should assist all aspects of system engineering with improved prediction and analysis of complex emergent behavior. Future system engineering tools should also support immersive technologies for visualization.

Following this newer vision, these technologies will also integrate with CAD/Computer Aided Engineering (CAE)/PLM environments into an enterprise management environment.

Grieves compared classical system engineering models such as *Waterfall*, *Spiral*, or *Vee* with the digital twin [1]. He argued that classical sequential approaches were needed in the past whenever complex prototypes had to be developed. With these approaches, lately detected design failures resulted almost always in expensive redesigns.

In contrast, the digital twin in use today is running simultaneously with each design step. This makes required corrections of the design more feasible. Modifications anywhere in the model can immediately be propagated to the current design phase. Also design variations are supported. Grieves definition [1, 3] and INCOSE's visions have general similarities. Although MBSE mainly focuses on *Creation* and *Production* phases, both MBSE and the digital twin share common properties. Among others, both are supposed to be beneficial along the lifecycle phases and should improve prediction of emergent behavior.

The lack of interoperability pointed out by the *Systems Engineering Vision 2025* [15] is nowadays addressed by several related standards. One of these is the System Modeling Language (SysML) [16] that provides a general purpose modeling language for system engineering. It was published as an international standard *ISO/IEC 19514:2017*. SysML is based on Unified Modeling Language (UML) [17], being a so-called profile of UML. UML is intended to support software developers and engineers with tools for analysis, design, and modeling. It provides syntax, semantics, and specification for notation with the goal to increase interoperability. The *OMG XMI 2 model interchange standard* [16, 18], for example, focuses on general model exchange across UML specific tools. A higher level exchange on system engineering information is defined in *ISO 10303 STEP AP233 data interchange standard*. Besides these standards, several different tools, frameworks, and methodologies implementing MBSE [19] exist.

MBSE has also been applied in Europe, e.g., at Airbus Defence and Space (Airbus DS) [20]. In this work, the authors developed and introduced the concept of Model-based Development and Verification Environment (MDVE). Their goals were the reduction of cost and managing system complexity. MDVE manages development processes through several lifecycle phases on the basis of consistent models and databases.

The term database is often used in conjunction with MBSE in space applications [21]. As stated by Fischer et al., a database usually consists of three main elements: a Conceptual Data Model (CDM), a system model as an instance of the CDM, as well as procedures to analyze the information. Examples of how such models and databases can support decision-making were presented by Deshmukh et al. [22]. Focusing on early design phases, they presented on-the-fly analysis of the model data. Their analysis provided instant feedback to engineers by, e.g., summarizing the spacecraft masses.

In space missions, a typical lifecycle of a spacecraft is divided in seven different phases [23]. These phases are named 0/A, B, C, D, E, and F, ranging from design over assembly to operations and the final disposal, as can be seen in Fig. 35.1. The first phase, also called 0/A, covers the early design being often called preliminary design. Phases B and C involve a preliminary and detailed definition of the system, respectively. Phase D involves the qualification and production of the spacecraft. In phase E, the system is actually in operation. And finally, phase F involves the disposal process. Each phase is associated with reviews that need to be passed to enter into the next sequential phase: MDR, PRR, SRR, PDR, CDR, AR, ORR, FRR, and ELR. Currently, different databases for the different phases of the lifecycle exist. Some of them are:

- Virtual Satellite [24] developed by German Aerospace Center (DLR)
- Open Concurrent Design Tool (OCDT) [25] developed by European Space Agency (ESA)
- Virtual Spacecraft Design (VSD) [26] developed by ESA
- RangeDB [27] developed by Airbus DS

Regardless of addressing different phases in the lifecycle of a spacecraft, these databases have in common that they are used to provide and manage a system model. OCDT addresses the early design phase, VSD and RangeDB address later phases focusing on the Assembly, Integration and Test (AIT) activities of a spacecraft. Virtual Satellite has been developed to implement MBSE over the whole lifecycle of a spacecraft [24]. In the MBSE context, it is important that

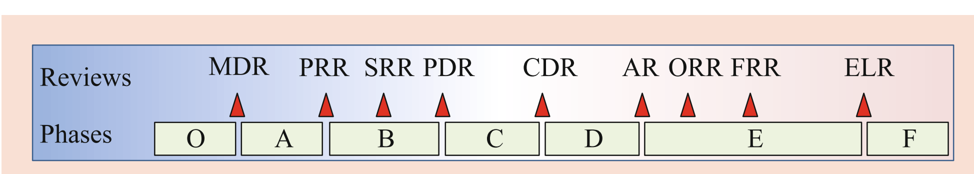


Fig. 35.1 Typical lifecycle of a spacecraft according to the ECSS. Phases 0, A, B, C, D, E, and F and the reviews that need to be passed to enter into the next sequential phase: MDR, PRR, SRR, PDR, CDR, AR, ORR. FRR, and ELR. (© DLR 2020, CC-BY 3.0)

such a database reflects a single point of truth in the projects, as stated by Fischer et al. [24]. Starting from preliminary design, the databases should allow information reuse along the whole lifecycle of the spacecraft. Any information that is relevant to more than one engineer should be stored in the database. Interfaces should offer capabilities for information exchange between domain specific models, tools, and processes. These concepts should allow to, e.g., import information from traditional design tools, such as a CAD software [28]. Roventa showed an extended CAD interface for a complete round-trip engineering process with an MBSE database [29]. Using this interface, parameters representing the position and orientation of spacecraft components could be shared using the system model in the database. Roventa argued that this approach allows information to be exchanged with other domains, e.g., in thermal analysis.

Fischer et al. [30] showed an example of how a MBSE database can be used in the phase B of the spacecraft lifecycle to configure simulation environments. These environments were used to validate the design and avoid potential issues that could lead to expensive changes in the future. Here, system models could be synchronized and interfaced with a simulation database, to drive simulator configurations. Cazenave and Fancelli [31] illustrated a wider approach of how RangeDB evolved in Airbus DS strategic digitalization goal. They pointed out the interoperability with other modeling tools and the PLM in particular. Such a connection was addressed by Grieves in [2] as well. This is another point that indicates the strong connection between MBSE and digital twins.

35.3 AR Digital Twins in Aerospace Applications

The digital twin can be understood as a combination of three main parts: conceptual models, a global collection of real data, and their respective simulation and analysis. Figure 35.2 depicts an overview of this model-based digital twin along the lifecycle phases. During the *creation*, the digital twin mainly consists of conceptual models. These models reflect the designed system, for example, within a system model. In the beginning of the *production* phase, a DTI naturally comes into existence for each real spacecraft. Already during AIT campaigns, virtual components are replaced by real ones. These real components are still connected to the system model due to their conceptual representatives and vice versa. For example, calibration data, which has been initially modeled from specifications, is now extended by real measurements. Modeled and measured calibration data can be compared to improve the system. This relates the system model, associated simulations, and processes from an early abstract representation to a later more specific and realistic representation. Through the lifecycle, the amount of information in the system model as well as the amount of real data is growing. Even after *disposal*, the DTI of a spacecraft virtually still exists. It can be reused, for example, to inspect past events or as basis to design new spacecrafts.

Virtual and augmented reality environments afford a natural perception and interaction with digital three-dimensional worlds and its elements. While Virtual Reality (VR) con-

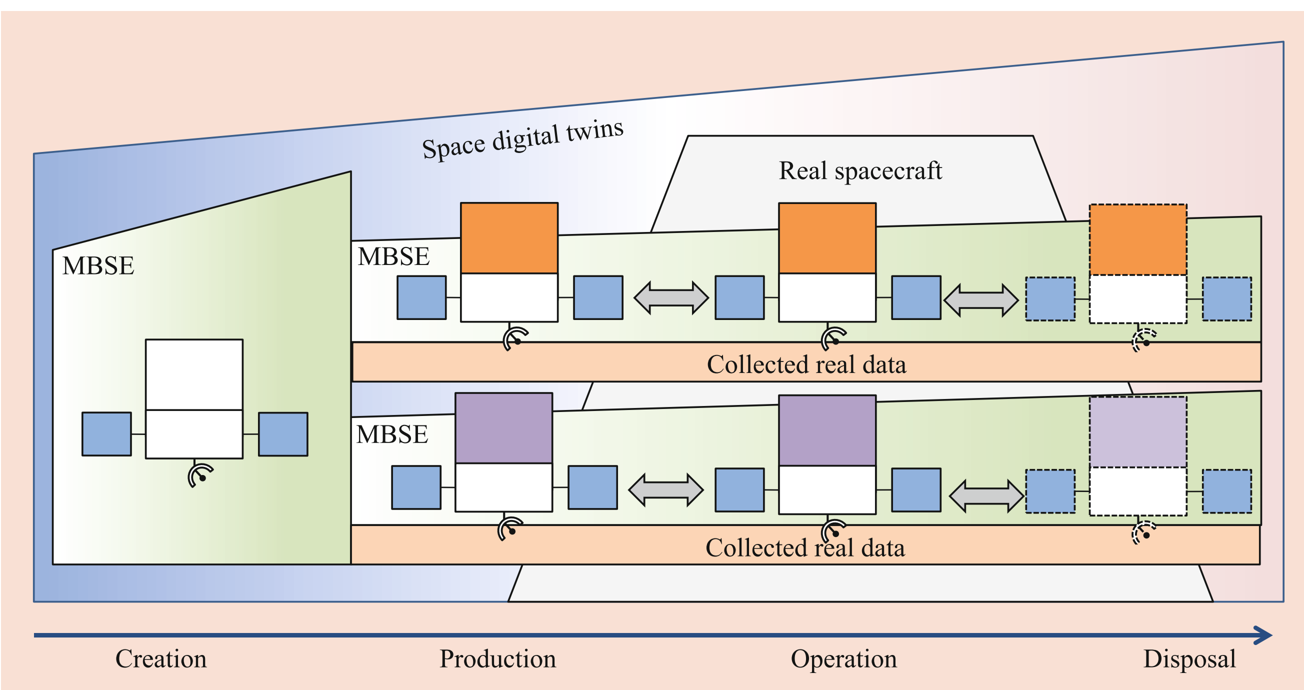


Fig. 35.2 Digital twins in design and operation phases of a spacecraft. (© DLR 2020, CC-BY 3.0)

ventionally comprises environments where the observer is completely immersed in a digital, virtual world, AR environments augment the real world with simulated, virtual elements [32]. Compared to VR, AR can better support the understanding and interaction with the real environments, connecting people, objects, and virtual information with the real world [33, 34].

In the last years, the release of AR goggles as the Microsoft HoloLens opened up a new world of possibilities for AR applications. These headsets can be used to add different kinds of available virtual information to the real environment, for example, to visualize invisible quantities as current temperature or ultrasound sensor data of a component on the fly (Fig. 35.3). AR has a natural connection to digital twins by merging virtual and real content in the right location in real time, making the sheer amount of complex information understandable. It is not only suitable to visualize overlaid information but also to precisely register any information in the DTI. The ability to register virtual information with accuracy related to physical coordinates on the real component while it is being visualized can be of great benefit. AR can support the functionality of a digital twin in the different lifecycle phases, for example:

- **Creation**: Understanding and verification of new ideas can be supported by visualizing design prototypes in AR. Engineers can interact with virtual models to validate the design while it is being created. Design problems can be adjusted and additional digital twin information can be created. For example, if a satellite part has to be reached by a tool used in later phases like production or operation, this reachability requirement can easily be verified by augmenting the already existing physical tool with the virtual prototype to be designed. An example application for the creation phase is discussed in Sect. 35.5.1 where AR can be used to improve the communication between engineers working in the Concurrent Engineering Facility (CEF).
- **Production**: AR can also support assembly tasks by overlaying graphics and information about the whole assembly process in the real environment. The different steps can be highlighted on their turn, and virtual hints to where the parts should be positioned can be shown on demand. With the help of computer-vision algorithms, differences between the *current* and *should be* status of the product can be overlaid for quality assurance. Since the connection between the real system and its digital twin is bidirectional, previously stored information can be corrected or extended while being visualized on the top of the component.
- **Operation**: Depending on requirements and application area, AR can be used in different forms to support operation. In aeronautics, where the components are commonly planned for a long lifetime, virtual user manuals and work instructions guidance can support the maintenance and training activities. An overlay of sensor data on the physical component can give cues about the real damage during maintenance. Data acquired during maintenance tasks can be used not only to document the task itself but also to improve future procedures. In space applications, however, maintenance activities usually involve high costs and safety risks. For these cases, the visualization of overlaid data is better applicable during the spacecraft's earth stay. During spatial operations, remote collaboration using AR mock-ups can be used for guidance and training to minimize safety issues and cost aspects. Exemplary applications for the operation phase in aeronautics and space are presented in Sects. 35.5.3 and 35.5.2, respectively.
- **Disposal**: As mentioned before, in this phase the physical product stops existing and individual DTIs gradually lose their importance. However, the acquired data during the lifecycle can be used for future analysis and projects. In this phase, AR mock-ups can be further used to support analysis and visualization of the gathered data.

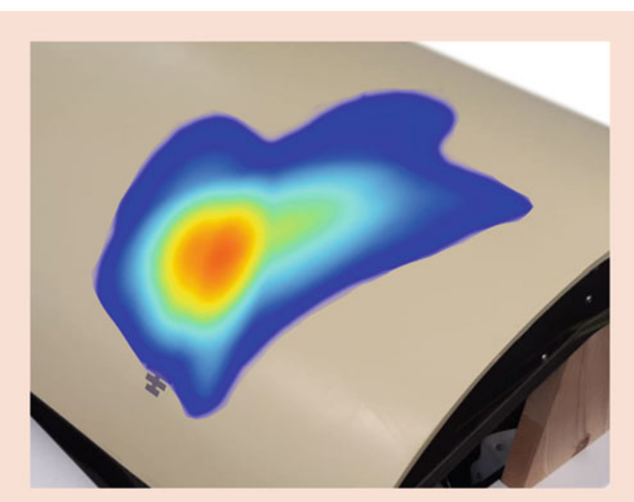


Fig. 35.3 Ultrasonic sensors data depicting a damaged location of an airplane component. (© DLR 2020, CC-BY 3.0)

35.4 Challenges in AR Digital Twin Implementations

AR can be used to improve the applicability of digital twins allowing intuitive visualization and interaction with virtual content. For aerospace applications, however, practical aspects such as overlay precision or user interaction design require special attention. In this section, we describe some challenges and discuss exemplary solutions that we developed during our research on AR and digital twins.

35.4.1 Overlay Precision

In the aerospace industry, we often deal with large aircraft components forming fuselage, wings, and engines. As mentioned before, AR can help to augment the components with useful information. This information can arise from different sources, e.g., CAD models, sensor data, or even manually registered information. However, an inaccurate visualization or registration of this information may lead to misunderstanding and wrong decision-making. The first step to visualize any information in AR is to create a connection between real and virtual worlds, i.e., to determine the exact position of the virtual content in the real environment (Fig. 35.4b).

In this section, we describe challenges faced in precisely aligning digital twin information with large aircraft components in AR. Thereafter, we present approaches to improve the alignment accuracy.

A common approach to align virtual objects with the real world is the use of image markers [35–37]. In this approach, one or more image markers are attached to the real component at predefined locations to support calibration between the real and virtual worlds. However, the use of fixed image markers in the real environment may be restricted, as the aerospace community perceives the need to fix markers on aircraft as a major AR limitation [38,39]. Furthermore, when it is necessary to align large real and virtual components, misalignment problems may occur. Especially component parts far away from the position where the marker was placed are sensitive to such misalignment.

Figure 35.5 shows an AR marker-based calibration between a real component and its virtual counterpart. A marker pose estimation library [35] captures images of a marker with the front camera of the AR headset and calculates the position and orientation of the marker relative to the AR headset. Then this information is used to determine the position and orientation of a virtual component in relation to the marker. In order to perfectly align the virtual and real component, the marker has to be placed on an exact position on the real component. This position is preliminary defined in a tracking setup. However, the perfect placement of the marker with perfect position and orientation is not a straightforward task. Especially for large objects, a small tracking offset in the orientation of the marker can result in a large misalignment on parts distant from the marker position. Similarly, manually placing the marker image precisely at required positions is prone to a positional offset due to human error.

One possibility of mitigating the effects of positional and orientational offsets is to use multiple predefined calibration positions. By pointing to different positions with a tracked tool, the displacement in adjacent regions can be reduced. The following example demonstrates the effect of misalignment by matching a 0.90 m × 1.64 m-sized air brake with its virtual counterpart. To overcome orientational misalignment, the calibration positions should be as distant from each other as possible, covering the extension of the component, e.g., at the corners of the air brake. Using this multi-point approach, Fig. 35.4a shows three positions, A, B, and C, on the air brake which are predefined for calibration. We use the positions to calculate orthogonal vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, forming a coordinate system. This can then be used to determine the orientation of the virtual air brake with higher precision. We

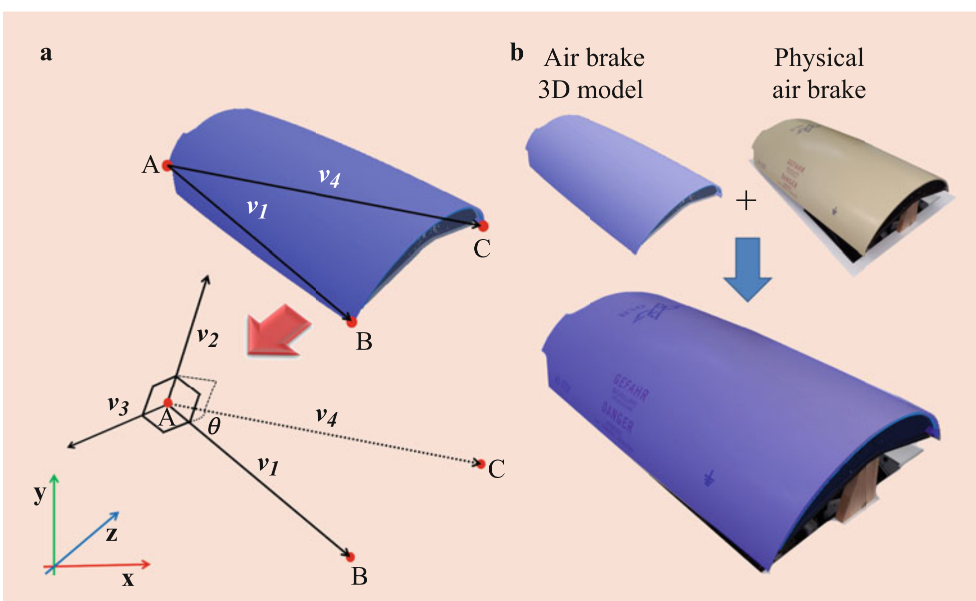


Fig. 35.4 (**a**) Diagram depicting the multi-point approach with calibration points **A**, **B**, and **C** specifying a local coordinate system. (**b**) Result showing the precise overlay of the air brake's 3D model on the top of the real air brake. (© DLR 2020, CC-BY 3.0)

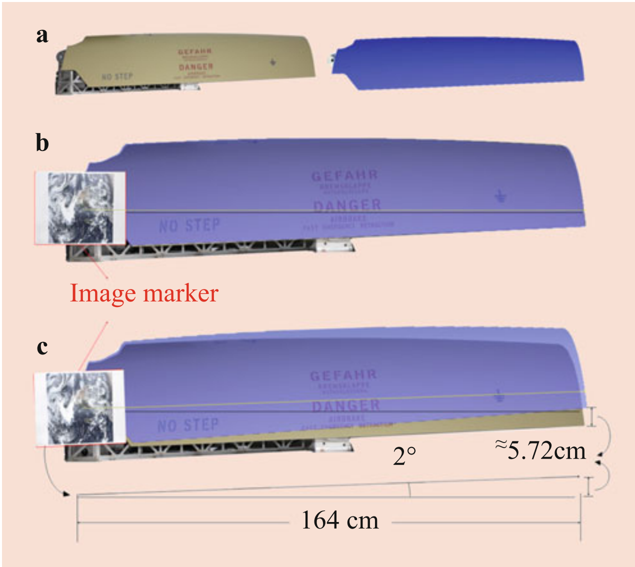


Fig. 35.5 Example showing inaccuracy in geometric alignment due to wrong orientation of the marker. (**a**) Real (left) and virtual (right) aircraft component (air brake). (**b**) Virtual component accurately superimposed on real component. (**c**) Misalignment caused due to an offset of 2 degrees in tracked orientation. (© DLR 2020, CC-BY 3.0)

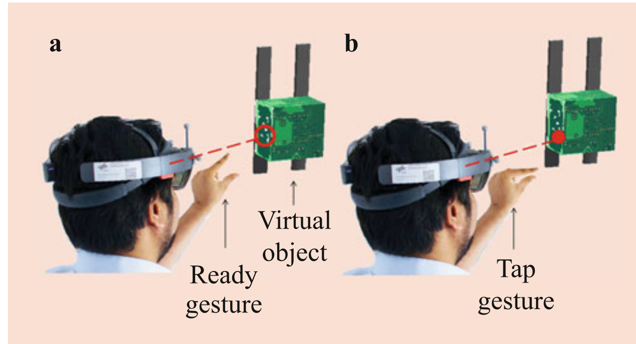


Fig. 35.6 (**a**) A user wearing Microsoft HoloLens gazing at an object with the *Ready* hand gesture. (**b**) User performing the *Tap* gesture to select a virtual object. The red line shows his gaze direction and the circle shows his gaze position. Note that the hollow circle changes into a solid circle as a feedback of the *Tap* gesture. (© DLR 2020, CC-BY 3.0)

can also mitigate positional error by averaging the virtual air brake's positions relative to each calibration point.

