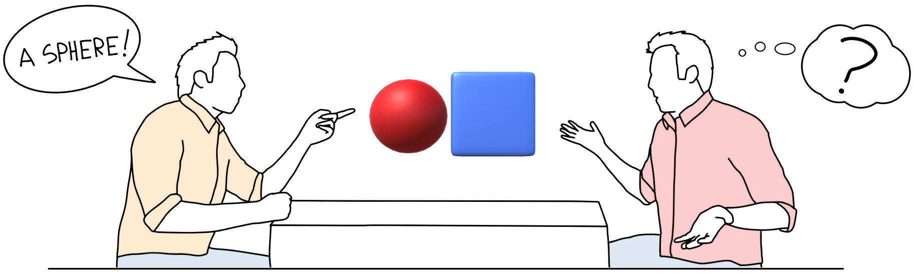
UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO



Perception Manipulation for Seamless Face-to-face Remote Collaboration

António Maurício Lança Tavares de Sousa

Supervisor: Doctor Joaquim Armando Pires Jorge

Thesis approved in public session to obtain the PhD Degree in Information Systems and Computer Engineering

Jury final classification: Pass with Distinction and Honour

2020

TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Perception Manipulation for Seamless Face-to-face Remote Collaboration

António Maurício Lança Tavares de Sousa

Supervisor: Doctor Joaquim Armando Pires Jorge

Thesis approved in public session to obtain the PhD Degree in Information Systems and Computer Engineering

Jury final classification: Pass with Distinction and Honour

Jury

Chairperson: Doctor João Emílio Segurado Pavão Martins, Instituto Superior Técnico, Universidade de Lisboa
Members of the Committee:
Doctor Anthony Steed, Faculty of Engineering Sciences, University College London, UK
Doctor Joaquim Armando Pires Jorge, Instituto Superior Técnico, Universidade de Lisboa
Doctor Pedro Filipe Pereira Campos, Faculdade de Ciências Exactas e da Engenharia, Universidade da Madeira
Doctor Carlos António Roque Martinho, Instituto Superior Técnico, Universidade de Lisboa
Doctor José Miguel de Oliveira Monteiro Sales Dias, recognized individuality in the scientific area of the thesis.

2020

Abstract

This thesis aims to improve remote collaboration in shared 3D workspaces. Current mixed reality technologies allow geographically distant collaborators to be together and share the same virtual space, making it possible for people to see each other through realistic virtual representations. Face-to-face telepresence also promotes a sense of presence and can improve collaboration by allowing the immediate understanding of nonverbal cues. Indeed, several approaches have successfully explored face-to-face remote interactions with 2D content. However, when collaborating in a 3D object-centered volumetric workspace, there is a decrease in awareness due to gesture ambiguities, occlusions, and different participants' viewpoints. In this dissertation, we contribute the use of perception manipulation to improve workspace awareness in computer-supported collaborative work in mixed reality telepresence environments by assuring that remote collaborators are always aware of what is happening in the workspace when communicating using nonverbal cues. We began by contributing the technological foundations to prototype remote interactions. And then, we proposed and evaluated perception manipulation techniques focused on allowing remote people always to share the same understanding of the workspace. And, at the same time, being aware of nonverbal communication. Results suggest that by purposefully changing the properties of the person-task space using geometric transformations, warping, and repositioning devices, we can counteract gesture ambiguities, eliminate workspace occlusions, and promote a shared understanding of the workspace. In conclusion, we have validated our thesis, stating that perception manipulation techniques increase workspace awareness and improve face-to-face remote collaboration in mixed reality 3D workspaces.