This is only one between different approaches that can be used to improve the precision overlay between real and virtual components. However, even when an accurate overlay is achieved, virtual objects are prone to shift over time due to the tracking limitations of headsets [40]. Therefore, in AR experiences that last several minutes, recalibration is often recommended. Furthermore, the aircraft components such as wings can also deform or bend due to their own weight, which may lead to an offset in the overlay. Accurate tracking and precise calibration are active research topics, and better algorithms leading to highly precise virtual-real alignment are expected in the future.

35.4.2 Interaction

In traditional AR applications, users may require only basic interaction such as pointing or clicking the virtual content. An AR digital twin in turn requires manipulation of the virtual content in different ways depending on the lifecycle phase of the product. The interaction should not only support information presentation and selection but also support editing operations such as creating new or removing and moving existing content within the AR experience. While some applications designed to support the *Creation* and *Production* phases require extensive manipulation to place subcomponents in the right position, others may require simpler operations such as purely visual inspection. This subsection describes popular techniques used to interact with the virtual content in AR, highlights the challenges faced during the interaction design process, and suggests how these challenges can be overcome.

Interaction in AR can be achieved through various channels of input such as gaze, gestures, speech, and manual input via hardware controllers, keyboards, and mouses. Modern AR headsets such as Microsoft HoloLens offer manual input via a clicking device as well as gaze, hand gestures, hand movement, and voice as input. An equally important component of interaction is the communication of feedback to the user through an output channel, which, for example, can be visual, auditory, and/or haptic. Multiple channels are often employed together to communicate the feedback. Unlike VR, traditional AR headsets are not directly supplied with 6Degrees of Freedom (DoF) tracked controllers. But appropriate mechanisms can be established to allow using them together. The HoloLens offers integrated "gaze-based ray-casting" (Gaze) as an interaction metaphor. A cursor appears in the Field Of View (FOV) of the user indicating a forward ray in head direction, which can be used for pointing real and virtual objects in the AR environment as shown in Fig. 35.6.

With hand gestures, objects can be selected and moved in a limited tracking area in front of the AR headset. Furthermore, with speech recognition techniques, it is possible for the headset to listen to voice input. These input methods are useful for applications that require simple interaction. However, complex manipulation tasks demanding repeated rotation and translation of virtual objects can be performed faster with a 6DoF manipulation controller [41].

In order to evaluate a 6DoF controller as an input choice for complex manipulation tasks, we combined the Microsoft HoloLens, with an external tracked 6DoF

Fig. 35.7 A user manipulating a satellite component using a tracked controller. (© DLR 2020, CC-BY 3.0)

controller. Advanced Realtime Tracking (ART) SmartTrack infra-red tracking device was used for external tracking along with ART Flystick 3 controller for 6DoF manipulation. We created a new interaction metaphor to facilitate the placement, i.e., translation and rotation of subcomponents in the AR application [41]. When the controller touches a movable virtual object, the object attaches itself rigidly fixed to the controller, i.e., it is rotated and translated in 6DoF along with the controller. The controller button can be pressed to detach virtual objects from the controller. Figure 35.7 shows a user manipulating a satellite component with such a controller.

Compared to the interaction metaphor with a controller, the HoloLens' default gaze- and gesture-based manipulation metaphor allowed for either 1DoF rotation or 3DoF translation at a time. 1DoF rotation can be performed by selecting a control point on the object through the gaze- and gesture-based technique and moving the hand in horizontal or vertical direction.

To evaluate the interaction methods, we implemented an AR application for the *Creation* phase of a satellite in which different virtual subcomponents of the satellite had to be manually placed in the satellite in a predefined position and orientation. We conducted a user study in which the users assembled the satellite once using the 6DoF controller and once using the HoloLens' gaze- and gesture-based method. The order of interaction methods presented to the participants was counterbalanced. The task completion time for each interaction method was also recorded for every user. Each user was also asked to decide his/her preferred interaction choice in the end. Through the results of the study, we found out the 6DoF controller was easy to use, time efficient, and a preferred choice of interaction for the users [41].

Another issue to be addressed during the interaction with virtual content is the experience of arm fatigue. During the user study discussed above, the participants reported arm fatigue after experiencing both interaction paradigms. Even though desired manipulation could be accomplished faster using the 6DoF controller, the fatigue still arose for both methods.

Arm fatigue occurs when a person wearing an AR headset keeps his arm in mid-air for a long time in order to perform interaction tasks. The hand gesture recognition sensors are located in front of the headset. This often requires users to bring their hand to the height of the headset's sensors which, when done repeatedly, causes arm fatigue. Ideally, users must be able to interact with the virtual information efficiently without having any fatigue or exhaustion.

In order to address the arm fatigue problem, we brought into consideration the domain knowledge about components involved in the satellite creation phase. Based upon the context information and requirements of the assembly task, e.g., balancing the center of gravity, a suggestion engine was developed which could locate the ideal position and orientation of a satellite component based on mathematical optimization. In addition to the optimization, the existing gaze interaction of the HoloLens and a HoloLens Clicker is made available in this interaction metaphor to allow the user to determine the final position of the component. When a component is gazed at and selected, it attaches to the cursor of the HoloLens. When the cursor hits a virtual surface of the satellite housing, the object orients itself parallel to the surface. When the desired position on the surface of the satellite is reached, the click button can be pressed to release the object. Using domain knowledge of satellite design, the geometric normals of the satellite surfaces and the components are preconfigured in the application so that it aligns itself to the correct orientation.

Similar to the previously described user study, this suggestion-based method was evaluated for the *Creation* phase of a satellite. In the user study, the participants had to place satellite components of different shapes, sizes, and weights inside the satellite housing. During this task, they had to keep in view the constraint that the center of gravity of the satellite should be as close as possible to the geometric center of the satellite. The user study participants were not aware of the fact that an optimizer was running in the AR application to provide optimum suggestions. It was observed that optimized suggestions enabled the users to perform the task significantly faster compared to when they were served with non-optimized suggestions and the precision of object placement increased. Furthermore, the mental workload felt by the users was significantly lower when they were helped with optimized suggestions.

The results proved quite promising for the satellite use case as the interaction was heavily tailored for the required task. However, in AR digital twin applications, it is not always possible to get detailed domain knowledge for every use case to tailor the interaction accordingly.

Based on the user studies, we can conclude that a good interaction design must comply for every aspect of the AR experience including physical effort, mental effort, and senses involved in the experience. With scientific progress, AR headsets are not only ergonomically improving, but they are also integrating more intuitive input channels such as finger and eye tracking. These advances are expected to open up new possibilities of AR interaction.

35.5 Dedicated AR Applications

In this section, we present various AR applications in aerospace that are based on the concept of digital twins and MBSE.

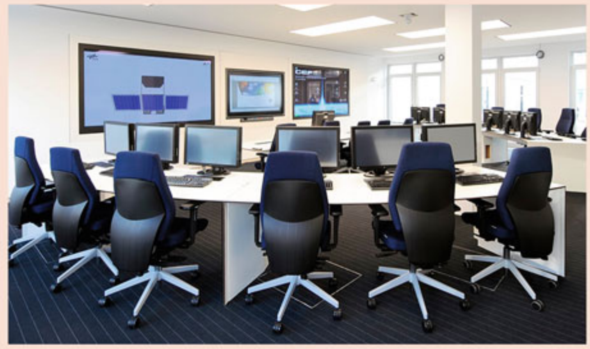


Fig. 35.8 The CEF at DLR Bremen is a specific meeting room for designing spacecraft. The CEF is optimized to improve communication between engineers. (© DLR 2020, CC-BY 3.0)

35.5.1 From Diagrams to AR in Spacecraft Design

One of the main tools for spacecraft design is a Concurrent Engineering Facility (CEF) such as the one at the DLR in Bremen. CEF is a special meeting room that is optimized to improve communication between engineers, as shown in Fig. 35.8. For a spacecraft design study, around 12 engineers from different domains are invited to this room. Usually they discuss the preliminary design for 1 week, but sometimes this process can take several weeks. During this time, the engineers discuss important aspects of the spacecraft. At this point, their goal is to create a good first draft of the spacecraft, matching the mission requirements. Such studies follow a defined process. The first day is about decomposing the system. Every engineer decides which components are needed for their subsystem, e.g., the power engineer selects two batteries. On the second day, each component is complemented by its mass. They are automatically summarized to determine the total masses of the subsystems and spacecraft. These accumulated parameters are the first important Key Performance Indicators (KPIs). In the next days, other system aspects are brought into the discussion. They range from power consumption of components to equipment positions in the structural configuration. This whole design study activity requires a lot of communication and information exchange. Changing one aspect of a spacecraft often impacts all other domains. Therefore, to keep all the data consistent, all the engineers enter their information into a single database called Virtual Satellite.

Figure 35.9 shows the user interface to the Virtual Satellite database as a desktop application. It shows parts of the decomposed system in the navigator on the left, a mass summary on a subsystem in the center editor, as well as the system structural configuration visualized as 3D model on the right. Every engineer interacts and designs with this interface. Within the CEF, each engineer uses the same application but with a reduced set of user interfaces according to his/her specific domain. During the study, the database is filled step by step with information about the spacecraft system. The data entered is exchanged with all other engineers and analyzed on the fly. Results can be simple warnings or, as shown, pie chart diagrams indicating the weight distribution. This process and database help the engineers to progress quickly and keep the spacecraft design consistent [21, 22].

The data stored in the database may be quite abstract. Diagrams such as a pie chart (Fig. 35.9, left) help to get a better understanding of the data and the design. Nevertheless, in some cases, only diagrams are often not enough to provide a good insight into the data. One example is the data involved in the structural configuration. Usually every engineer knows the best where his/her equipment has to be placed. For example, science and payload engineers know in which direction a camera has to point. However, they may have little domain knowledge about other parts of the spacecraft, and their expectation for these parts may be conflicting in size and position. A possible solution for this issue is to have a dedicated engineer configuring the whole spacecraft. In the traditional approach, such an engineer asks all other experts for relevant information to start designing a first version of the spacecraft in a CAD tool. This avoids positional conflicts but requires a tedious process of gathering all domain knowledge necessary to place the equipment. Additionally, this process can be very time-consuming since the configuration engineer has to synchronize the structural configuration between the CAD tool and database manually. Our idea is to collaboratively place the parts in the database and visualize the result for everyone. This means all engineers are placing their components themselves using an AR application. The configuration information is exchanged in the database and visualized. A simple 3D model can be shown in each instance of the Virtual Satellite desktop application or as hologram on

35

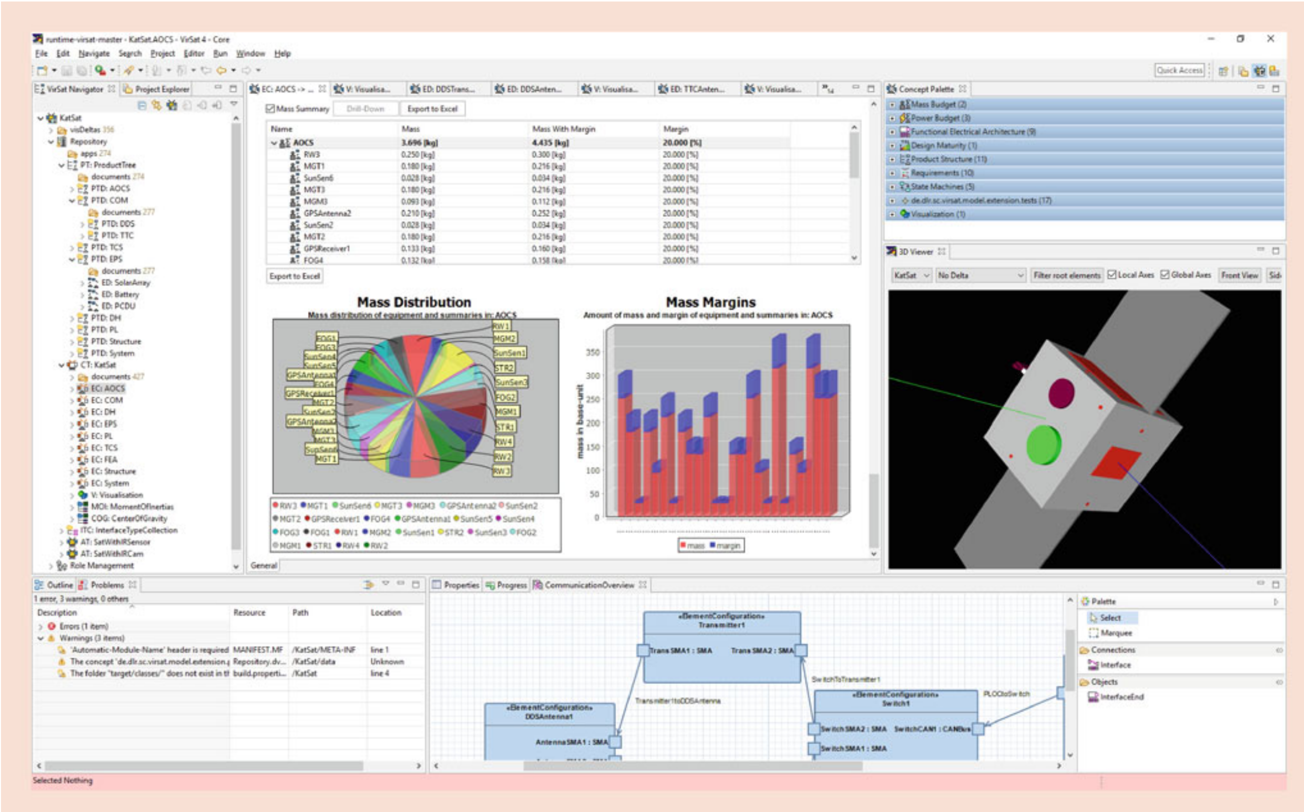


Fig. 35.9 The database Virtual Satellite as desktop application. It shows the decomposition of the system (left), a weight distribution of a subsystem (center), a 3D visualization of the spacecraft (right), and a block diagram of an electrical architecture (bottom). (© DLR 2020, CC-BY 3.0)

the table. In case of conflict, the participating engineers can discuss a solution together [42].

The construction and visualization of a 3D model are an important tool not only during early design but also in later design phases. During early design, a preliminary sketch of the spacecraft helps to provide physical context to abstract numbers. In later phases, providing physical context is still very useful, but the design detail increases. From a structural configuration perspective, the 3D model is augmented with a great deal of domain-specific information. Besides position and orientation of parts, mounting brackets, screws, and other information become relevant. Further properties about materials, such as alignment constraints, can be interesting as well. From a system engineering perspective, usually only the positions and orientations are of interest. This reduced information is eventually shared with, e.g., thermal domain engineers. They can use the positions and sizes of equipment for their specific simulation tools. In the frame of MBSE, it is therefore interesting to establish an engineering round trip between the system and the CAD models. In whatever model a position is changed, it has to be synchronized into the other model. Still the synchronization does not corrupt the augmented detailed information about other parts such as screws, mounting brackets, etc. Accordingly, the system model remains the single source of truth for all other domains [29].

Figure 35.10 shows an example of a cube sat. At the top, the visualization of the system model can be seen, mainly with simple primitives. At the bottom, the domain-specific structural configuration is shown. It shows a more detailed version from a 3D model. Both models are synchronized, but the 3D model is augmented by, e.g., mounting brackets (yellow) under the solar panels.

On system engineering level, the visualized model can also be augmented with additional information. One example is to overlay the visualized model with heat map based on preselected properties, as shown in Fig. 35.11. Overlaying such kind of data directly on top of the model can increase perception, as it helps to illustrate properties along with the physical context. This approach can be extended by not just displaying the current design state but a delta to a previous state. Liu et al. [43], e.g., showed thrusters which have been modified and a box which was replaced with a smaller one compared to the previous version.

Displaying the spacecraft on traditional desktops is already of advantage for design discussions. However, Deshmukh et al. [44] showed that displaying it in an interactive stereoscopic 3D environment can additionally

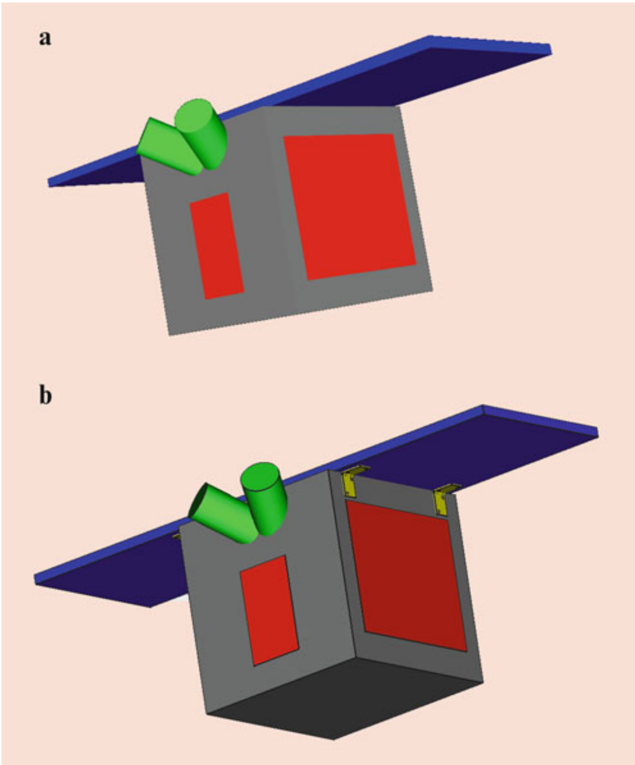


Fig. 35.10 MBSE round trip engineering with CAD on the example of a cube sat. The spacecraft in Virtual Satellite (**a**) is kept synchronized with the one in CAD (**b**). Additional engineering information (yellow mechanics) is kept in the domain tool and is not synchronized. (© 2018 DLR, CC-BY 3.0)

improve design discussion. One advantage was on the use of head tracking. Engineers do not necessarily need to turn the model but could move naturally around it to inspect hidden parts. Further, the increased space in front of large stereoscopic environments allowed for a discussion with more engineers compared to a desktop screen. An interesting aspect was the possibility of direct interaction with the visualized model. Engineers could select a spacecraft part manually and move it into a new position. The system model was directly synchronized with changes applied to the visualization. This natural interaction is far different from just using mouses and keyboards in front of a desktop.

Based on these approaches, AR devices such as the Microsoft HoloLens open up new possibilities for collaboration and interaction. With such a device, engineers can walk around a holographic representation of the spacecraft during the design process. With hand tracking and gesture recognition, an engineer can directly interact with the model. Figures 35.12 and 35.13 show the interaction with a satellite model on an engineer's desk and on a design room, respectively. Rather than displaying the real size spacecraft, a miniature version can be displayed besides the computer. Changes to the design are again directly synchronized between the system model in the database and the visualization [45]. These applications are the baseline for future applications of remote collaboration. This can range from remote collaborative design sessions to remote assistance in AIT campaigns. In the later years, 3D holographic models can be overlaid with the real spacecraft in the integration center. This indicates where, e.g., parts of the spacecraft need to be mounted or cables have to be connected. Further information such as documentation or formerly paper-based mounting instructions can be directly displayed to the engineers.

35.5.2 On-Orbit Servicing

The accumulation of space debris over the last years is endangering future space missions. When a satellite breaks down or runs out of fuel, it usually remains in the Earth's orbit increasing the amount of debris. In the best case, remaining fuel can be used to steer the satellite toward the Earth's atmosphere where it would burn up. A better solution, however, would be to repair or refuel satellites. But those service operations can still just be carried out accurately by astronauts. Additionally, human space flights are extremely dangerous and very expensive. Therefore, space agencies are looking for reasonable alternatives. A promising solution is the usage of a service spacecraft with a connected robot arm. It is controlled from the Earth to approach the defective satellite. After a successful maneuver and docking, the robot arm can be used remotely from the Earth to repair the damaged satellite.

The VR-OOS project by DLR [29] aims at the evaluation of On-Orbit Servicing (OOS) in virtual environments. Instead of one arm, a complete robot equipped with stereo camera and two arms comes into operation. The setup is shown in Fig. 35.14. The counterpart is an operator who wears a head-mounted display where the video stream from the robot camera is projected. The head is tracked to control the position and orientation of the robot camera. Additionally, the operator is using a bimanual haptic input device (called Haptic User Gert (HUG) [46]) which captures the movements of her or his arms to directly control the movement of the robot arms accordingly (as can be seen in Fig. 35.15). The HUG has been developed at DLR and consists of two lightweight robot arms which are very similar to the ones attached to the servicing robot. The input device can also provide haptic feedback. When a robot hand collides with the satellite, the force information is transmitted to the HUG. The device uses the received data to apply forces against the hands of the operator which enables to feel the satellite. This allows for an intuitive control of the servicing robot.

In VR-OOS, instead of the real space environment, a virtual satellite and virtual robot arms are rendered from the robot's perspective (cf. Fig. 35.16). The scene is a photorealistic reconstruction of the space environment. The basis of

35

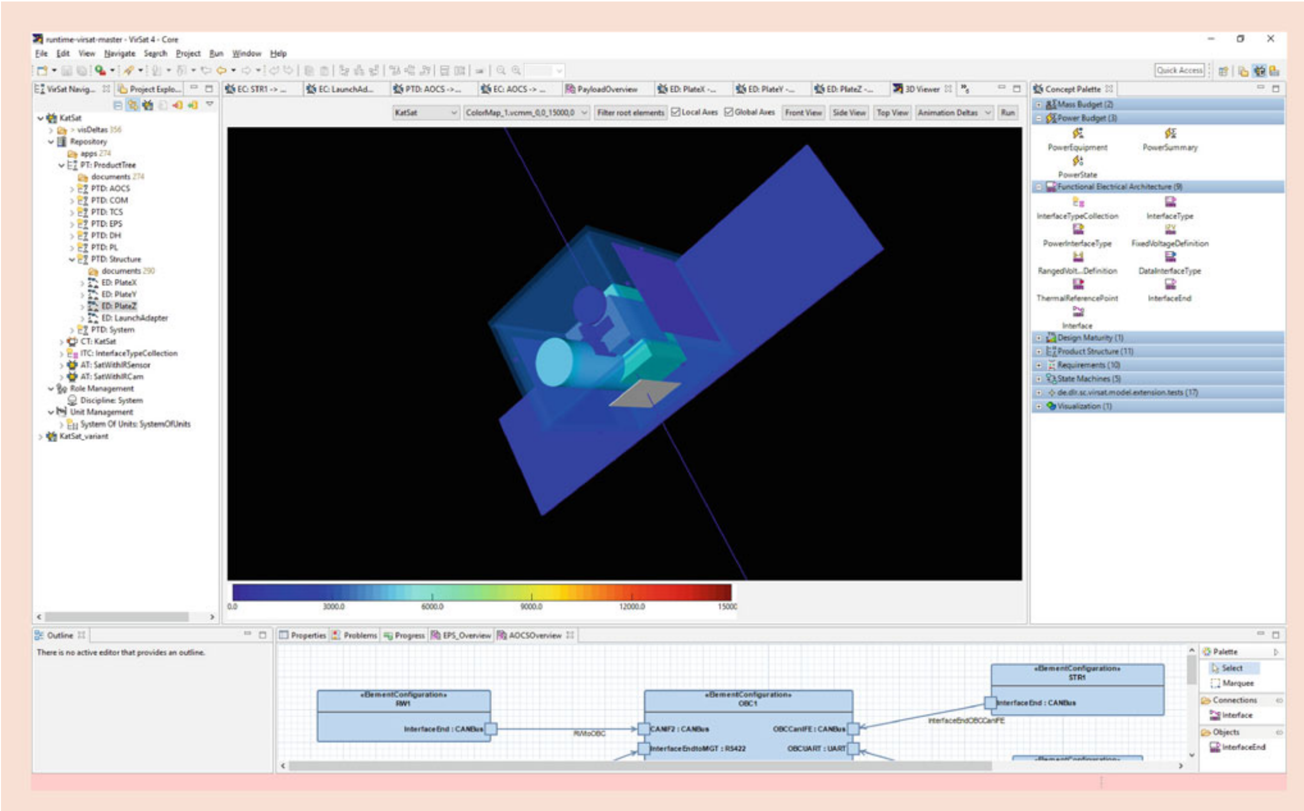


Fig. 35.11 Visualization of a spacecraft with augmented information highlighting a specific distribution in a heat map. (© DLR 2020, CC-BY 3.0)

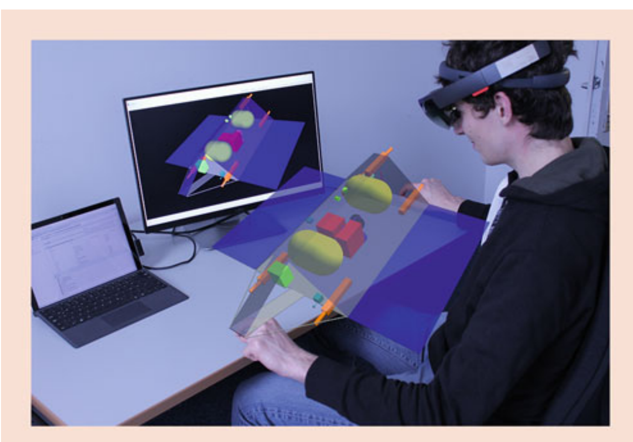


Fig. 35.12 3D holographic visualization of a spacecraft based on Microsoft HoloLens. (© DLR 2020, CC-BY 3.0)

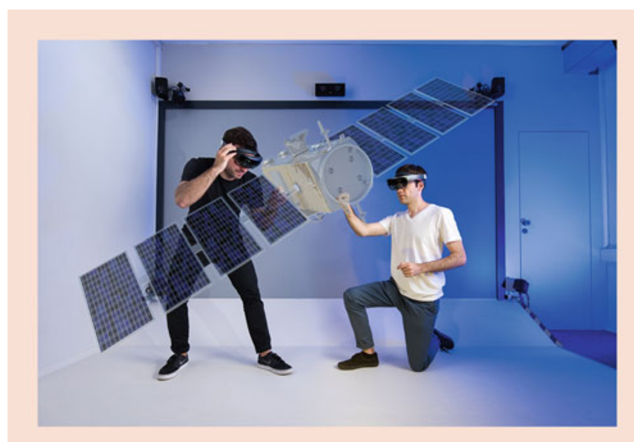


Fig. 35.13 Two engineers inspect a satellite together. (© 2018 DLR, CC-BY 3.0)

this application is a digital twin of the satellite that, besides a 3D model, provides also additional information such as documentation or annotations for every component of the satellite as well as for the robot.

AR-Based Guidance and Training

Additional to the physics-based VR-integration of the OOS scenario, AR can be used for guidance and training tasks. To support the HUG operator, an expert wearing an AR headset stands next to the operator. Connected to the DTI, the AR application receives data from the robot-satellite assembly. It allows the expert to visualize the virtual satellite model directly in front of the exoskeleton, simulating the robot-satellite assembly in space. Using AR has the advantage that the expert can still see the operator and can immediately inform him about mistakes or potential risks. Since

Fig. 35.14 The servicing robot is attached to a broken satellite. (© DLR 2018, CC-BY 3.0)

Fig. 35.15 The operator sits in an exoskeleton. (© DLR 2018, CC-BY 3.0)

the expert can freely walk around the model, he/she is able to see components that are hidden from the perspective of the operator. Additionally, parts of the virtual model can be made transparent to be able to see hidden satellite parts. The expert can help in avoiding collisions between the robot and sensible satellite parts, which may result in damaging or losing satellite parts.

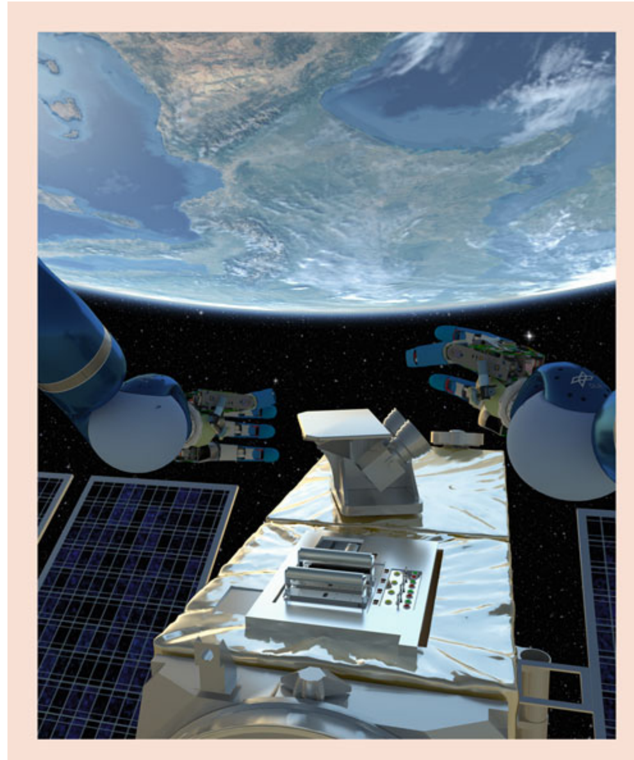


Fig. 35.16 View through the robot's cameras to the satellite in orbit. (© DLR 2018, CC-BY 3.0)

Fig. 35.17 An expert creates a virtual barrier. (© DLR 2018, CC-BY 3.0)

To constrain the servicing robot movement, the expert can place virtual barriers as can be seen in Fig. 35.17. These barriers can act as virtual walls through which the robot's arm cannot be moved.

To ensure a successful operation, the operator must first train the repair. The training scenarios are very similar to the real operation, with the difference that it can be repeated as often as necessary. In this context, a trainer takes the lead and guides the operator.

35.5.3 Remote Collaboration and Maintenance

This subsection describes an AR application to facilitate maintenance of aircraft components by bringing multiple users into a shared virtual space showing the DTI.

Maintenance, Repair and Overhaul (MRO) includes the repair, service, or inspection of an aircraft or aircraft component. It consists of all the maintenance activities that take place to ensure safety and airworthiness of all aircrafts by international standards. Once an aircraft component comes into operation, its MRO requires detailed inspection in order to exclude any possible internal or external damages.

At DLR, aircraft components are redesigned by inventing new composite structures. They can significantly decrease production effort [47]. Even though such composite structures are quite robust, they can be vulnerable to certain damage types. Regular visual and sensor-based inspections are essential for composite maintenance. Material defects can grow over time when the integrity of a composite structure is compromised. Impacts caused by loose objects on a runway or by tools accidentally dropped on a composite surface might look harmless at first glance. Since composites are layered, even hidden material layers can crack without being noticed. Structural Health Monitoring (SHM) can be very helpful to address such hidden damages. Via piezoelectric sensors emitting and receiving lamb waves, SHM can compare the resonance of an intact structure with a damaged one and therefore autonomously detect impacts. Because these sensors can be directly installed in the composite structure, they can provide a live view of an aircraft component. Further classical approaches are the use of X-ray, ultrasonic scans, and volumetric captures that represent the damaged state of a component.

In order to support MRO using AR, we interviewed experts to identify the most common tasks involved in performing a physical inspection. Then a prototype application was developed to demonstrate how AR can facilitate MRO by streaming information from the DTI and visualizing it on top of a real composite aircraft component. It demonstrates an end-to-end MRO process and especially how AR helps to locate damaged spots and investigate and document them during inspection. As use case, we modeled the inspection of a composite air brake component. Accessible is the DTP data specified in the design phase. This can be compared with the data stored in its DTI. Additionally, image markers and SHM sensors are attached to it. The damage reports and the SHM data are fed directly into the DTI.

In order to visualize existing damage entries within the DTI on top of the physical air brake, we show information windows anchored to relevant damage spots in 3D. Anchored windows are also used by an inspector to request relevant DTI information during the physical inspection or repair task. An attached image marker or a calibration tool helps to align

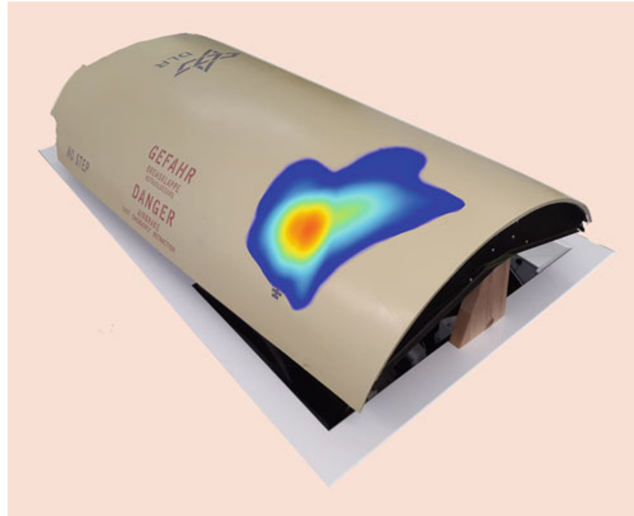


Fig. 35.18 SHM data from the ultrasonic sensors highlighting a damaged location of the air brake. (© DLR 2020. CC-BY 3.0)

our DTI overlay including information windows with the physical air brake. Whenever SHM data is available, it is streamed onto the AR headset and visualized as heat map textures on the air brake's surface. The heat map in Fig. 35.18 overlaid on the air brake indicates the distributed damage probability.

To analyze the airworthiness of the air brake, a Finite Element Method (FEM) simulation can be applied to the digital twin [48]. However, high-fidelity simulations need the exact damage location of an impact. We put the inspector into this simulation loop by requesting a manual probe: Using an ultrasonic device, the inspector is able to pinpoint the exact damage location even below the surface. AR-based systems may also include real-time visualization of ultrasound images within the physical component. Therefore, the inspector can interact with both the real air brake and its virtual scan in a shared field of view. The in-place visualization of ultrasonic scan data is shown in Fig. 35.19

In case a new damage location is manually selected, it is sent to a remote machine where the damage analysis is carried out. The DTI again supports the communication and synchronization between information visualization and simulation. As soon as the damage simulation result is ready, it is visualized in AR anchored on the air brake's surface. The inspector can navigate through different material layers which are colored separately. The visualization of a cross-section of the air brake is depicted in Fig. 35.20.

Analyzing the risk of crack growth in certain layers may require expert knowledge. For example, the inspector could require assistance by an expert for structural analysis or by the manufacturer of the air brake. Our AR application handles this requirement by tracking communication channels in the DTI. Once the inspector and a remote expert are

Fig. 35.19 A manual ultrasonic scan (C-Scan) visualized on an air brake depicting the affected parts. (© DLR 2020, CC-BY 3.0)

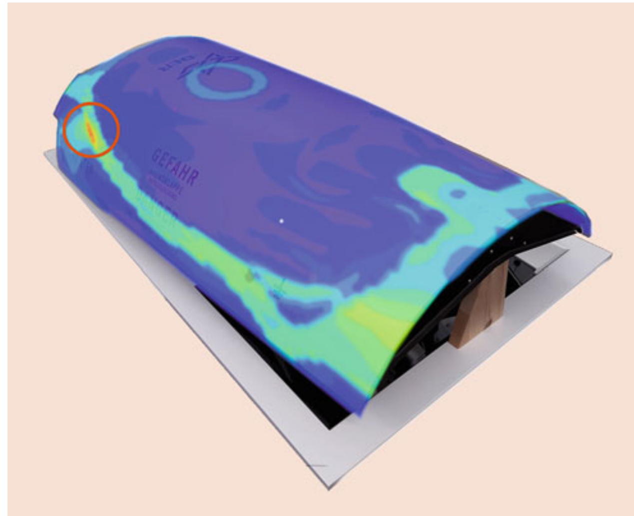


Fig. 35.21 Visualization of strain simulation on the fiber layer of the air brake caused by high stress. (© DLR 2020, CC-BY 3.0)

Fig. 35.20 AR Visualization of the cross-section of the air brake. (© DLR 2020, CC-BY 3.0)

connected, they can discuss the air brake's status based on all the information available in the DTI. The expert sees a virtual air brake projected into his or her work environment. Information from the design phase which is also part of the digital twin can also be visualized. For example, Fig. 35.21 shows the simulated strain on carbon fiber layers of the air brake caused by applied stress. The red-circled region in Fig. 35.21 highlights the part of the air brake that is prone to damage when high stress is applied to it. Additionally, all collected data including information windows, overlays, and findings are superimposed. As the DTI is shared on the fly, voice and avatar information can be exchanged. This enables the inspector and expert to talk to each other and also see each other as 3D avatars relative to the air brake. Hence, the inspector sees the expert's avatar next to the physical air brake, and the expert can see the inspector's avatar next to the virtual air brake.

The remote collaboration session can be closed whenever the inspector and expert reach a common conclusion. However, during most of inspections, an expert may not be available right away. In this case, the inspection session can be recorded and sent to the expert, who can view it independently. Such an asynchronous remote collaboration could also benefit the repair task in order to recap the damage analysis.

Closing the inspection job triggers the following repair task or labels the air brake to be ready for service. It means that a whole MRO task chain is mapped. The data collected in the DTI does not only embody an official documentation and status history of the air brake; it also proofs a correct inspection procedure for legal requirements.

Based upon the inferences drawn from the above implementation, we identified the most common attributes required to carry out technical inspection.

It can be observed from Table 35.1 that the combination of AR and digital twin supports most of the attributes of a physical inspection and enables to carry out the process remotely.

35.6 Closing Remarks

This chapter discussed advantages of connecting digital twins, MBSE, and AR to support different phases of the aerospace system lifecycle. We addressed important aspects and limitations of state-of-the-art AR headsets for the usage in digital twin applications. In this context, improvements for

Table 35.1 Inspection attributes and their support offered by each of the systems

Attributes	Physical inspection only	Physical inspection + AR	Remote inspection + AR
View component (physical)	✓	✓	X
View component (3D model)	X	✓	✓
Interact with the component	✓	✓	✓
Visualize information in place	X	✓	✓
Interact with fellow experts/technicians	✓	✓	✓
Touch/haptic feedback	✓	✓	X
Record inspection session	X	✓	✓

overlay precision and different interaction metaphors were investigated. Furthermore, we demonstrated the usefulness of AR digital twins in three exemplary aerospace applications: satellite design, satellite servicing, and maintenance of aircraft components. The first one has been presented as a prototype application for the *Creation* phase. The latter two were examples for the *Operation* phase.

Although AR is less widespread than traditional visualization displays, many applications aiming at different phases of the spacecraft lifecycle can already benefit from upcoming AR devices. Nevertheless, a large-scale use of AR is still hindered by the aforementioned limitations. This includes the interconnection of AR devices with the digital environment, the connection to MBSE, and the integration with operational data and information of the whole PLM/Product Data Management (PDM) process. All these issues have to be addressed to leverage AR technologies in the aerospace domain. Once solved, the incorporation of the full potential of digital twins and the intuitive AR-based interaction into the aerospace working environment should just be a last small step.