Resumo

Esta tese visa melhorar a colaboração remota em espaços de trabalho 3D compartilhados. As atuais tecnologias de realidade mista permitem que colaboradores geograficamente distantes estejam juntos e compartilhem o mesmo espaço virtual, possibilitando que estes se vejam por meio de representações virtuais realistas. A telepresença frente-a-frente promove uma sensação de presença e pode melhorar a colaboração, permitindo a compreensão imediata de sinais não-verbais. De fato, várias abordagens exploraram com sucesso interações remotas frente-a-frente com conteúdo virtual 2D. No entanto, ao colaborar num espaço de trabalho volumétrico com objetos 3D, existe uma diminuição na percepção do espaço devido a ambiguidades de gestos, oclusões, e devido também aos pontos de vista dos participantes. Nesta dissertação, contribuímos técnicas de manipulação da percepção para melhorar a consciência do espaço de trabalho na colaboração em ambientes de telepresença de realidade mista, garantindo que os vários colaboradores remotos estejam sempre cientes do que acontece no espaço de trabalho quando comunicam usando gestos. Começamos por contribuir as bases tecnológicas para prototipar interações remotas. De seguida, propusemos e avaliamos técnicas de manipulação da percepção focadas em permitir que pessoas remotas consigam compartilhar o mesmo entendimento do espaço de trabalho. E, ao mesmo tempo, estarem cientes de toda a comunicação não-verbal. Os resultados sugerem que, alterando intencionalmente as propriedades do espaço da pessoa-tarefa usando transformações geométricas, deformações e reposicionamentos, pode-se neutralizar as ambiguidades dos gestos, eliminar as oclusões e promover um entendimento comum do espaço de trabalho. Concluindo, nós validamos a nossa tese e conseguimos afirmar que as técnicas de manipulação da percepção aumentam a conhecimento do espaço de trabalho e melhoram a colaboração remota frente-a-frente em espaços de trabalho 3D de realidade mista.

Keywords

Remote Collaboration
Workspace Awareness
Nonverbal Communication
Perception Manipulation
Mixed Reality

Palavras Chave

Colaboração Remota
Consciência do Espaço de Trabalho
Communicação Não Verbal
Manipulação da Percepção
Realidade Mista

Publications

The work developed during this dissertation produced multiple publications accepted in top tier peer-reviewed journals and conferences. Here, we highlight the relevant publications listed in chronological order by date of release from oldest to newer.

Main Publications

Materials, ideas, and figures from this dissertation have appeared previously in the following publications. Also, at the end of each chapter, we also provide a written indication of the corresponding publications.

[P6] Maurício Sousa, Daniel Mendes, Rafael Kuffner dos Anjos, Daniel Simões Lopes, and Joaquim Jorge. **Negative Space: Investigating Workspace Awareness in 3D Face-to-face Remote Collaboration.** ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI), 2019

[P5] Maurício Sousa, Rafael Kuffner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. **WARPING DEIXIS: Distorting Gestures to Enhance Collaboration.** In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 12 pages. DOI: https://doi.org/10.1145/3290605.3300838
★ **Featured in the 'Best of CHI 2019' event** by IndiaHCI, co-sponsored by ACM SIGCHI Asian Development Committee and HCI Professionals Association of India.

[P4] Maurício Sousa, Daniel Mendes, Rafael Kuffner dos Anjos, Daniel Simões Lopes, and Joaquim Jorge. **Investigating Workspace Awareness in 3D Face-to-Face Remote Collaboration.** In International Conference on Graphics and Interaction (ICGI 2018), Lisbon.
★ BEST POSTER AWARD

[P3] Maurício Sousa, Daniel Mendes, Rafael Kuffner dos Anjos, Daniel Medeiros, Alfredo Ferreira, Alberto Raposo, João Madeiras Pereira, and Joaquim Jorge. **Creepy Tracker Toolkit for Context-aware Interfaces.** In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17). ACM, New York, NY, USA, 191-200. DOI: https://doi.org/10.1145/3132272.3134113

[P2] Maurício Sousa, Daniel Mendes, Soraia Paulo, Nuno Matela, Joaquim Jorge, and Daniel Simões Lopes. **VRRRRoom: Virtual Reality for Radiologists in the Reading Room.** In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 4057-4062. DOI: https://doi.org/10.1145/3025453.3025566

[P1] Maurício Sousa, João Vieira, Daniel Medeiros, Artur Arsenio, and Joaquim Jorge. **SleeveAR: Augmented Reality for Rehabilitation using Realtime Feedback.** In Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16). ACM, New York, NY, USA, 175-185. DOI: https://doi.org/10.1145/2856767.2856773

Other Publications

During my Ph.D. programme, I collaborated with colleagues in other concurrent research projects. The resulting publications are listed below.