References

1. Grieves, M., Vickers, J.: Transdisciplinary perspectives on complex systems—new findings and approaches. In: Chapter Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems, pp. 85–113. Springer, Switzerland (2017)
2. Grieves, M.: Virtually Perfect: Driving Innovative and Lean Products through Product Lifecycle Management. Space Coast Press LLC, Cocoa Beach, Florida (2011)
3. Grieves, M.: Digital Twin: manufacturing excellence trough virtual factory replication. White Paper **1**, 1–7 (2014)
4. Negri, E., Fumagalli, L., Macchi, M.: A review of the roles of digital twin in CPS-based production systems. Procedia Manufacturing **11**, 939–948 (2011). https://doi.org/10.1016/j.promfg.2017.07.198
5. Haag, S., Anderl, R.: Digital Twin—proof of concept. Manufacturing Letters **15**, 64–66 (2018). https://doi.org/10.1016/j.mfglet.2018.02.006
6. Boschert, S., Rosen, R.: Digital Twin—the simulation aspect. In: Hehenberger, P., Bradley, D. (eds.) Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and their Designers, pp. 59–74. Springer, Cham (2016)
7. Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., Wang, L.: Modeling, Simulation, Information Technology and Processing Roadmap. NASA Headquarters, Washington, DC 20546 (2012)
8. Miranda, D.: 2020 NASA Technology Taxonomy (2019)
9. Tuegel, E.J., Ingraffea, A.R., Eason, T.G., Spottswood, S.M.: Reengineering Aircraft Structural Life Prediction Using a Digital Twin. Int. J. Aerosp. Eng. **2011**(Article ID 154798), 14 (2011). https://doi.org/10.1155/2011/154798
10. Aeronautics, N.N., Administration, S.: NASA Systems Engineering Handbook Revision 2 (2017)
11. ECSS Secretariat: ECSS-E-ST-10C Rev.1 System Engineering General Requirements. Noordwijk, Netherlands (2017)
12. Schlager, K.J.: Systems engineering-key to modern development. IRE Trans. Eng. Manag. **EM-3**(3), 64–66 (1956)
13. Sillitto, H., Martin, J., McKinney, D., Griego, R., Dori, D., Krob, D., Godfrey, P., Arnold, E., Jackson, S.: Systems Engineering and System Definitions. CRC press, New York (2019)
14. Operations, I.T.: Systems engineering vision 2020, pp. 1–32. INCOSE, San Diego, CA, accessed Jan 2020 (2007)
15. I. I. C. on Systems Engineering: Systems engineering vision 2025 (2014)
16. OMG: System Modelling Language—Version 1.5 (2017)
17. OMG: Unified Modelling Language—Version 2.5.1 (2017)
18. OMG: XML Metadata Interchange (XMI) Specification—Version 2.5.1 (2015)
19. Estefan, J.A.: Survey of Model-Based Systems Engineering (MBSE) Methodologies (2008)
20. Hendricks, R., Eickhoff, J.: The significant role of simulation in satellite development and verification. Aerosp. Sci. Technol. **9**, 273–283 (2005)
21. Fischer, P.M., Deshmukh, M., Maiwald, V., Quantius, D., Gomez, A.M., Gerndt, A.: Conceptual data model: A foundation for successful concurrent engineering. Concurrent Engineering Research and Applications **26**(1), 55–76 (2018)
22. Deshmukh, M., Schaus, V., Fischer, P.M., Quantius, D., Maiwald, V., Gerndt, A.: Decision support tool for concurrent engineering in space mission design. In: Stjepandic, J., Rock, G., Bil, C. (eds.) Concurrent Engineering Approaches for Sustainable Product Development in a Multi-Disciplinary Environment, vol. 1, pp. 497–508 Springer, London (2013)
23. ECSS Secretariat: ECSS-M-ST-10C Space project management—Project planning and implementation. Noordwijk, Netherlands (2009)
24. Fischer, P.M., Lüdtke, D., Lange, C., Roshani, F.-C., Dannemann, F., Gerndt, A.: Implementing model-based system engineering for the whole lifecycle of a spacecraft. CEAS Space Journal **9**(3), 351–365 (2017)
25. de Koning, H.P., Gerenü, S., Ferreira, I., Pickering, A., Beyer, F., Vennekens, J.: Open Concurrent Design Tool—ESA Community Open Source—Ready to Go!. In: Proceedings of the 6th International Conference on Systems and Concurrent Engineering for Space Applications, Stuttgart (2014)
26. ESA: Virtual Spacecraft Design (2013)
27. Eisenmann, H., Cazenave, C., Noblet, T.: RangeDB the product to meet the challenges of nowadays System Database. In: Simulation

and EGSE for Space Programmes (SESP). Noordwijk, Netherlands (2015)
28. ScopeSet - the tools experts: Mapping and importing Catia Properties. http://www.vsd-project.org/download/ssdevideos/CatiaImport.swf accessed Apr 2016 (2016)
29. Bernstein, A.-C., Utzig, S.: Ich sehe was, was du auch siehst. DLR Magazin **159**, 44–47 (2018)
30. Fischer, P.M., Fuchs, J., Eisenmann, H.: Functional Verification by Simulation based on Preliminary System Design Data. In: Proceedings of the 6th International Workshop on Systems and Concurrent Engineering for Space Applications (SECESA 2012) (2014)
31. Cazenave, C., Fancelli, L.: Digitalization Lessons Learnt for Spacecraft Engineering. In: Workshop on Simulation and EGSE for European Space Programmes (SESP 2019) (2019)
32. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: A class of displays on the reality-virtuality continuum. In: Telemanipulator and Telepresence Technologies, vol. 2351, pp. 282–292 (1995)
33. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virt. **6**(4), 355–385 (1997)
34. Azuma, R.T.: Making augmented reality a reality. In: Applied Industrial Optics: Spectroscopy, Imaging and Metrology, pp. JTu1F–1 (2017)
35. Vuforia Engine. https://engine.vuforia.com/engine, accessed Jan 2020 (2020)
36. Hincapié, M., Caponio, A., Rios, H., Mendívil, E.G.: An introduction to Augmented Reality with applications in aeronautical maintenance. In: Proceedings of the 2011 13th International Conference on Transparent Optical Networks, pp. 1–4 (2011)
37. Utzig, S., Kaps, R., Azeem, S.M., Gerndt, A.: Augmented Reality for Remote Collaboration in Aircraft Maintenance Tasks. In: 2019 IEEE Aerospace Conference, pp. 1–10 (2019)
38. De Crescenzio, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., Salti, S.: Augmented reality for aircraft maintenance training and operations support. IEEE Comput. Graph. Appl. **31**(1), 96–101 (2010)
39. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. Robot. Comput. Integr. Manuf. **49**, 215–228 (2018)
40. Vassallo, R., Rankin, A., Chen, E.C., Peters, T.M.: Hologram stability evaluation for Microsoft HoloLens. In: Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment, vol. 10136, p. 1013614 (2017)
41. Bahnmuller, A., Muhammad, A.S., Albuquerque, G., Gerndt, A.: Evaluation of Interaction Techniques for Early Phase Satellite Design in Immersive AR. In: Proceeding of the 2020 IEEE Aerospace Conference (to appear) (2020)
42. Fischer, P.M., Wolff, R., Gerndt, A.: Collaborative Satellite Configuration Supported by Interactive Visualization. In: IEEE Aerospace 2012 (2012)
43. Liu, Y., Deshmukh, M., Wulkop, J.C., Fischer, P.M., Gerndt, A.: Real-time immersive visualization for satellite configuration and version comparison. In: Workshop on Simulation and EGSE for European Space Programmes (SESP 2017) (2017)
44. Deshmukh, M., Wolff, R., Fischer, P.M., Flatken, M., Gerndt, A.: Interactive 3D visualization to support concurrent engineering in the early space mission design phase. In: Challenges in European Aerospace (2015)
45. Baranowski, A., Utzig, S., Fischer, P.M., Gerndt, A., Herder, J.: 3D spacecraft configuration using immersive AR technology. In: Herder, J., Geiger, C., Dörner, R., Grimm, P. (eds.) Virtuelle und Erweiterte Realität – 15. Workshop der GI-Fachgruppe VR/AR, pp. 71–82. Shaker Verlag, Herzogenrath (2018)
46. Hulin, T., Hertkorn, K., Kremer, P., Preusche, C., Schätzle, S., Artigas, J., Sagardia, M., Zacharias, F.: The DLR bimanual haptic device with optimized workspace. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3441–3442 (2011)
47. Hanke, M., Froese, S.: Eurofighter airbrake–a demonstration of advanced composite design. Jahresbericht Wehrwissenschaftliche Forschung des BMVg, englische Ausgabe **2017**, 44–45 (2018)
48. Garbade, M.: Efficient simulation of the through-the-thickness damage composition in composite aircraft structures for use with integrated SHM systems. In: Proceedings of the 9th International Conference on Computational Methods (ICCM2018) (2018)

Georgia Albuquerque is a research scientist at the department “Software for Space Systems and Interactive Visualization” at the German Aerospace Center (DLR). She received a bachelor’s degree in computer science from the University of Pernambuco, Brazil, in 2003, and a master’s degree in computer science from the TU Braunschweig, Germany, in 2007. She received a PhD degree at the Computer Graphics Lab from the TU Braunschweig in 2013, and continued as postdoctoral researcher until 2019. Her main research interests focus on augmented and virtual reality as well as on visual analytics, machine learning, and computer graphics.

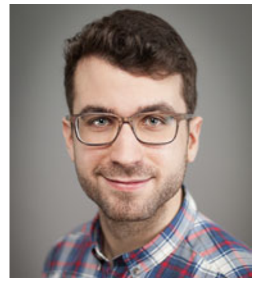
Philipp M. Fischer is scientist and software engineer at DLR’s Simulation and Software Technology Lab, where he is in charge of the “Modelling and Simulation” group. His research interests are in model based systems engineering (MBSE) and data modelling, or, more specifically, applying MBSE to the complete lifecycle of space systems. As part of this activity, he supported several Phase A studies at DLR’s Concurrent Engineering Facility. Currently he is involved in the digital transformation from MBSE towards digital twins.

Philipp M. Fischer received his diploma degree in electrical and computer engineering from the Leibniz University Hannover in Germany in 2007. During his studies he spent one year abroad focusing on computer science at the Swinburne University of Technology in Melbourne/Australia.

During employments at Toyota Motorsports and Airbus Defense and Space and others, Philipp M. Fischer contributed his experience in software engineering and MBSE to various projects and R&D activities. He acted as session chair and co-chair in several workshops and conferences related to these R&D topics.

Syed Muhammad Azeem is a research scientist at the department "Software for Space Systems and Interactive Visualization" at German Aerospace Center (DLR). He has BE "Computer Engineering" from National University of Sciences and Technology (NUST), Pakistan, and MS "HCI & Robotics" from Korea University of Science and Technology (UST), South Korea. He moved to Germany in 2017 to join DLR.

Andreas-Christoph Bernstein is a research scientist at the department "Software for Space Systems and Interactive Visualization" at the German Aerospace Center (DLR). He received a diploma in computer science from Bauhaus-University Weimar in 2008, where he then worked as a researcher until 2017. His main research interests focuses on computer graphics, virtual and augmented reality, and machine learning.



Sebastian Utzig received a M.Sc. in Computer Science and Media from Bauhaus University Weimar, Germany, in 2016. He specialized in Virtual Reality and Interactive Visualization and became researcher at the German Aerospace Center (DLR) investigating VR and AR applications in scientific environments. Today, he works at a Berlin-based agency, where he develops immersive solutions for various companies.

Andreas Gerndt is a Professor for High Performance Visualization at University of Bremen and the head of the department "Software for Space Systems and Interactive Visualization" at the German Aerospace Center (DLR). He received his degree in computer science from Technical University, Darmstadt, Germany in 1993. In the position of a research scientist, he also worked at the Fraunhofer Institute for Computer Graphics (IGD) in Germany. Thereafter, he was a software engineer for different companies with focus on Software Engineering and Computer Graphics. In 1999, he continued his studies in Virtual Reality and Scientific Visualization at RWTH Aachen University, Germany, where he received his doctoral degree in computer science. After two years of interdisciplinary research activities as a postdoctoral fellow at the University of Louisiana, Lafayette, USA, he returned to Germany in 2008.

Index

A
Abaqus, 534
Above and around the keyboard surface, 117
Absolute orientation, 215
Accelerometers, 58, 217, 218
Access control, 163
Action Research Arm Test (ARAT), 748
Active defense, 166
Active haptic systems, 114
Active markers, 227
Activities of Daily Living (ADL) training, 744
Actor, 435, 436, 441–443, 446, 448
AdaGrad algorithm, 68
ADD metric, 93
Adopting a I4.0 strategy, 865
Adoption of DTs, 866
Adversaries, 160
Aerospace
  aero glass, 479
  aircraft, 474
  airport security, 488
  applications, 898
  AR applications, 489
  AR technology, 476, 487
  artificial intelligence (AI), 489, 490
  augmented reality (AR), 489, 490
  automotive vehicle, 475
  aviation industry, 474
  cognitive tunnelling, 478
  communication, 475
  COVID-19, 488, 491
  crew task Support, 488, 489
  driving experience, 475
  emerging applications, 474
  engineering, 483, 489
  engineering design process, 482, 483
  flight instruments, 476
  gunsight, 473, 474
  haptic feedback, 490, 491
  In-Flight Entertainment and Communication (IFEC), 489
  instruments, 475
  maintenance, 483–485
  manufacturing, 483–485
  mental exhaustion and fatigue, 476
  navigation routes, 474
  navigation tasks, 474
  physiological challenges, 478
  primary tasks, 474
  remote maintenance, 485
  simulations, 477
  smart glasses, 478
  space operations, 486, 487
  tele-maintenance, 485
  3D symbology, 478
  3D visual information, 478
  traffic control and management, 474
  visualization, 482, 483
  workload, 476
Affordances, 817
Agent avatar, 844
AI controlled characters
  agent's actions, 133
  ALIVE system, 133
  AR environments, 132
  architecture, 138
  assistants, 137, 138
  author control vs. autonomy, 147
  basic structures, 139
  blue flowers, 139
  categorization, 136
  coaches, 136, 137
  cognition, 138
  cognitive abilities, 135
  companions, 137, 138
  component knowledge, 139
  computer vision techniques, 133
  coordinated look, 145
  decision trees, 148, 149
  design guidelines, 144
  emotions, 134
  enemies, 138
  facial and body animation, 133
  FSMs, 149
  gestures, 134
  goal oriented behavior (GOB), 149, 151
  human-AI communication, 140, 143
  human- human communication, 140, 141
  human user, 144
  immersive applications, 132
  integration, 132
  mixed reality (MR), 133
  movement, 145–147
  non-verbal communication, 141
  opponent, 138
  option manager, 148
  options, 139
  perception, 138
  personality, 144
  physical space, 138
  planning algorithms, 147

A. Y. C. Nee, S. K. Ong (eds.), *Springer Handbook of Augmented Reality*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-67822-7

AI controlled characters (*cont.*)
PlayStation VR Device, 134
real barriers, 139
real/virtual objects, 134
reasoners, 147
research community, 132
smartphones, 131
social components, 134
social facilitation and inhibition, 135
social inhibition, 135
social interaction, 135
social presence, 135, 140
speech, 134
state-of-the-art game engines, 133
storyline/workflow, 144
subject of an exanimation, 137
trainee, 136, 137
trainer, 136, 137
user interaction, 139
user interface (UI), 133
user's presence, 144
utility AI, 151–153
verbal communication, 142, 143
virtual and real objects, 135
virtual characters, 133
virtual content, 138
virtual humans, 132
visual agents, 132
visualization modes, 134
Air brake, 896
Air Force Research Laboratory (AFRL), 41
Air Traffic Control Officer (ATCO), 474, 480, 481
Air Traffic Management (ATM), 474, 480, 481
Airbus DS, 885
Aircraft, 881, 896
manufacturing, 620
Airport security, 488
Airport Surveillance Radar (ASR), 481
A2J, 99–100
AlexNet, 67
ALICE Project, 434, 435
Alignment problem, 274
ALIVE system, 804
AlphaPose, 229
Amazon Alexa, 801
Amazon Echo Show, 801
Amazon Sumerian, 44
Amazon Web Services (AWS), 44
Ambient intelligence, 803
Ambient lighting, 401
Analysis of Variance (ANOVA) tests, 759
Anatomical knowledge, 690
Anchoring, 843
Android, 376
Ani library, 379
Annotated training sets, 91
Anomalous reflection, 199
ANSYS, 534
ANSYS Parametric Design Language (APDL), 555
Appearance layer, 140, 144
Apple, 800
Apple's ARKit, 131
Apple Siri, 801
Application
layer, 292
scale, 625
AR-assisted healthcare exercising systems, 744
ANOVA, 759, 760
ARAT, 748
AR-based systems, 746, 747
ARTHE system, 749, 750, 759, 760
bare-hand tracking, 748, 749
daily living activities, 745, 755
bookcase arrangement, 756
faucets, 756
opening doors, 756
scoring module, 756
data glove, 751
exercise module, 753
experiment procedure, 758, 759
experimental *vs.* control group, 759
FMA, 748
glove-based devices, 748, 749
haptic-based systems, 746
motion tracking, 751
motor functions recovery, 745, 746
non-dominant hands, 758
participants, 758
pre-test and post-test design, 759
reach-to-grasp movement, 745
ROM exercises, 745
scoring module, 753
flaccid stage, 753
HPT, 754
reach-to-grasp, 754
ROM, for finger, 753
ROM, for shoulder, 753, 754
scoring modulegrip stability, 755
scoring modulesynergy stage, 754
synergy stage, 754
sensor-based tracking, 747, 748
task-oriented training, 745
upper-extremity, 744
usability experiment, 758
vibration wrist band, 752
VR-based systems, 745, 746
WMFT, 748
AR-assisted healthcare systems, 744
AR-based Guidance, 601
AR-based system, 359
AR-based virtual showcase, 601
Archaeology learning
3D navigation, 316
mobile cultural access, 316
modality, 316
morphology, 316
Architecture, 311, 809, 810
Architecture, Engineering, Construction, and Operations (AECO), 496
ARCore, 44, 659, 660
Arich, 394
AR-IoT, 809, 810, 817
architecture, 809
context-based AR-IoT interactions and content interoperability, 814–815
embodied AR agent insights, 819–820
embodied AR-IoT agent prototypes, 818–819
embodied interactions, in AR-IoT environments, 817–818
framework evaluation, 815–816
object-centric AR-IoT data management, 810–812
potential use cases, 820–821
scalable AR-IoT recognition and tracking, 812–813
ARKit, 44, 660

AR-PEclass, 359–361, 364
ARQuake, 42, 711
AR Remote Maintenance Support Service (ARRMSS), 630
Articulated hand tracking (AHAT) mode, 767
Artifact evaluation phase
  building material, 511
  building material store, 518
  float replacement, 520
  NASA TLX questionnaire, 511
  perception filters, 522–525
  total sample characterization, 511, 512, 516, 517
  university campus, 519
  workload and factors analysis, 517–519, 521, 522
Artificial intelligence (AI), 47, 132, 325, 407, 479, 489, 490, 769, 798, 800, 837
  and augmented reality, 804–805
  and ubiquitous computing, 803–804
Artificial neural networks (ANN), 536
ARToolKit, 86, 545, 659, 660
Arts education
  cognitive process, 307
  communication, 308
  digital process, 306
  ecosystem, 307
  educational technology, 306
  expeditions, 308
  gesture control, 307
  hegenomous culture, 308
  participative cultures, 307
  pedagogy, 309
  physical computing, 306
  pluralism, 308
  road map, 308
  significant artefacts, 306
  universal human rights, 308
  user experiences, 306
Assembly, Integration, and Test (AIT), 885, 893
Asset Lifecycle, 860
Assistive systems, 783, 785–787, 789
Assistive technologies, 765
Associative phase, 357
Asymmetric links, 284
Attack scenarios, 173
Attention, relevance, confidence, and satisfaction (ARCS), 364, 365
Attractiveness, 375
Audio portion of AR, 638
Audio-visual language, 320
  body language, 321
  editing & setting, 321
  film, 321
  macrolanguage, 321
  texts, 321
Augmented reality (AR)
  advertising, 45
  assembling, 44, 45
  cultural heritage, 45
  forensics, 46
  guidance, 645
  headsets, 889
  maintenance, 44, 45
  medical uses, 45, 46
  military applications, 45
  mock-ups, 887
  repair, 44, 45
  services, 454, 456, 458
  sightseeing, 45
  technologies, limitations, 386, 619, 624, 625, 629, 631
  Tennis, 43
  ToolKit, 42–44
Augmentationism, 771
Augmented and mixed reality (AR/MR)
  applications, 644
  architectures for shipbuilding applications, 649–651
  challenges, 661, 664
  commercial developments, 645
  communications, 647
  design and construction assistance, 645
  device layer, 649
  devices, 647
  guided manufacturing, 646
  hardware and software for shipbuilding applications, 651–661
  hidden area visualization, 647
  IIoT devices, monitoring and interaction with, 648
  markers, 658
  potential shipbuilding tasks, 646–649
  predictive maintenance, 647
  product and tool maintenance, 649
  quality control, 646
  shared information, 645
  shipbuilding painting, 645
  shipbuilding welding, 644
  step-by-step guidance, 645
  structure visualization, 648
  training, 648–649
  user experience, 646
  warehouse management, 647
Augmented and virtual reality (AR/VR), 109
Augmented autoencoders, 96
Augmented Immersive Team Trainer (AITT), 624
Augmented Live Interactively Controlled Environment (ALICE) Project, 435
Augmented reality (AR), 4–7, 110, 111, 113–115, 117, 118, 131, 136, 244, 246, 420, 421, 427–430, 576, 669, 691, 710, 744, 798, 832–834, 837, 840, 841, 882, 883
  academic tourism community, 452
  accessibility, 733
  acquisition process, 314
  age considerations, 731, 732
  aided servicing, 602
  aim, 598
  applications, 35, 43, 336, 377, 710
  applications in research of human memory and cognition, 771–776
  AR-assisted maintenance guidance system, 576
  AR-assisted maintenance systems, 579
  AR-assisted maintenance training systems, 576
  ARCore, 43
  in art, 718, 719
    appreciation and experience, 719
    future, 720
    limitation, 719, 720
    young children, 719
  and artificial intelligence, 804–805
  AR/VR-assisted maintenance systems, 578
  ARGs, 710
  ARKIT SDK, 43
  ARToolKit, 43
  audio, 631, 638
  auditory information, 338
  augmentation, 46
  basics, 35, 36
  beamsplitters, 188
  biomimetics, 47

Augmented reality (AR) (*cont.*)
  camera technology in, 767
  cautions, 724
  character marker, 40
  children, 348
  cluster, 464, 465
  clustering algorithm, 461
  in cognitive load research, 776–781
  commercial applications, 691
  communication, 324
  communicative activities, 452
  complexity, 345
  computer-generated information, 534
  computer vision, 600
  computing paradigms for, 768–769
  concept, 710
  construction phases, 598
  consumer, 336, 465
  consumer mobile devices, 346
  COVID-19 pandemic, 733
  culture gourmets, 463, 464
  culture-seeking activities, 462
  cutting-edge technology, 598
  definition, 35, 597, 618
  developments, 335
  devices and tools for, 766–768
  didactic information, 314
  digital assistants, 346
  digitally mediated experiences, 337
  digital twin (*see* Digital twin)
  disabled, 724
  discomfort, 203
  diversity, 336
  in education, 335, 376, 720
    college, 722
    context, 377
    elementary school, 720, 721
    field, 47
    graduate education, 722
    high school, 721, 722
    medical school students, 722, 723
    middle school, 721
    patients, 723, 724
    physicians and clinicians, 723
    preschool, 720
    use, 337
    vocational training, 722
    vocations, 724
  enabling technologies, 36
  experience data, 377
  explorers, 462, 464
  external FEA program, 549
  games, 45, 631
  gender consideration, 732, 733
  German tourists, 462
  growth, 46
  guidance, 692
  haptic, 631
  haptic channels, 338
  hardware display, 187
  HCL tool, 535
  head mounted displays, 766
  health and safety concerns, 345, 346
  hedonists, 463
  history, 40, 766
  HMD, 598
  holograms, 188
  human capability, 335
  human cognition, 346
  IIoT, 600
  imaging and anatomical modeling, 693–694
  industry, 336
  Industry 4.0, 43, 600, 601
  information appliance concept, 347
  information appliances, 346
  ingress, 336
  interface, 347
  inverse stiffness matrix, 548, 549
  IoT, 598
  issues and challenges in application for cognitive impairments, 788–789
  laparoscopic adrenalectomy, 691
  learning, 335
  learning activities, 348
  learning contexts, 376
  learning structural mechanics, 537
  limitations, 898
  linear stress-strain relationship, 537
  location based positioning systems, 338
  location detection, 343, 344
  Magic Leap ONE, 44
  maintenance, 575, 576, 598, 599
    application, 577, 580
    automated authoring, 590
    benefits, 577
    closed-loop maintenance systems, 588, 589
    context-awareness, 589, 590
    contextual awareness, 581, 587
    digital twin integration, 587
    effectiveness, 577
    ergonomic assessment, 588
    ergonomic consideration, 590
    hardware, 584, 585
    HCI tool, 577
    human motion tracking, 582, 583
    industrial, 586
    Industry 4.0, 590, 591
    IoT, 587, 590, 591
    literature review, 578, 579
    remote, 587, 590
    remote maintenance, 580
    tracking, 581, 582, 590
    training platforms, 579
    unified and robust design approach, 587
    uniform implementation method, 586
    user interface and interaction, 583, 584, 586
    user tracking, 588
    virtual information, 581
  malware, 788
  marker based augmented reality, 429
  markerless augmented reality, 429
  marketing aspects, 336
  measurement system, 337
  mediated reality, 46
  in medical practice, 725
    educational opportunities, 725
    health/wellness goals, 726
    medical procedures, 725
    mental health, 726
    nonprocedural medical uses, 726
    patient communication, 725, 726
    therapeutic uses, 725

methodology, 459, 460
Microsoft Hololens 2, 44
mixed reality classification, 337
mobile, 620, 621, 624–626, 637
mobile platforms, 536
mobile technologies, 451
modality, 314
monodimensional real-versus-virtual continuum, 46
nanotechnology, 47
navigation, 46
new generation, 47
non-visual formats, 337
observations, 378
operational aspects, 336
in operation theater, 692
optical design, 190
optical see-through HMD, 39
overlay display, 600
pandemic consideration, 733
PCG method, 549
pedagogical enablers, 348
pedagogy, 345
pediatric patients, 710
platonic solids, 376, 378
positional tracking, 36–38
positioning information, 338
process of learning, 346
projection-based, 40
questionnaire, 377
real-time FEA, 548, 549
real-time numerical simulators, 537
remote video assistance, 600
repair, 598
repair procedures, 602, 603
research design, 459, 460
research questions (RQs), 458, 459
respective spatial constructions, 376
for restoring perception, 781–788
robustness to environment, 626
safety precaution
distracted driver prevention, 728
driving, 727, 728
DVT, 729
eye strain, 729
indoor considerations, 727
mental health, 728–730
outside considerations, 727
overuse injuries, 729
spatial ware and body position, 728, 729
sales, 336
scientific visualization techniques, 536
segments, 462
sensing and measurement, 536, 537
simulation, 337
Sketchpad application, 40
smart assistants, 346
smart glasses, 451
smartphone applications, 336, 464
social and ethical challenges, 348
social media networks, 714
Facebook, 715
future, 716, 717
Instagram, 715
Messenger, 715
Pinterest, 716
precautions and risks, 716
snapchat, 714, 715
TikTok, 715, 716
social network applications, 47
spatial, 768
spatial and temporal registration, 337
specifications, 376
in sports, 717
athletic training, 717
future, 718
limitations, 718
spectator interactions, 718
stakeholders, 452
statistics, 461
students, 376
system for training, 623
tablet, 464
technical issues, 342, 343
technology, 376, 452, 534
telepresence, 47, 337
time delay, 537
tourism organizations, 452, 465
tourism sector, 452
tourist segments, 461, 464, 466, 467
training aspects, 336
types, 35
and ubiquitous computing, 805–806
urban respondents, 460
usability, 344
usage preference, 463
use cases, 463, 635
user experience, 452, 460
user's perception, 451
uses, 766
value of augmentation, 335
Varjo XR-3, 44
video see-through HMD, 39
virtual and real objects, 534
virtual assistants, 603, 604
virtual elements, 36
virtual information, 188
virtual objects, 337
virtual reality, 337
visual display, 338
visual positioning systems, 338
visualization, 38
voice commands, 600
wearable, 622
young adults, 348
Augmented reality, for cultural heritage, 398, 399, 407
ambient lighting, 401
ARETE, EU project, 2019–2023, 398
calibration, 403
CEMEC project, 397
connectivity, 402
content optimization, 403
drawings on glass at Carnuntum (Austria), 2011, 396
efficacy and effectiveness, 404–407
FoV, 400, 402
historical background, 393–394
iMARECULTURE, EU project, 2016–2020, 398
Jumieges 3D, Normandy (France), 2013, 396
latency, 401
monitor-based AR display, 394–395
multi-sensorial AR experiences, 403–404
occlusion, 401
Olympia, 2000, 395–396

Augmented reality, for cultural heritage (*cont.*)
  projector-based AR / spatial AR, 394–396
  Revealing Flashlight, EU project, 2013, 397
  S. Maria Antiqua Videomapping, Rome (Italy), 2015, 397
  see-through AR displays, 394
  technical limitations and challenges, 400–403
  VAC, 400
  ViaggiArte, Fontanelle Cemetery, Naples (Italy), 2019, 397
  virtual lighting, 401–402
Augmented reality, for surgery
  physics-based modeling, 697–700
  registration and pose estimation, 696–697
  surgical planning, 702–703
  surgical vision, 694–696
  training and knowledge transfer, 701
  visualization and perception, 700
Augmented reality games (ARGs), 710
  ARQuake, 711
  benefits, 712
  future, 714
  popularity, 712, 713
  safety precautions, 713, 714
  smartphone technology, 711, 712
Augmented Reality Manual (MAR), 505
Augmented reality (AR) marketing
  academics, 453
  advertising, 452
  communicative activities, 452
  elements, 452
  financial value, 453
  hotel building company, 453
  internal and external stakeholders, 452
  on-site AR experiences, 454, 456
  pre-booking and information-gathering stage, 453, 454
  retailers, 453
  stakeholders, 453
  tourism, 453, 456, 458
  user evaluation, 454
Augmented Reality for Product (or Quality) Assurance (ARPASS), 486
Augmented Reality for Remote Assistance (ARRA)
  AR approach, 670
  assumptions, 671
  CAD model, 671
  hardware development, 674
  image processing, 682
  MRO applications, 670
  novice maintainer, 674
  piping systems, 675
  quantitative spatial referencing error, 681
  spatial referencing, 672, 675
  spatial referencing error, 676–683
  tool, 670
  unity application, 674
  validation purposes, 675
  vs. video-call support, 675, 680, 682, 683
  virtual objects, 671
  working, 672, 673
Augmented Reality Visualization of the Common Operational Picture (ARVCOP), 619
Augmented virtuality (AV), 136, 799, 832
Augmenting humans, 119
Augmenting on and around the keyboard, 117
Authentication, 170
Authoring maintenance, 628
Authorization, 168, 170
Autism Spectrum Disorder (ASD), 135
Automated identification system (AIS), 627
Automatic phase, 357
Availability, 168
Avatar, 599, 817
Awareness, 818, 819

B
Backchannel (BC), 143
Back-propagation, 68
Baggage, 488
Balance of power, 169
Bandwidth requirements, 275
Bare-hand tracking method, 748, 749
Battery, 661
  life, 657
Battle Damage Assessment and Repair (BDAR), 485
Battlefield Augmented Reality System, 621
Battlespace visualization, 628
Bayes' theorem, 256, 779
BB8, 93–96, 98
Beginning of Life (BOL), 860
Behavioral model, 604
Bending degrees, 749
Bernoulli mixture model, 295
BigHands dataset, 92
Birdbath, 196
Block-based matching, 61
Blockchain, 651
Bluetooth, 344
Body-worn, 242
Boeing Augmented Reality Kit (BARK), 484
The Boeing Company, 474
Boston Consulting Group, 453
Boundaries, 169
Bridge of ship, 619
Brightness constancy, 71
Bring Your Own Device (BYOD), 313
Brushless direct current (BLDC), 27
Bufferbloat, 285, 287, 292
Bug-eye reflectors, 198
Building information modeling (BIM), 497
Building maintenance
  artifact design, 505, 506
  artifact development
    Adobe Photoshop, 508
    Apple and Android devices, 505
    application screen, 511
    AR model, 509
    AR view screens, 512
    Autodesk's BIM Revit, 506
    characteristics, 508
    floater engine reference, 507
    graphical interface, 509
    LAR/MAR marker rating, 509
    MAR smart glasses version, 515
    MAR tablet version, 513
    mobile devices, 509
    smart glasses versions, 510
    stages, 506
    tablet version, 510
    3D Studio Max, 508
    3DS Max, 506
  artifact publication
    Android and iOS operating systems, 510
    application publishing, 516

application testing, 517
TestFlight tool, 510
AR visualization, 527
evaluation, 526
generalization
activity, 526
AR application prototypes, 527
AR incorporation framework, 528
AR model process, 526
component demands, 526
DSR method, 526
physical demand, 527
workload factors, 527
Building maintenance instruction manual (BMIM)
applications, AR, 497
AR visualization, 497
BS 8210: 2012 cites, 496
building operations, 497
building representation, 496
contractual guarantee, 496
delivering/selling, 496
design, 497
European Union (EU), 496
evaluation, 497
facilities management, 497, 498
facilities manager, 497
factors, 496
hydraulic installations, 497
multidisciplinary, 497
research challenge, 496
technical instructions, 496
tracking and detection technologies, 497
two-dimensional (2D), 496
user perception and interaction, 497
virtual reality (VR), 497
visualization technologies, 496
web systems, 497
Bundle adjustment, 66
Business models, 873
Bystander privacy, 159, 173
Bystanders, 162

C
Calibrated accelerometer signal, 58
Calibrated gyroscope signal, 57
Calibration, 56, 888
coordinate system, 888
inaccuracy, 889
multi-point approach, 888
re-calibration, 889
Calibration methods
absolute orientation, 215
hand–eye calibration, 214
IMU, 217–219
notations, 213
OST-HMD, 221–223
projector–camera calibration, 220
tip tool calibration, 213
tracking targets, 214
California Institute of Technology, 486
Calm technology, 800
Camera(s), 54, 360
calibration, 56
extrinsic parameters, 55
image distortion, 55
intrinsic parameters, 54
perspective camera model, 54
pixel coordinates, 55
rolling shutter effects, 57
spherical camera model, 57
Camera-based motion capture systems, 227–228
CamIO system, 784
Camouflaging techniques, 170
Capacitive sensing keyboard, 117
Carbon fiber layers, 897
Carpal tunnel syndrome, 729
Cathode ray tube (CRT), 7, 188
Cellular phone, 226
Center of gravity, 890
Center-to-edge swipe, 249
Central educational authority, 312
Challenges, 848
Chatbots, 801
Chromaglasses, 785
Cinema of poetry, 324
Ciphertext-policy attribute-based encryption, 164
City View AR-system, 394
Clarity, 481
Classful traffic, 286
Classifiers, 173
Clickjacking, 167
Client, 808
Close air support (CAS), 623
Closed-loop maintenance systems, 588, 589
Cloud, 649, 810
computing, 276, 279, 296, 769
and edge-assisted MR processing, 159
Cloudlets, 651
CMU Panoptic hand dataset, 92
Coarse-to-fine strategy, 61
CodeCubes
activities, 380
application, 379, 381
AR marker, 379
computer vision detection, 379
data collection, 380
digital games, 379
easiness, 381
enthusiastically, 381
game, 381
game screen, 379
group of students, 380
interface, 379, 380
intuitiveness, 381
learning, 380, 381
levels, game, 379
limitations, 381
objectives, 380
performance, 381
physical programming blocks, 379
player, 379
procedure, 380
program instruction, 380
questionnaires, 381
solutions, 381
strategy, 379
visual programming, 379
Co-development, 174
Cognitive, 374
disability, 773
phase, 357
processes, 476

Cognitive augmented reality assistant (CARA), 783
Cognitive impairments, 773
  issues and challenges in application of AR for, 788–789
  perception related, 781–783
Cognitive load
  AR application in research of, 779–781
  example, 777
  measurements techniques, 778
  subjective techniques, 777
  types of, 776
Collaboration, 375, 820, 883
  AR, 279
  environment, 603
  interactions, 166, 169–170
  SA, 621
Collaborative Working Environments, 599
Collaborative, Research, Engineering, Analysis, and Training in Immersive Virtual Environments (CREATIVE) Lab, 630
Color perception, 632–637
Color sequential, 194
Color vision deficiency, 782, 785
Combiner, 196
Commercial AR, 292, 295
Commercial developments, 645
Communication technology, 312
Communities, 833
Comparison, 883
Competencies, 374
Computational, 809
  resource, 240
  thinking, 373
Computational fluid dynamics (CFD), 483
Computed tomography (CT), 689–691, 693
Computer aided design (CAD), 482, 883, 885, 886, 888, 891
Computer-aided engineering (CAE), 885
Computer Aided Medical Diagnostics and Surgery System (CADMASS) project, 487
Computer-assisted surgery, 703
Computer games, 419, 420, 424
Computer-guided interventions, augmented reality
  imaging and anatomical modeling, 693–694
  physics-based modeling, 697–700
  registration and pose estimation, 696–697
  surgical planning, 702–703
  surgical vision, 694–696
  training and knowledge transfer, 701
  visualization and perception, 700
Computer multimedia, 358
Computer Science Laboratories (CSL), 42
Computer vision techniques, 86, 600, 766, 887
  deep learning, 66–69
  depth image tracking/ICP, 66
  direct image alignment, 64
  epipolar constraint, 63
  feature matching, 60
  feature tracking, 60–62
  line tracking, 64
  pose estimation from feature correspondences, 62–63
  SfM, 65
  3D pose estimation, 87–89
COMSOL, 534
Conceptual data model (CDM), 885
Conceptual design, 884
Conceptual models, 886
Conceptualization, 883
Concurrent Engineering Facility (CEF), 887, 891
Congestion, 287
Conservative plane releasing, 166
Constitutive law, 698
Constrained interfaces, for augmented reality, 240
  constrained hardware configuration, 241
  freehand pointing on constrained hardware, 244–246
  HIBEY, 253–258
  indirect touch on miniature interfaces, 241
  optimized text entry layout design, 246–247
  physical constraint from mobility, 241
  text entry on constrained screen real estate, 246
  TiPoint, 248–253
  TOFI, 258–265
Construction learning, 327
Consumer Protection Code, 496
Content caching server, 650
Content hiding, 168, 170
Context, 834, 836, 840, 841
Context Aware Access Control (CAAC), 839
Context-adaptive, 817
Context-aware, 118, 839
Context-awareness, 584, 798, 803
Contextual reality (CoRe) frameworks, for XRI system design, 841–842
  proof of concept designs, 843–846
  XRI taxonomy perspectives, 844–847
Continuous/multi-factor authenticators, 173
Contrast ratio (CR), 194
Control group, 362
Convergence, 802–808, 813, 817, 818, 820, 822
Convolutional neural network (CNN), 66, 97
Cooperation, 375
Cooperative gestures, 169
Coordinate system transformations, 54
Co-presence, 819
Co-registration, 36
Cornea, 190
Corner-to-corner swipe, 249
Coronavirus disease pandemic 2019 (COVID-19), *see* COVID-19 pandemic
Cost optimization, 601
Coulomb-Contensou model, 102
Counter-intuitive user interfaces, 535
Countermeasures, 173
COVID-19 pandemic, 47, 324, 488, 491, 710, 721, 722, 733
Crawling waves, 24
Creative expression
  cognitive values, 316
  collaborative dimension, 317
  communication, 317
  didactic pathways, 317
  imagination, 317
Crisp imaging, 200
Cross-authentication, 170
Cross-platform interoperability, 160
Cross-reality, 807, 833
Cube sat, 892
Cultural awareness and expression, 314
Cultural diversity, 311
Cultural education, 309
Cultural heritage, augmented reality, 398, 399
  ambient lighting, 401
  ARETE, EU project, 2019–2023, 398
  calibration, 403
  CEMEC project, 397
  connectivity, 402

content optimization, 403
Drawings on glass at Carnuntum (Austria), 2011, 396
efficacy and effectiveness, 404–407
FoV, 400, 402
historical background, 393–394
iMARECULTURE, EU project, 2016–2020, 398
Jumieges 3D, Normandy (France), 2013, 396
latency, 401
monitor-based AR display, 394–395
multi-sensorial AR experiences, 403–404
occlusion, 401
Olympia, 2000, 395–396
projector-based AR / spatial AR, 394–396
Revealing Flashlight, EU project, 2013, 397
S.Maria Antiqua Videomapping, Rome (Italy), 2015, 397
see-through AR displays, 394
technical limitations and challenges, 400–403
VAC, 400
ViaggiArte, Fontanelle Cemetery, Naples (Italy), 2019, 397
virtual lighting, 401–402
Cultural re-elaboration, 327
Cultural sensitivity and expression, 309
CyberCode, 42
Cyber foraging, 276, 279
Cybersecurity, 651, 664
Cyborgs, 18
CycleGAN, 101, 102