[P19] Maurício Sousa, Daniel Mendes, and Joaquim Jorge. **Safe Walking in VR.** ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI), 2019

[P18] Rafael Kuffner dos Anjos, Maurício Sousa, Daniel Medeiros, Daniel Mendes, Mark Billinghurst, Craig Anslow, and Joaquim Jorge. **Adventures in Hologram Space: Exploring the Design Space of Eye-to-eye Volumetric Projection-based Telepresence.** ACM Symposium on Virtual Reality Software and Technology (VRST), 2019

[P17] Ezequiel Zorzal, Maurício Sousa, Daniel Mendes, Rafael Kuffner dos Anjos, Soraia Paulo, Pedro Rodrigues, José Mendes, Vincent Delmas, Jean-Francois Uhl, José Mogorrón, Daniel Simões Lopes, and Joaquim Jorge. **Anatomy Studio: a Tool for Virtual Dissection Through Augmented 3D Reconstruction Sessions.** Computers & Graphics, 2019

[P16] Daniel Medeiros, Maurício Sousa, Alberto Raposo and Joaquim Jorge. **Magic Carpet: Interaction Fidelity for Flying in VR.** IEEE Transactions on Visualization and Computer Graphics (TVCG), 2019.

[P15] Daniel Mendes, Maurício Sousa, Rodrigo Lorena, Alfredo Ferreira, and Joaquim Jorge. **Using custom transformation axes for mid-air manipulation of 3D virtual objects.** In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17). ACM, New York, NY, USA, Article 27, 8 pages. DOI: https://doi.org/10.1145/3139131.3139157

[P14] Daniel Mendes, Daniel Medeiros, Maurício Sousa, Eduardo Cordeiro, Alfredo Ferreira and Joaquim Jorge. **Design and evaluation of novel out-of-reach selection techniques for VR using iterative refinement.** Computers & Graphics, 2017

[P13] Daniel Mendes, Daniel Medeiros, Eduardo Cordeiro, Maurício Sousa, Alfredo Ferreira and Joaquim Jorge. **PRECIOUS! Out-of-reach Selection using Iterative Refinement in VR.** IEEE Symposium on 3D User Interfaces (3DUI), 2017

[P12] Daniel Mendes, Daniel Medeiros, Maurício Sousa, Ricardo Ferreira, Alberto Barbosa Raposo, Alfredo Ferreira and Joaquim Jorge. **Mid-air Modeling with Boolean Operations in VR.** IEEE Symposium on 3D User Interfaces (3DUI), 2017

[P11] Luís Bruno, Maurício Sousa, Alfredo Ferreira, João Madeiras Pereira and Joaquim Jorge. **Hip-directed walking-in-place using a single depth camera.** International Journal of Human-Computer Studies (IJHCS), Elsevier, 2017

[P10] Maurício Sousa, Daniel Mendes, Daniel Medeiros, Alfredo Ferreira, João Madeiras Pereira and Joaquim Jorge. **Remote Proxemics.** Chapter in Collaboration Meets Interactive Spaces, Springer, 2016

[P9] Daniel Medeiros, Eduardo Cordeiro, Daniel Mendes, Maurício Sousa, Alberto Raposo, Alfredo Ferreira and Joaquim Jorge. **Effects of Speed and Transitions on Target-based Travel Techniques.** ACM Symposium on Virtual Reality Software and Technology (VRST), 2016

[P8] Daniel Medeiros, Maurício Sousa, Daniel Mendes, Alberto Raposo and Joaquim Jorge. **Perceiving Depth: Optical versus Video See-through.** ACM Symposium on Virtual Reality Software and Technology (VRST), 2016

[P7] Daniel Simões Lopes, Daniel Mendes, Maurício Sousa and Joaquim Jorge. **Expeditious Illustration of Layer-Cake Models On and Above a Tactile Surface.** Computers & Geosciences (in press), 2016