D

Damage analysis, 896
Damage model, 884
Dancing in Cyberspace, 42
Dashboards, 832
Data augmentation, 95
Data-centric categorization, 162
Data fusion, 637
Data glove, 751
Data literacy, 327
Data routing layer, 649
Datasets, 93
for hand pose estimation, 92
for object pose estimation, 91–92
Data slicing, 546
DCCP, 281
Decision making, 133, 136–139, 143, 147
Decision trees, 148, 149, 612
Deep learning, 66, 90, 94, 244, 769
convolutional neural networks, 66
network training, 67–69
for SLAM, 79–80
in tracking and mapping, 69
Deep learning-based trackers, 72
Deep network training, 90
Deep neural network, 173
Deep venous thrombosis (DVT), 729
DeepIM, 94, 96
DeepPrior++, 99
Defense applications, 618
Degraded visual environments (DVE), 477
Degree of freedom (DOF), 87, 548, 889
Deliberate mediation, 10
Dense map, 77
Depth cameras, 58, 113
Depth image tracking, 66
Depth maps, 98–100
Depth perception, 619, 621
Depth relationship, 618, 619
Depth sensors, 58–59
Design(s), 843
and construction assistance, 645
space, 117
Design science research (DSR) method, 497
Deweyan reflection, 323
Didactic mediation, 307
Diffractive coupling method, 201
Diffractive optical element, 200
Diffractive structures, 199
Diffuse curatorship, 320
Digital artefacts, 318
cognitive process, 319
communication, 319
knowledge process, 319
neural networks, 320
Digital biomarker-based prognostic models, 781
Digital communication, 312
Digital competences
audience development, 325
curators, 325
didactic methodology, 325
didactic pathways, 326
digital innovation, 324
discovery learning, 328
museum educators, 325, 326
museums, 325
onlife, 326
third space, 327
Digital engineering, 669
Digital environments, 307, 312
third space, 317
Digital Game Based Learning (DGBL) approach, 382
Digital heritage, 306
Digital Interactive Experience Developer, 325, 327
Digital living, 822
Digital objects, 786
Digital storytelling technique, 324
Digital twin, 648, 669, 807, 854, 859, 882–883
in aerospace applications, 883–884, 886–887
applications, 862, 866
creation phase, 883
digital twin environment (DTE), 883
digital twin instance (DTI), 883, 894, 896
digital twin prototype (DTP), 883
disposal phase, 883
implementations, 887–891
interaction, 889–891
operation phase, 883
overlay precision, 888–889
production phase, 883
unified repository, 882
Digital twin instance (DTI), 896, 897
Diminished reality, 46, 800
Direct image alignment, 64
Direct linear transformation (DLT) algorithm, 62
Direct manipulation, 267
Direct methods, 71
Direct SLAM, 77
Direct trackers, 71
Directionality, 169
Disassembly process, 872
Discovering, 375
Disk jockey (DJ), 9
Dismounted infantry training, 621–626

Dismounted personnel, 620
Display leakage, 172
Display-relative calibration (DRC), 222
Disposal, 886
Distortion correction software algorithms, 192
Distortion parameters, 55
Distributed, 798, 806
  computing, 626
  environments, 883
Distributed ledger technologies (DLTs), 651
Domain adaptation, 67
Domain knowledge, 891
Domain randomization, 104
DPOD, 97
6D Pose metrics, 93
Drift, 65
Drone Research and Application Centre (DRAC), 480
Drone Soar, 480
Dynamic decision strategies, 771

E
Edge computing, 173, 276, 279, 296, 651, 652
Edge server, 808
Edge trackers, 71
Education, 414, 415, 421, 427, 428, 430, 805
Educational experience(s)
  aesthetic experience, 323
  development, 312
  digital communication, 323
  digital media, 322
  images and learning, 322
  media education, 322
  modality, 323
  multimedia connotation, 324
  sectoral study, 321
Effectiveness, 375
E-learning, 358
  platform, 359
Electric machines and SWIM, 27, 29
Electrocardiogram, 691
Electro-magnetic propagation, 664
Electromagnetic sensor, 226, 307
Embedded cameras, 657
Embedded training, 622
Embodied agents, 135, 137, 803
Embodied autonomous agents, 132
Embodied conversational agents, 132
Embodied interaction, 242, 265, 799
  AR-IoT environments, 817–821
Embodiment, 118, 242
Empathetic substitution, 422, 423, 428
Empathy, 422
Encryption-based techniques, 164
End of life (EOL), 860
Engineering model
  spiral, 885
  Vee, 885
  waterfall, 885
Entertainment, 373, 805
  applications, 233
Enthusiasm, 386
Environmental understanding, 158
Epileptic seizures, 167
Epipolar constraint, 63
Epochs, 68
Equivariance, 66
Ergonomics, 588
Error reduction, 601
ESA's European Astronaut Centre, 486
Essential matrix, 63
Ethical considerations, 158
Ethics, 413–416, 422, 428, 430
European Commission, 310
European Cooperation for Space Standardization (ECSS), 884
European Geostationary Navigation Overly Service (EGNOS), 59
European level, AR
  assessment, 310
  cultural dimension, 310
  cultural reality, 310
  didactics, 312
  disciplines, 310
  economic/academic success, 310
  innovation, 310
  mobile devices, 312
  museums, 311
  school teaching, 310
  transparency, 313
European Space Agency (ESA), 486
Evaluation methods
  objective measurements, 223–224
  subjective measurements, 224
Everysight Raptor, 199
Exemplary ray trace, 189
Exoskeleton, 894
Experimental group, 362
Experimentation, 374
Extended Kalman filter (EKF), 218
Extended reality (XR), 46, 370, 854
Extended response, 8
External display(s), 170
  protection, 172
Extraneous cognitive load, 776
Extrinsic parameters, camera, 55
Extrinsic policy enforcement, 162
Extrinsic protection, 169
Eye clearance, 191
The Eye of Judgment, 42
Eye resolution, 190
Eye strain, 729
Eye tracking, 204, 223
Eye-wear form factor, 170
Eyebox, 190
  expansion, 201
EyeTap, 41

F
5G, 283
Facebook, 715, 800
Facial de-identification, 164
Facial tracking, 230
Fast point-feature histograms (FPFHs), 217
Fatigue, 890
Feature descriptors, 60
Feature detection and tracking, 653
Feature matching, 60
Feature tracking, 60, 70
  optical flow estimation, 61–62
Feed-through signalling, 169, 170
Fiducial markers, 812
Field demonstration, 627
Field of view (FoV), 188, 241, 400, 402, 406, 474, 661, 889
Film, 417, 418

Filtering methods, 638
Filtering of information, 623
Filtering-based fusion, 78
Fine-grained permissions, 172
Finer-grained policy enforcement, 175
Finger-worn, 267
Finite element analysis (FEA), 483
  ANN, 536
  AR integrated system
    architecture, 539, 540
    coordinate systems, 541
    design considerations, 538, 539
    hardware configuration, 540
    interaction tools, 540, 541
    object selection techniques, 541–543
    visualization and interaction, 543
  AR technology, 535, 538
  CAD system, 536
  counter-intuitive user interfaces, 534
  data manipulation and exploration, 546–548
  drawbacks, 535
  efficient interaction methods, 569
  engineering fields, 533, 534
  integration of VTK and AR, 545
  interactive model modification, 553, 554
    geometric models, 554
    local mesh refinement, 554, 555
    methods, 569
  interactivity, 535
  load acquisition, 549
    management and conversion, 550
    virtual loads, 552
    WSN configuration, 551–553
  mobile AR platforms, 570
  occlusion handling, 545, 546
  piecewise polynomial interpolation scheme, 534
  prototype application, 563, 565
  proving ring, 563
  real-time finite element modeling, 537, 538
  real time simulation, 570
  scalar stresses, 534
  scientific visualization, 544
  software systems, 534
  stages, 534
  step ladder, 555
    data visualization and exploration, 556, 558
    local mesh refinement, 563
    model preparation, 555, 556
    real-time simulation, 558–560
    structural stiffening, 560, 563
    system response time, 563
  stiffness matrices, 534
  structural health monitoring, 570
  system implementation, 555
  system workflow, 552, 553
  time synchronization, 552, 553
  traditional computer setups, 535
  visualization with VTK, 544
  VR technology, 535, 536
Finite element method (FEM), 699, 896
Finite state machines (FSM), 149, 612
First-Person Hand Action dataset, 93
First-Person View (FPV) displays, 479
Fitts, Paul M., 617, 618
Flaccid stage, 760
Flex sensors, 226
Flipped classroom, 358
Flow experience, 374
Fluoroscopy, 690
Fog computing, 651
FootMath
  benefits, 383
  camera, 383
  challenges, 383
  characteristics, 383
  classroom, 386
  curiosity, 385
  digital artifact, 382
  educational activities, 386
  educational tool, 383
  experiments, 385
  function, 383, 384
  graph, 383
  interaction, 385
  interest, 385
  joysticks, 383
  learning, 386
  learning mathematics, 382
  learning opportunity, 386
  logical reasoning, 384
  mathematical concepts, 382
  mathematical functions, 382, 383
  motivation, 385
  parameters, 383
  Platonic Solids, 386
  playful, 383
  plot, 383
  procedure, 385
  score, 383
  serious games, 383
  solutions, 384
  strategies, 385
  students, 382
  teachers, 382
  technical development, 382, 383
  3D football, 383
  triggering, 383
  type of application, 386
Force-assisted interaction, 262
Forward air controllers (FACs), 623
FountX AsR, 485
Framework, 659, 799
  Convention, 309
Freeform optics, 200, 201, 204
Freehand interaction, 244
Free-space reflective combiner, 197–199
FreiHand dataset, 93
Front-run simulation, 883
Fugl-Meyer Assessment (FMA), 748
Fugl-Meyer score, 754
Full-color imaging, 201
Fundamental matrix, 63
Fun environments, 386
Fun game experience, 383
Furness, Tom, 619
Future work, 848, 849

G
Game AI, 139, 143, 146, 148, 153
Game based learning (GBL), 373
Game-changing technology, 203
Game engine, 132, 435

Game engineering, 132, 133, 149
Gameplay, 419, 420, 424
GamePlay Experience (GX), 375
GANerated hand dataset, 93
GANerated Hands, 101
Gaps, 833
Gartner, 614
Gauss-Newton minimization, 64
Gaze, 889
  interaction, 244
Gaze-based interaction, 114
Gaze-contingent displays, 114
Generative adversarial networks (GANs), 101, 166
Genetic algorithm, 260
Geo-referencing, 313
Geo-Spatial Navigation Sensors, 59
Geometry-based virtual acoustic space (GbVAS) algorithm, 784
Germane cognitive load, 776
Gesture(s), 889
  recognition, 748, 893
Gesture-based controls as part of the UI, 638
Gesture-based interaction, 112–113
Global positioning system (GPS), 59, 499
Globalization, 308
Goal oriented behavior (GOB), 139, 149, 151
Google, 800
Google Assistant, 801
Google Cardboard, 306
Google Glass, 196, 370, 786
Google Play Console platform, 510
GoogleNet, 67
Google's ARCore, 131
Graceful degradation, 287
Gradient descent, 68
Graphic intelligence, 322
Graphic user interfaces (GUIs), 407, 817
Graphics processing units, 217
Gravity, 217
Grip stability, 755
Ground-based augmentation systems (GBAS), 59
Guided manufacturing, 646
GunnAR, 627
Gyroscopes, 57
Gyro sensor, 218

H

Hand–eye calibration, 214
Hand-held device (HHD), 244, 649, 661
Hand held displays, 585
Hand-object interaction, 93
Hand pose estimation
  BigHands dataset, 92
  CMU Panoptic hand dataset, 92
  early datasets, 92
Hand tracking, 113, 230, 747, 748
Hands-free interaction, 244
Haptic, 805
  augmentation, 583
  feedback, 114
  interaction, 114–115
Haptic User Gerät (HUG), 893
Hardware, 40–44, 138, 143, 187–188, 203, 211–233, 241, 244, 246, 252–253, 343, 429, 540, 584–585, 618, 620, 651
Head mounted display (HMD), 10, 45, 110, 113, 132, 212, 213, 243, 306, 585, 598, 609, 661, 670, 798
Head-up display (HUD), 45, 473, 474, 476, 617, 618, 620
  aircraft pilot, 619
  dismounted personnel, 620
  ship captain, 619
Head-worn display, 204, 814
Healthcare, 820, 821
Health monitoring, 884
Hearing disorders, 783
Helmet-mounted display (HMD), 474
Helmet-mounted visualization, 188
Heritage education, 307
Heterogeneous methods of protection, 175
Heterogeneous sensing management framework, 172
Heuristic measure, 166
HIBEY, 242, 253, 265–267
  application scenarios, 258
  character keys, 254
  gesture of pointing hand, 254
  implementation and user performance, 256–258
  keyboard-less environment, uncertainty on, 254–256
  probabilistic method for handling imprecision, 256
  system design, 253–254
Hidden area visualization, 647
Hierarchical FSMs, 149
Hierarchical task networks (HTN), 134
High dynamic range (HDR), 8, 10
High-level representation, 80
High-performance computing (HPC), 884
History of the SWIM, 23–24
HO-3D dataset, 93
Hole Peg Test (HPT), 754
Holistic protection, 175
Holocaust education
  anti-Semitic sentiments, 421
  applications, 427–429
  challenge, 416
  classroom, 415
  cultural differences, 414
  cultural values, 414
  digital reconstruction, 424–427
  digital technologies, 414
  educational computer applications, 413
  ethics, 414
  freedom of speech, 416
  institutions, 417
  Jewish community, 415
  learning, 415
  learning institutions, 414
  lifestyle requirements, 415
  memorials, 422–424
  multimedia content, 417–420
  political spectrum, 415
  public-school approach, 414
  structures, 421
  sympathy, 416
  teaching, 415
  technology-enhanced learning, 413
  3D character model, 420
  tolerance, 416
  types of school, 414
  virtual environment, 414
  visualization, 421
Holographic displays, 394
Holographic processing unit (HPU), 44
HoloLearn, 773
HoloLens, 159

HoloMoL, 774
Home automation, 814
Homography matrix, 62
Homography pose estimation, 62
HOnnotate, 103
HOPS-Net, 102
Horizontal situation display (HSD), 477
HSPA+, 282
HTTP, 281, 293, 551
Hue-saturation-value (HSV), 544
Human-centred collaboration, 861
Human computer interaction (HCI), 240, 246, 535, 746, 832–834, 840, 847
Human factors, 618, 619, 621, 627, 631, 636, 638, 832, 834, 841, 847
Human Factors Engineering Technical Advisory Group (HFE-TAG), 636
Human-human communication, 140, 141
Humanistic intelligence (HI), 17
Human-machine collaboration, 861
Human memory
  AR applications in research of, 771–776
  augmentation systems, 771
  augmentation, AR applications, 772
  components, 769
  process of indexing the information, 772
  working memory capacity, 770
Human-object-relationship (HOR), 835, 837, 840, 841
Human organisational factors (HOF), 835, 837, 840, 841
Human pose tracking, 228–230
Humans-in-the-loop (HitL), 833, 841
Human space flights, 893
Human-system interface, 628, 636
Human visual system, 189, 342
  accommodation, 190
  cornea, 190
  eye resolution, 190
  FOV, 190
  fovea, 190
  image formation, 190
  optical system, 190
Human-wearable symbiosis, 240
Hybrid approach, 38
Hybrid display, 116
Hybrid reality, 807
Hybrid Synthetic Vision, 42
Hybrid trackers, 72
Hybrid XRI objects, 843
Hype Cycle forecasts, 614
Hyperelastic model, 698
Hyperreality, 799

I
Identifiability, 170
Identity privacy, 164
Identity spoofing, 170
Image distortion, 55, 192
Image-guided therapy, 690
Image markers, 888
  misalignment, 888
Image processing techniques, 582
Image pyramids, 61
Image recognition, 279
Image segmentation, 249
Imaging optics, 196
i-MareCulture project, 316
Imbalance of power, 170
Immersive adaptive environments, 848
Immersive AR, 38
Immersiveness, 392
Immersive Semi-Autonomous Aerial Command System (ISAACS), 480
Immersive technologies, 884
In-air gestures, 112, 113
Induction coils, 226
Industrial 4.0, 802
Industrial Internet of Things (IIoT), 600, 644, 648–651
Industrial requirements, 661
Industry 4.0, 599, 600, 644, 646, 648, 650, 651, 831, 853
Inertial approaches, 583
Inertial measurement units (IMUs), 213, 217, 218, 226
  bias, 218
  IMU–camera calibration, 219
  sensor fusion, 218–219
  sensors, 767
Inertial navigation, 58
Inertial sensors, 57
  accelerometers, 58
  gyroscopes, 57
  inertial navigation, 58
Inertial tracking, 37
Inflight, 488
In-Flight Entertainment and Communication (IFEC), 489
Influence, 820
Informal education, 308
Information abstractions, 174
Information and communications technologies (ICT), 497
Information and Computer Technology (ICT), 383
Information filter, 623, 637
Information overload, 623, 637, 638
Information Surveyed Point for Observation and Tracking (InfoSPOTS), 498
Information-of-interest, 158
Infrared-based systems, 438
Infrastructure, 798
Initialization, 72, 75
Inspections, 897
Instagram, 715
Institutionalization, 309
Integrate data from multiple sources, 637
Integrated Helmet and Display Sighting System (IHADSS), 188
Integrity, 168
Intelligent, 817
  agents, 801
  AR assistant, 821
Intelligent virtual agents (IVAs), 132, 135, 143, 801, 803–805
Intensity constraint, 61
Interaction, 391, 393, 397, 398, 404, 406, 407, 584, 817, 889
  controllers, 889
  design, and AR-IoT framework, 808–816
  metaphors, 890
  paradigm, 604
  rotation and translation, 889
Interaction-based methods, 221–222
Interaction-free calibration, 222–223
Interaction-free display calibration (INDICA), 222, 224
Interaction techniques for mixed reality (MR), *see* Mixed reality interaction techniques
Interactive darkroom, 11
Interactive design, 438

Interactive engine (iEngine), 604
  applications, 604
  architecture, 605, 606
  areas, 605
  benefits, 604
  decision making layer, 605
  framework, 604
  game engine, 604
  hardware layer, 605
  low-level software layer, 605
  medium level software layer, 605
  module architecture
    modules, 606, 607
    package, 608
    plugins, 608
    services, 608
  MultiModule, 605
  novelty, 608
  on site support phase, 608
  three-dimensional virtual objects, 604
  training phase, 608
  user layer, 605
Interactivity, 819
Interdisciplinary, 798
Interfaces, 805, 806
Interference, 653
Intermediary protection layer, 175
International Council on Systems Engineering (INCOSE), 882, 884
International Network for Contemporary Performing Arts, 434
International organisations, 309
International Space Station (ISS), 486
Internet, 358
Internet of Things (IoT), 408, 598, 798, 799, 801–802, 806, 831–833, 836, 839, 841, 843, 848, 849
Internet Protocol smart Objects (IPSO), 801
Interoperability, 808, 885
Interpupillary distance, 203
Interventional radiology, 689
Intrinsic cognitive load, 776
Intrinsic parameters, 54
Intrinsic policy enforcement, 162
Intuitiveness, 814
Inverse transformation, 54
iOS, 44, 376, 505
IoT-Avatars, 843
Iqbal's method, 100
iTacitus, 394
Iterative closest point (ICP), 66, 216–217

J

Jacobian matrix, 69
Jaguar Land Rovers project, 789
Japan Airline, 485
Jewish Day schools, 414
Jitter, 275, 289
Jurassic World Alive, 711
Just-in-time (JIT), 498

K

Keyboard-less text entry, 243, 254
Keyframes, 70, 76
Key performance indicators (KPI), 891
Keypoint based SLAM, 76–78
Kinect, 435, 436, 438, 442–445, 447, 449
Kinetic energy, 341
Knowledge-based Augmented Reality for Maintenance Assistance (KARMA), 41
Knowledge transfer, 701

L

Laboratory application, 857, 868
Lambertian scattering, 195
Laparoscopic adrenalectomy, 691
Laparoscopic surgery, 690, 694, 701
Laser beam scanning system, 195
Laser Scanning, 425, 426
Laser scanning system, 188
Latency, 267, 620, 811
  trade-offs, 173
Layar, 43
LeapMotion, 428
Learned depth, 79
Learning, 649, 700
  active learning process, 371
  activities, 343
  applications, 371, 372
  AR technology, 369, 372
  AR-supported mobile applications, 371
  characteristics, 371
  classroom, 371
  Computational Thinking, 372
  devices, 371
  different domains, 372
  digital games, 372
  digital information, 369
  dynamic processes, 371
  education, 371
  educational learning approaches, 373, 374
  environments, 386
  experimentation, 371
  films, 373
  gameplay, 376
  immersion, 372
  interaction experiences, 370
  location-based AR, 371
  marker-based AR, 371
  markerless AR, 371
  mathematics applications, 372
  mobile platforms, 371
  outcomes, 359, 360, 362
  pedagogical contexts, 372
  platforms/resources, 372, 373
  real-time interaction, 372
  serious game, 371
  tangible interfaces, 372
  tangible programming blocks, 372
  technology requirements, 372
  usability, 372
  video games/digital games industry, 373
  virtual environments, 369
Learning motivation, 356–361, 363–365
  scale, 361
Least privilege, 163
Legislation, 174
Levenberg-Marquardt optimization, 63
Lifecycle, 856
  creation phase, 887
  disposal phase, 887
  operation phase, 887
  perspective, 864

phases, 887
production phase, 887
Lifeplus, 393–394
Light Detection and Ranging (LiDAR) technology, 716
Light-emitting diodes (LED), 194
Light engine, 193
Lighting conditions, 624
Light propagation, 201
Lightpack, 44
Lightwear, 44
Likert scale, 816
Line tracking, 64
Linear elastic law, 698
Linear elasticity, 699
LineMOD dataset, 91
Liquid crystal on a silicon (LCoS) display, 194
Living Augmented Reality (LAR), 505
Locality sensitive hashing, 295
Location based positioning systems, 338
Location detection, 343, 344
Location-responsive services, 343
Lock-in amplifier (LIA), 12, 13, 24, 27
Long short-term memory (LSTM), 79, 103
Long-term memorization, 321
Long-term memory, 321, 769
Loop closing, 77
Loosely coupled, 78
Loss recovery, 289
LSD-SLAM, 77
LTE, 282, 296
Lumilo, 805

M

Mach drills, 361
Machine learning (ML), 47, 60, 66, 75, 77, 80, 90, 769, 837
MagicBooks, 341
Magnetic fields, 341
Magnetic tracking, 37
Magnetometer, 57
Maintenance and repair, 620–621, 628–631, 637, 638
authoring, 628, 629
Maintenance requirement cards (MRCs), 629
Maintenance, repair and overhaul (MRO), 896, 897
operations, 670
Management software, 648
Mann Continuum, 11
MANO model, 101, 102
Manual Augmented Reality (MAR), 504
Manufacturing, 646, 883
Manufacturing execution systems (MES), 882
Mapping, 54, 55, 65, 69, 73, 75, 77, 78
March, AR project, 394
Marine Corps, 624
Marine training application, 625–626
Marker-based AR, 401, 582, 664, 812
Markerless AR, 401
Markerless tracking, 37, 228–230
algorithm, 748
methodologies, 582
Marker trackers, 69
Market requirements, 866
Mask-RCNN, 103
Massachusetts Institute of Technology (MIT), 499
MATLAB AR Drone Simulink Development Kit, 480
Mauristhius museum, 315
Mechanical links, 225
Mechanical tracking, 37
methods, 582
MediaPipe, 101
Mediated reality, 10, 370, 800
continuum, 11
Mediated virtuality, 370
Mediating conflicts, 169
Medical application, augmented reality, 232
imaging and anatomical modeling, 693–694
physics-based modeling, 697–700
registration and pose estimation, 696–697
surgical vision, 694–696
visualization and perception, 700
Medical imaging, 689
Medical system, 812
Mental workload, 890
Merged Reality, 370
Messenger, 715
Metacognition, 327
Metaform optics, 205
Metasurfaces, 204, 205
Method of Loci (MOL), 774
Microdisplays, AR
aberration correction, 193
color filtering, 194
focal length, 193
FOV, 193
imaging optics, 193
self-emit, 193
superluminescent LED, 194
Microelectromechanical systems (MEMS) technology, 37
Microsoft, 800
Microsoft Cortana, 801
Microsoft HoloLens, 138, 159, 338, 478, 480, 485–487, 647–649, 657, 659–661, 767, 773, 784, 887, 889, 893
Microsoft HoloLens 2, 653, 657, 767, 768
Microsoft Kinect, 59
Microsoft's Minecraft Earth, 711
Microsoft's Teams, 347
Microsurgery, 702
Mid-air, 253
click, 249
interaction, 113, 242, 248–249
Middle of Life (MOL), 860
Milgram continuum, 11
Milgram's taxonomy, 834
Military maintenance, 620–621
Military operations, 627–628
Military training, 622, 626, 637
Mini Bluetooth Adapter, 751
Minimally-invasive therapies, 699
Mini Mental State Examination (MMSE) scores, 774
Mirror World, 807
Mission-critical applications, 172
Mission requirements, 891
Mixed reality (MR), 36, 157, 158, 176, 337, 345, 348, 370, 393, 710, 832–833, 836, 839–841, 843
continuum, 7
data manipulation and transformations, 163–167
device protection, 170–171
encryption-based techniques, 164
environments, 306, 604
facial de-identification, 164
feed-through signalling, 169–170
humans, 118
input protection, 162–167

Mixed reality (MR) (*cont.*)
mobility, 159
open research challenges, 171–173
output protection, 167–168
output reliability and user safety, 167–168
protecting collaborative interactions, 169–170
protecting device access, 170
protecting external displays, 168
protecting physical interfaces, 170
protecting sharing initialization, 170
protecting user interactions, 168–170
secure multi-party visual processing, 164
security and privacy risks, 159–161
system, 481
3D data attacks and protection, 166
3D spatial data transformations, 166–167
visual information abstraction, 163
visual information access control, 163
visual information sanitization, 162
Mixed reality interaction techniques
gaze-based interaction, 114
gesture-based interaction, 112–113
haptic interaction, 114–115
keyboard and mouse, 117–118
multi-display interaction, 116
multimodal interaction, 115–116
pen-based interaction, 113–114
tangible and surface-based interaction, 110–112
virtual agents, 118
Mnemonic strategies, 776
Mobile AR, 620
Mobile Augmented Reality Application (MARA) project, 43
Mobile Augmented Reality for Space Operation Procedures (MARSOP), 486
Mobile computer, 619
Mobile devices, 240, 360, 447
Mobile internet devices (MIDs), 670
Mobile learning, 358
Mobile phones, 42
Mobile platforms, 382
Mobile use, 624
Mobile users, 621
Mobile VR, 113
Mobility, 625, 626, 637
Modality, 119
Model-based instruction (MBI), 484
Model based systems engineering (MBSE), 882, 884–886, 892, 893, 897, 898
Model-based tracking, 69
3D model trackers, 70–73
marker trackers, 69–70
non-rigid and articulated objects, 74
Modeling, 637
Modelview matrix, 55
Momentum, 341
Monocular-based approaches, 695
Monolithic imaging system, 200
MonoSLAM, 76
Motion capture, 435, 438, 446, 448
Motion-to-photon latency, 274, 294, 296
Motion tracking, 265
Motor skill, 361
learning, 357, 360
Movement, 133, 134, 136, 138, 139, 145–147
layer, 140
MPEG-TS, 281
MRI, 693, 696, 701
Multi-display interaction, 116
Multi-factor authentication (MFA), 170
Multi-physics modeling, 884
Multi-sensorial AR experiences, 403–404
Multi-user coordination, 169
Multi-user MR environment, 160
Multiculturalism, 308
Multidimensional multimediated reality, 19, 20
Multimedia, 417
information, 649
Multimediated reality, 11, 17–18, 20, 21, 29, 800
continuum, 21–23
darkroom, 11–17, 21
measuring instruments, 17
multidimensional, 19, 20
multisensory synthetic synesthesia, 19
photographic darkrooms, 11
sensory attenuation technologies, 11
Multimodal, 318, 805, 818
interaction, 115–116
languages, 325
Multipath, 281, 289, 290
Multiple external representations (MERs), 780
Multiple hypotheses, 71
Multiplexed holograms, 199
Multisensor integration, 212
Multisensory output, 115
Multisensory synthetic synesthesia, 19
Mummification process, 315
Museums, AR
exhibition, 314
physiognomy, 315
sensible organism, 315
Music & graphics, 321
Mutuality, 169

N

Naish, J. Michael, 618
NASA-TLX scale, 779
NASA's Langley Research Centre, 478, 479
NASA's Jet Propulsion Laboratory (NASA JPL), 486
National Aeronautics and Space Administration (NASA), 478, 884
National Holocaust Centre and Museum, 421
National Operative Programme (NOP), 312
National Science Foundation, 802
Natural features, 228
Natural interfaces, 244
Naval Air Warfare Center – Aircraft Division (NAWC-AD), 630
Naval domains, AR
augmented ship transits, for improved decision-making, 627
battlespace visualization, 628
color perception, 632–637
dismounted infantry training, 621–626
GunnAR, 627
human factors, 631
ISEAotF, 628
local maintenance, 629
Ocean AR, 628
rapid prototypes, 631
remote maintenance aid, 630–631
SMART, 628
vestibular therapy, 631
virtual scientist, 628

Naval Information Warfare Systems Command (NIWC), 627
Naval Research Laboratory, 621–623
Naval Undersea Warfare Center Division, Keyport (NUWCKP), 628
Navigation mesh, 146
Navigation system, 692
Navy Health Research Center (NHRC), 631
Near-eye displays, 172
Nearest neighbor search, 173
Network protocol, 280
Network training, 67
  back-propagation algorithm, 68
  epochs, 68
  loss function, 67
  optimization, 68
  training data, 67
Neural networks, 307
Neuro motor index (NMI), 781
NeuroPage, 772
Neurosurgery, 695
NeverMind, 774
NGRAIN, 485
Nine Hole Peg Test (NHPT), 754
Non-cyborgs, 18
Nondeterministic finite automata, 612
Nonlinear optimization, 62
Non player character (NPC), 132
Non-repudiation, 168
Non-rigid object tracking, 74
Numerous scholastic experiments, 311
NyARToolkit, 379
NYU dataset, 92

O

Oberweger's method, 96
Object alignment, 888
Object-centric, 798
  AR-IoT data management, 810–812
Objective measurements, 223–224
Object level, 80
  abstraction, 163
  permissions, 163
Object pose estimation, 54, 72
  LineMOD dataset, 91
  Occluded LineMOD dataset, 91
  T-LESS dataset, 91
  YCB-Video dataset, 91
Object tracking, 71, 72, 74, 80
ObMan dataset, 93, 102
Occluded LineMOD dataset, 91
Ocean AR, 628
Off-hand posture, 265
Off-the-shelf controllers, 113
Office of Naval Research, 623–624
Office of the Vice Chancellor for Research and Graduate Education, 435
Offloading, 161
OmniTouch, 112
On-body interaction, 244
On-device interaction, 244
On-orbit servicing (OOS), 893–895
Ontology, 174
  and taxonomy of objects, 171
Open Concurrent Design Tool (OCDT), 885
Open source, 659, 660
  framework, 648
OpenCV, 295
OpenDX, 544
OpenGL, 541
OpenGL Utility Toolkit (GLUT), 545
OpenPose, 229
OpenXR, 768
Operation
  airplane, 884
  interface, 360
  methodology, 319
Optical architectures, AR
  aberration correction, 199
  birdbath architecture, 200
  bug-eye, 198
  coaxial eyepiece, 196
  flat beamsplitter, 196
  FOV, 197
  freeform optics, 200
  light engine, 197
  microdisplay, 198
  nose mirror, 198
  reimaging, 201
Optical components, AR
  combiner, 193
  light engine, 193
  microdisplays, 193
Optical design
  arcmin resolution, 193
  eyepieces, 190
  FOV, 190
  imaging system, 190
  pupil swim, 192
  screen-door effect, 193
  virtual image, 191
Optical fibers, 226
Optical flow, 295
  estimation, 61–62
Optical head-mounted displays (OHMDs), 243
Optical see-through (OST), 400, 632, 633, 635, 814
Optical see-through head-mounted displays (OST-HMD), 110, 116, 212, 221
  eye tracking, 223
  interaction-based methods, 221–222
  interaction-free calibration, 222–223
Optical strategies, 168, 170
Optical tracking, 37
  approach, 582
  sensors, 749
Optimal eye clearance, 191
Optimization, 260
OptiTrack motion system, 751
OptiTrack motion tracking system, 760
ORB, 295
ORB-SLAM, 77
Organic light-emitting diodes (OLED), 193
Original SWIM, 23
Out-coupler, 202
Out-of-band, 170
Outliers, 63
Output access control framework, 168
Output confidentiality, 168
Output rendering priority, 168
Overfitting, 67
Overlay, 893
  offset, 889

P
Packet transaction type analysis, 161
Painting
 shipbuilding, 645
Parameters of a tracking system, 36
Passenger Service Unit (PSU), 489
Passive haptics, 114
Passive markers, 227
Pathfinding algorithms, 146
Pedagogical framing
 activity oriented educational applications, 339
 audio annotation, 339
 collaborative aspects, 339
 content knowledge, 338
 content orientation, 339
 definition, 339
 education and training settings, 338
 experiences, 338
 learner activity impact, 339
 location, 339, 340
 materials, 338
 motivation, 338
 pedagogical theory, 339
 resource impact, 339
 role focused affordances, 341
 task-oriented, 341
Pedagogy, 415
Pen-based interaction, 113–114
Perceived utility, 174
Perception, 622, 626, 632–635
Perception impairments
 AR-enhanced applications in research of, 783–788
 types of, 782
Perception related cognitive impairments, 781–783
Perceptual and sensory threats, 158
Perceptual challenges, 618, 619
Performances, 866
Peripheral vision loss (PVL), 786
Persistence-of-exposure, 11
Personal area network (PAN), 163
Personal digital assistant (PDA), 42, 113, 771
Perspective camera model, 54
Pervasive, 801
Pervasive computing, *see* Ubiquitous computing (UbiComp)
Phase-coherent, 25
Phases of system lifecycle, 864
Phenomenality, 22
Phenomenological augmented reality (PAR), 4, 6, 22–29
Phenomenological reality, 12, 13
Philosophy of Cinema, 323
Phonagnosia, 783
Photogrammetry, 425, 426
Photometric error, 64
Photometric stereo, 694
Photometry, 195
Photonics Chip, 44
Phygital museum, 325
Physical education (PE)
 application, 358, 359
 AR-assisted learning system, 359, 360, 362, 363
 audiovisual media, 356
 augmented reality (AR), 356
 challenges, 357, 358
 computer multimedia materials, 356
 conventional method, 358
 efficient and innovative teaching method, 357
 immersive learning, 356
 information technology (IT), 355
 instructors, 355, 357
 Internet, 355
 Laban movement analysis, 357
 learning motivation, 364–366
 learning outcomes, 357, 358
 motor skill acquisition, 363, 364
 motor skills, 355
 practical skills, 356
 real and virtual domains, 356
 research methods, 360, 361, 363
 skill demonstrations, 355
 teaching, 357
 teaching academic knowledge, 356
 teaching styles, 357
 techniques, 355
 three-dimensional (3D) space, 356
 3D dynamic model, 356
 3D motor skills, 356
 types of teaching, 356
 verification, 360, 361, 363
 video-assisted learning, 357
 video learning model, 356
 virtuality–reality integration, 357
 visual and auditory information, 356
 visualization techniques, 356
Physical markers, 382
Physical objects in a virtual space, 854
Physical-virtual interactivity, 819
Physics-based modeling, 697–700
Physiological attacks, 167
Physiological cognitive load estimation, 777
Pie charts, 891
Piezoelectric sensors, 896
Pinterest, 716
Pix2Pose, 97
Pixel density, 343
Pixel mapping, 220
Pixelated display, 194
Place-dependent augmented reality, 340
Planar waveguide, 202
Plate beamsplitter combiner, 196
Platform-as-a-service (PAAS), 842
Platonic solids, 378
Player, 374
Playful AR technology activities, 387
Point-feature-based alignment, 217
Point-feature histograms (PFHs), 217
Pokémon GO, 43, 712, 734, 800
Policy compliance, 168
Policy enforcement, 174
Policy-makers, 174
Poro-hyperelastic model, 698
Pose estimation and tracking fundamentals, 53
 cameras, 54–57
 coordinate system transformations, 54
 depth sensors, 58–59
 6DoF pose of rigid bodies, 54
 Geo-Spatial Navigation Sensors, 59
 inertial sensors, 57–58
Pose refinement, 72
PoseCNN, 95
Positional tracking, 36–38
Post-capture protection, 175
Postmemory, 423, 425

Postures, 112
Practical DT implementation, 868
Pre-conditioned conjugate gradient (PCG) solver, 539, 565
Pre-integration, 78
Predict the physical world, 855
Predicted behaviours of manufacturing entities, 855
Predictive maintenance, 490, 647
Prefuse, 544
Preoperative MR images, 691
Presence, 340, 834
Primary Flight Display (PFD), 475
Priority class, 287, 288
Prisma, 394
Prism-air interface, 200
Privacy, 819
  approaches, 158
  guarantee, 166
  lamps, 169
  leakage, 173
  preservation, 160
Privacy-leaning, 175
Private Interaction Panels (PIPs), 170
Problem-solving, 374
Processing, 379
Product data management (PDM), 898
Product design process, 861
Product lifecycle, 882, 883
  creation phase, 883
  disposal phase, 883
  operation phase, 883
  production phase, 883
Product lifecycle management (PLM), 882, 885, 898
Professional judgement, 311
Progressive markers, 320
Project MOON, 484
Project OnSight, 487, 488
Project Sidekick, 486
Projection matrix, 56
Projection-based AR, 110, 394, 396
  sensing, 230–231
Projector–camera calibration, 220
Projector–camera system, 112, 220
Proof-of-concept, 815
Proprioception, 265
Props, 110
Prosopagnosia, 782
ProtoSpace, 487
Psychometric tests, 777
Pupil-center corneal-reflections (PCCR), 223
Pupil-matched eyepiece, 191
Pupil replication, 196
Pupil-steering, 204
PVNet, 96

Q
QoS classes, 286
Quality control, 646
Quality of experience (QoE), 174
Quantitative validation process, 675
QUIC, 281

R
Radial distortion, 192
Radio frequency identification technology (RFID), 231, 801
Radiometric brightness analysis
  broadband laser, 195
  environment brightness, 194
  eye-tracking, 196
  light source, 194
  luminance, 194
  narrow spectrum, 196
  OLEDs, 195
  retina, 195
  speckle, 196
Radio signals, 226
Random decision forests, 90
Random sample consensus (RANSAC), 63
Range of Motion (ROM), 744
RangeDB, 885
Rapid prototyping, 631
Reach-to-grasp exercise, 754
Reachability, 887
Realism of focused types of training, 626
Reality, 22
Reality-virtuality continuum model, 136, 157, 306, 370, 834, 840, 844
Real object training phase, 756
Real-time computation, 699
Real-time finite element (FE) modeling, 537, 538
Real-time location systems, 646
Reasoning layer, 140
Rectified linear unit (ReLu), 67
Reflective beam splitters, 201
Reflective markers, 751
Registration, 694, 696, 697
  accuracy, 888
  error, 619, 620, 623
Registration methods, 216
  ICP, 216–217
  point-feature-based alignment, 217
Reinforcement learning, 168
Relationships, 840
Religion, 416
Relocalization, 76
Rembrandt reality, 315
Remembrance agent, 771
Reminiscent, 201
Remote, 814
  maintenance, 577
Remote assistance (RA), 669
Remote collaboration
  asynchronous, 897
  and maintenance, 896–897
Remote procedure call (RPC) protocol, 551
Rendering process, 36
Reprojection error, 62
Research methodology
  artifact evaluation, 503, 504
  constructive research, 501
  design, 502, 503
  development, 502, 503
  DSR method, 503
  generalization, 505
  identification, 502
  learning, 505
  NASA TLX protocol, 502
ResNet, 67
Respiratory motion, 692
Retina, 190
Retina Scanning Displays (RSD), 585
Retroreflective materials, 227
RGB-D SLAM, 78
Rigid link, 225

Risk(s)
- analysis, 173
- assessment, 175
- with MR data processing, 159–160

Risk analysis and risk reduction (RA/RR) processes, 437
Robotic(s), 893
- applications, 233

Rolling shutter, 168
- effects, 57

Royal Navy Buccaneer, 474
RSVP, 280
RTCP, 280
RTP, 280
RTSP, 281
Ruggedized, 653

S

Sensorama, 40
Safe limited acceleration (SLA), 437
Safe limited position (SLP), 437
Safe limited speed (SLS), 437
Safety, 444, 447, 448
- and danger warning, 172
- requirements, 661

Sanitization, 162
SARS-CoV-2, 710
Satellite-based augmentation systems (SBAS), 59
Satellites, 343
Satisfaction, 816
Saturday science, 445
Scalability, 77, 799
Scalable AR-IoT recognition and tracking, 812–813
Scale, 625
Scale invariant point features (SIPFs), 217
Scanning, 70
Scanning ambiguous keyboard (SAK), 247
Scene revelation, 166
Screen real estate, 240
SCTP, 281
SDR SWIM setup, 25
Seamless, 822
Secret sharing, 164
Secure multi-party computation (SMC), 164
Security, 804
Security and privacy risks, mixed reality
- mobility and privacy, 160
- MR data processing, 159–160

Security-leaning, 175
See-through AR displays, 394
See-through display, 618, 620, 622, 624, 633
- *vs.* video, 622

See-through head-mounted display, 624
Segmentation-driven method, 96
Self-emissive microdisplay, 193
Self-supervised training, 104
Semantic(s), 885
- understanding, 79

Semi-automated forces (SAF), 621
Semi-direct approach, 77
Sense-making, 842
Sense-Think-Act cycle, 147
Sensing systems, 438
Sensor(s), 585
- data, 692
- fusion, 218–219

Sensor-based tracking, 744
Sensor integration, tracking systems, 224
- camera-based motion capture systems, 227–228
- electromagnetic sensors, 226
- flex sensors, 226
- IMUs, 226
- markerless tracking, 228–230
- mechanical links, 225
- projection-based sensing, 230–231
- radio signals, 226

Sensor-level management framework, 172
Sensory attenuation, 11
Sensory issues, 343
Sensory memory, 769
Sensory threats, 158
Sequential Wave Imprinting Machine (SWIM), 15, 23, 30
- and electric machines, 27, 29
- history of, 23–24
- principle of operation, 24
- visualizing radiowaves with SWIM and SDR, 25–27

Serious games (SG), 373
- augmented reality (AR), 375
- designs, 374
- types, 374
- user experiences, 374, 375

Service Maintenance Augmented Reality Tools (SMART), 628
Shared environment, 170
Shared information, 645
Shared ontology, 175
Shared space, 168
Sharing control module, 172
Shipbuilding, AR/MR
- advanced communications architectures, 651
- advanced industrial software, easy interaction with , 648
- AR/MR devices for shipbuilding applications, 653–657
- challenges, 661, 664
- commercial developments, 645
- communications, 647
- design and construction assistance, 645
- guided manufacturing, 646
- hidden area visualization, 647
- ideal AR/MR device characteristics, 651–653
- IIoT devices, monitoring and interaction with, 648
- painting, 645
- predictive maintenance, 647
- product and tool maintenance, 649
- product and tool tracking, 646
- quality control, 646
- shared information, 645
- software AR/MR frameworks, 658–660
- step-by-step guidance, 645
- structure visualization, 648
- traditional AR/MR communications architectures, 649–650
- training, 648–649
- warehouse management, 647
- welding, 644

Shipyard 4.0, 644
Shoulder-surfing attacks, 168
Side-channel attacks, 170
Siloing, organizational, 883
Simulated reality, 799
Simulation, 383, 422, 490
Simultaneous localization and mapping (SLAM), 74, 86, 118, 227, 813
- algorithm, 480
- deep learning, 79–80
- RGB-D SLAM, 78

visual SLAM, 75–77
visual-inertial SLAM, 77–78
Single-point active-alignment method (SPAAM), 221, 222
Singular value decomposition (SVD), 216
Sitting wave, 24
Situation awareness (SA), 476, 619, 621, 622, 625, 627, 628, 636
collaborative, 621
Six degree-of-freedom (6DOF), 36, 499
Sketching, 113
Skills, 374
Small and medium-sized enterprises (SMEs), 857
Smart, 802
buildings, 802
cities, 802
connected environments, 802
factories, 802
glasses, 137
homes, 802, 806
manufacturing, 651
objects, 802
spaces, 836, 840, 841
wearables, 241
Smart Swim digital glasses, 767
Smartglasses, 240, 241, 244–247
Smartphone, 115, 240, 451, 456
Snapchat, 319, 714, 715
Social cognition, 403
Social competency development, 311
Social connections, 804
Social influence, 822
Social interaction, 136
Social media networks, 714
Facebook, 715
future, 716, 717
Instagram, 715
Messenger, 715
Pinterest, 716
precautions and risks, 716
snapchat, 714, 715
TikTok, 715, 716
Social perception, 140, 144
Social presence, 132, 133, 135, 140, 143
Social realism, 820
Social responsibility, 309
Socio-technical system, 832
Soft tissue biomechanics, 697–699
Software-defined radio (SDR), 25–27
Software Development Kits (SDKs), 131, 382, 658, 659
Solids, 376
Sound language, 321
Sound spatialization, 406
Spacecraft, 881, 885
disposal, 885
operation, 885
phases, 885
preliminary design, 885, 891
production phase, 885, 886
reviews, 885
Space debris, 893
Space environment, 893
Space missions, 885, 893
Space-multiplexed approach, 110
Space operations, 486, 487
Sparse connectivity, 66
Spatial, 803
calibration, 220
division code, 231
gestures, 112
inference attacks, 173
mapping, 158
reasoning, 619
scanning, 231
Spatial user interface (SUI), 768
Speech, 889
and gestures, 115
Spherical camera model, 57
Spine surgery, 691
Spoof, 168
Sport-related technologies, 358
Spreadsheet interaction, 113
SSD-6D, 95
Standing waves, 24
Static decision making, 770
Statistical decoding, 256
Steering behaviors, 145
STEP, 885
Step-by-step guidance, 645
Stereo video display, 620, 621
Stereoscopy, 892
Stochastic and geometric measurements, 173
Stochastic gradient descent (SGD), 68
Storytelling, 400, 404, 405, 407
Strain simulation, 897
Strapdown inertial navigation, 58
Strategic level, 866
Structural health monitoring (SHM), 896
Structural life prediction, 884
Structure from motion (SfM), 65
bundle adjustment, 66
triangulation, 65
Structured light, 694
Student-centred learning, 317
Students, 375
Subjective measurements, 224
Super Cockpit, 619
Supervised training, 91
Surface dyslexia, 782
Surface Movement Radar (SMR), 481
Surface-to-plane generalizations, 166
Surgical navigation, 691, 693
Surgical stereoscopy, 695
Surgical vision, 694–696
Sustainable supply chains, 873
Sutherland, Ivan E., 618
SWIMulator™, 29
SwipeZone, 246
The Sword of Damocles, 40, 188
Synergistic approach, 175
Synergy stage, 760
Synesthesia, 19
Synthetic data, 67
Synthetic reality, 799
Synthetic synesthesia, 19
Systematic literature review (SLR)
AECO activities, 500
AR applications, 499
artifacts categorization, 500–502
DSR method, 499
primary studies, 499
protocol, 500
quantitative analysis, 499

Systematic literature review (SLR) (*cont.*)
quantitative meta-analysis, 499
search strategies, 499
search strategy, 499
sensors and markerless methods, 499
temporal distribution, 501
temporal evolution, 499
tracking methods, 499
tracking types, 501
urban installations, 500
Systematic security and privacy analysis, 173
System-level, 175
System model, 886
synchronization, 893
System Modeling Language (SysML), 885
Systems engineering, 884

T
Tablets, 451
Tamper, 168
Tangible and surface-based interaction, 110–112
Tangible augmented reality (TAR), 110
Tangible interface, 102, 379
Tangible magic lenses, 110
Tangible programming blocks, 379
Tangible user interfaces (TUIs), 110
Tap the artwork (ARTAP), 324
Target acquisition, 240–244, 247, 248, 265
Task augmentation, 341
Taxonomy, 834, 835, 884
of seven realities, 30
of X-reality, 9
XRI, 834–841
TCP, 281, 288
Teaching-learning paradigms, 374
Technical training, 701
Technical University of Munich, 478
Television, 419
Templates, 72
Temporal analysis, 173
Text entry, 242, 265
on constrained screen real estate, 246
Theatre, 433, 434, 441, 443, 446–449
Theatrical performance
accessibility, 447, 448
actor, 433
actors, 438, 440, 446
agency, 435
ALICE Project, 435, 448
application, 449
audience, 446
audience interaction, 435, 440
body and gesture tracking, 434
2D space, 439
challenge, 439
chicken game, 441, 442
community, 434
data, 438
data considerations, 445
design process, 438
digital projections, 449
film, 433
flying system, 440
hologram, 434
human action/interaction, 434
human movement, 435
immersion, 446, 447
interactive elements, 441
interactive performance research, 446
interactive production, 440
live performance events, 434
machine learning models, 441
minimum requirements, 433
motion capture, 446
motion pictures, 440
non-linear storytelling, 440, 446
parameters, 439
penguin game, 442
performance, 436
performance methodology, 449
performers, 438
public engagement, 445
public interaction, 443, 444
recognition systems, 434
requirements, 449
safety, 436, 437, 449
sensor, 438
sensor considerations, 444
simulation, 441
space, 434
stage-based interaction, 434
storytelling affordances, 443
storytelling environment, 449
swan game, 443
technical components, 435
technology, 435
technology's capabilities, 438
Unity game engine, 440, 441
video game engine, 446, 449
virtual actors, 434
virtual agents, 448
virtual object, 440
young audiences, 440, 441
Thermal analysis, 886
Thermal tracking, 230
Third space, digital environments
collaborative construction, 318
digital artefacts, 318
intelligent space, 318
physical environment, 318
Three-dimensions (3D), 376, 648
audio, 631
avatars, 897
data, 160, 166, 173, 174
dynamic model, 360
graphic content, 406
modeling, 628, 637
movements, 115
orientation estimation, 57
pose estimation, 103
rigid motion, 87
scan, 112
sculptures, 319
virtual character, 361
visual teaching materials, 363
visualization, 378
wireframe model, 813
3D hand pose estimation, 98
A2J, 99–100
DeepPrior++, 99
from depth maps, 98–100
GANerated Hands, 101

Iqbal's method, 100
manipulating virtual objects, 102
MANO model, 101
MediaPipe, 101
from RGB image, 100–102
V2V-PoseNet, 99
Zimmerman's method, 100
3D model(s), 891
requirement, 88
retrieval, 98
technology, 359
3D model trackers, 70
deep learning-based trackers, 72
direct trackers, 71
edge trackers, 71
feature trackers, 70
hybrid trackers, 72
object model acquisition, 70
tracking by detection / pose initialization, 72
3D object(s), 244
selection, 116
3D object pose estimation
Augmented Autoencoders, 96
BB8, 93–95
DeepIM, 96
DPOD and Pix2Pose, 97
Oberweger's method, 96
PoseCNN, 95
PVNet, 96
segmentation-driven method, 96
SSD-6D, 95
YOLO-6D, 95
3D model requirement, 88
ambiguities, 88
changing illumination conditions, 88
cluttered background, 88
early approaches and limits, 89–90
for object categories, 97–98
high degrees of freedom, 87
material and textures, 88
occlusions, 88
3D spatial data
risks with, 159, 160
transformations, 166–167
user perception, 160
3D tracking, 813
methods, 89
Thumb-to-finger space interaction, 260
Thumb-to-finger text entry, 242
TikTok, 715, 716
Time-of-flight, 694
Tip tool calibration, 213
TiPoint, 242, 248
application scenarios, 252–253
fast repetitive mode, 249
freehand mode, 248, 249
implementation, 249–252
mid-air interaction strategy, for small-screen display, 249
system requirements, 248
system-wide implementation, 252
T-LESS dataset, 91
TOFI, 243, 258, 265, 266
application scenarios, 265
keyboard layout, optimization of, 260–264
system implementation, 259
user performance, 264–265
Toolbox, 604
Total internal reflection (TIR) prism, 200
Touch, 112
Touchscreen, 240, 241, 243
Touring Machine, 620
Tracking, 36, 747, 805
devices, 890
eye tracking, 891
finger tracking, 891
hand tracking, 893
methods, 89
Tracking systems, 602
for sensor integration, 224–231
Traditional AR/MR communications architectures, 649–650
Traditional museum, 318
Training, 623, 638, 648, 700
military, 621, 623–626, 637, 638
military, evaluation, 624
realism, 625, 626
set, 91
Transreality, 799, 802, 807, 808, 818, 820, 822
Transreality interaction platform (TrIP), 807
Trends, 394
Triangulation, 65
Trust, 818
Trustworthiness, 161, 174
Tukey estimator, 64
TUM Medical Education Center (TUM MEC), 137
Two-dimensional eyebox expansion, 202
2D interfaces, 244
2D optical flow, 61
2D physical markers, 382

U

Ubiquitous computing (UbiComp), 241, 798, 800, 802
and artificial intelligence, 803–804
and augmented reality, 805–806
Ubiquitous virtual reality (U-VR), 807
UDP, 291
UI design, 836, 840, 841
Ultimate display, 40, 799
Ultrasound, 690, 694, 887, 896
Unauthorized access, 170
Underwater Augmented Reality (UWAR), 398
Undigital, 23
Unforeseen privacy risks, 160
Unified modeling language (UML), 885
Unintentionally modified reality, 10
Unity, 376, 660
Unity 3D, 132, 360, 607
Universal mCloud company, 485
Unmanned aerial vehicle (UAV), 479, 480
Unmodified mobile devices, 113
Unmodified physical keyboards, 117
Unreal Engine, 132
Upper-extremity, 743, 744
US Air Force, 41
US Air Force's Super Cockpit program, 474
Usability, 803, 814, 816
User acceptance, 620
User-centered design, 622, 627, 628, 638
User experience (UX), 375, 392, 401, 404, 834, 840, 847
User-identifiable information, 175
User interaction, 584
User interface, 136, 631
gesture, 638

User manuals, 887
User Markup Manual Language (UMML), 814
User-perceived sensitivity, 160
User-perspective rendering, 112
User privacy preferences, 175
User study(ies), 619, 621, 633, 891
Utility, 151–153, 166

V
Vampire mirrors, 169
Vascular surgery, 690, 691, 701
Venn diagram depicting the various realities, 29
Verbal communication, 142, 143
Vergence-accommodation conflict (VAC), 203, 400
Verification and validation (V&V), 884
Vertical situation display (VSD), 477
Vestibular therapy, 631
VGGNet, 67
Vibrotactile, 805
Video-assisted instruction (VAI), 356
Video game engine, 435, 436, 441, 446
Video learning, 359, 362
Videoplace, 41
Video see-through (VST), 110, 138, 400, 814
Virtual agents, 118, 801, 803, 817, 818
Virtual animals, 820
Virtual assistant, 603, 801
Virtual avatars, 799, 802, 817, 818, 822
Virtual cloth try-on, 164
Virtual collaboration, 628, 630
Virtual continuum, 41
Virtual digital factory, 861
Virtual dog, 820
Virtual elements, 375
Virtual environments, 136, 387, 447, 892
Virtual fixtures, 41
Virtual force-feedback, 631
Virtual Human Director (VHD), 133
Virtual humans, 132–135, 143, 144, 819
Virtual lighting, 401–402
Virtual object training phase, 756
Virtual-physical interaction, 857
Virtual private region, 170
Virtual reality (VR), 7, 35, 111, 113–115, 117, 118, 187, 306, 345, 370, 451, 534, 597, 644, 744, 798–800, 832–834, 836, 837, 840, 841, 886
Virtual reality environment (VRE), 604, 609
Virtual reality markup language (VRML), 536
Virtual reconstruction, 166, 314
Virtual Satellite, 885, 891, 892
Virtual screen space, 116
Virtual spacecraft design (VSD), 885
Virtual technologies, for cultural heritage legibility, 391–393
Virtual training environment (VTE) system, 609
  application
    decision trees, 612
    finite state machines, 612
    result, 612
  HMD, 609
  iEngine modules, 610, 611
  platform supports, 609
  software architecture, 609
  use, 609
  workflow, 609
Virtual transparency, 112
Virtuality, 22
  continuum, 800
Virtuality–reality integration, 359
Virtually enhanced senses (VES), 784
Visco-hyperelastic model, 698
Visible light-based stereo systems, 438
Visible spectrum, 195
Vision-based trackers, 539
Visual and haptic multimediated reality, 13
Visual augmentation, 172, 587
Visual channel, 168
Visual cryptographic techniques, 170
Visual cues, 170
Visual display, 337
Visual elements, 344
Visual-inertial fusion, 78
Visual-inertial SLAM, 77–78
Visual information access control, 163
Visual information protection, 163
Visual inspection, 889
Visual instrument design, human visual system, 189
Visual instruments, 321
Visual intelligence, 321
Visualization, 36, 420, 421, 423, 424, 429, 883
Visualization Toolkit (VTK), 535
Visually-Coupled System (VCS), 40
Visual markers, 344
Visual perception disorders, 782
Visual positioning service (VPS), 160
Visual positioning systems, 338
Visual resolution, 192
Visual secret sharing (VSS), 168
Visual SLAM, 75
  direct SLAM, 77
  keypoint based SLAM, 76–78
  SLAM architecture/main components, 76
Visual studio, 44
Voice, 784, 889
  assistance, 820
  commands, 240
  recognition technology, 448
Voice-based home assistant prototypes, 803
Voice-based interaction, 659
Voice user interface (VUI), 407
vtkActors, 545
vtkCamera, 545
vtkRenderer, 545
Vuforia, 44, 382, 659, 660
  platform, 376
Vuforia Developer Portal Target Manager, 509
V2V-PoseNet, 99

W
Ward's Method, 461
Warehouse management, 647
Warfighter Augmented Reality (WAR), 624
Warning systems, 172
Wave propagation, 4, 14, 17, 24
Waveguide architecture, 201, 202
Wearable, 173, 817
  computers, 240–242
  display, 619
  projector-depth-camera system, 112
  remembrance agents, 771
  sensors, 748
Wearable augmented reality (WEAR), 486, 773, 776, 780, 783

Welding, shipbuilding, 644
Wi-Fi, 226, 282
  802.11ac, 282
  802.11ax, 284
  802.11n, 282
  networks, 313
  Wi-Fi Ad-Hoc, 283
  Wi-Fi Direct, 283
Wide Area Augmentation System (WAAS), 59
Wikitude AR SDK, 44
Wikitude World Browser, 43
Windows, icons, menus, and pointers (WIMP), 242
Wireless communications, 661
Wireless localization, 170
Wireless sensor network (WSN), 539, 812
Wisconsin Alumni Research Foundation (WARF), 435
Wisconsin Institute for Discovery (WID), 443, 445
Wolf Motor Function Test (WMFT), 748
Working memory, 769–770
World-, object-, head-, body- /device-referenced, 116

X
X-ray, 694, 896
X-reality (XR), 8, 833, 834
XML, 551
XR-IoT (XRI), 833, 834, 848
  CoRe frameworks, 841–847
  multi-disciplinary taxonomy for, 834–841
Xsens MVN motion capture system, 359
XY-reality, 9

Y
YCB-Video dataset, 91
YOLO-6D, 95
Young audience, 442, 443

Z
ZFNet, 67
Zimmerman's method, 100
Zoom, 347