Acknowledgements

The research presented in this dissertation would have been an impossible attempt without the support and trust of many people. First, I would like to give a special thanks to my advisor Professor Joaquim Jorge. I am very grateful for all your support and guidance and for helping me become a scientist. I would like to thank the other members of my thesis committee: Pavão Martins, Anthony Steed, Pedro Campos, Carlos Martinho, and Miguel Sales Dias. I thank my colleagues and friends that worked hard next to me for all these years: Daniel Medeiros, Daniel Mendes, Rafael Kuffner dos Anjos, and Soraia Paulo. Next, I thank the VIMMI Group that welcomed me and provided a platform to develop my crazy ideas. I thank all my colleagues, professors, researchers, students, friends, and co-authors that supported me: Alberto Raposo, Alfredo Ferreira, Artur Arsenio, Bruno de Araújo, Craig Anslow, Daniel Gonçalves, Daniel Simões Lopes, Edmilson Rodrigues da Silva, Eduardo Cordeiro, Ezequiel Zorzal, Fernando Fonseca, Filipe Relvas, Francisco Venda, Hugo Nicolau, Inês Santos, João Frazão, João Guerreiro, João Madeiras Pereira, João Viera, Karan Singh, Luis Bruno, Mark Billinghurst, Paula Barrancos, Pedro Garcia da Silva, Pedro Lopes, Ricardo Ferreira, Rodrigo Lorena, Rodrigo ”Rogério” Pinheiro, Rui Prada, Sandra Gama, and Sandra Sá. Finally, I also would like to thank my family for all their support and help through all my academic life.

Research Acknowledgements

Three research projects partially supported the research presented in this thesis document. All projects addressed several challenges related to researching, designing, and evaluating novel interaction techniques in mixed reality 3D user interfaces using head-mounted displays, walls, tabletops, mobiles, large scale displays, and wearables:

- **Project TECTON 3D** focused on researching new multimodal interaction techniques suitable for architectural 3D modeling design and review tasks in mixed reality.
- **Project IT-MEDEX** studied new user interfaces and interactive experiences in workspaces where medical professionals are engaged in collaborative tasks around 3D medical images.
- **Project LARA** engaged in evaluating novel interaction techniques for minimally invasive laparoscopic surgery using augmented reality technologies.

“Your eyes can deceive you. Don’t trust them.”
– Master Obi-Wan Kenobi (0 BBY)

Contents

List of Figures

List of Tables

List of Video Figures

Additional Information for Video Figures

All video figures are encoded using the MPEG-4 H.264 codec and saved in a ".mp4" file format. The video figures' thumbnail images are hyperlinks for downloading the media file using a browser.

1
Introduction

In this dissertation, we present a new direction for improving workspace awareness in face-to-face remote collaboration with 3D digital content. Contemporary remote collaboration approaches are not yet capable of providing the necessary means for people to stay aware of one another when interacting in shared 3D workspaces. Awareness is a crucial component of collaboration but is inadequate when people do not share the same physical space. Therefore, our overarching long term goal is to enable people to seamlessly collaborate at a distance by allowing people to communicate remotely as quickly as when they are together.

Immediately after the debut of the television in the late 1920s, pundits and researchers envisioned a future where remote face-to-face communication would be indispensable in our daily lives. In the following decade, Bell Labs demonstrated its first attempt at a two-way television system [71]. At the same time, allusions to videoconferencing started to show up in the mainstream pop culture, particularly in Dick Tracy's comic strip, featuring a two-way wrist-worn television [59]. Later, as a consequence of shifting from analog technologies to digital electronics, combined with the dynamic evolution of digital computing, video-conferencing, and telepresence technologies, revolutionized the way people communicate in the present days. Indeed, video-conferencing and telepresence allow virtual encounters to take place and expedite the communication between multiple geographically separated people.

Videoconferencing and telepresence technologies are proven approaches in establishing collaborative virtual meetings between distributed teams of experts in multiple domains. After all, virtual meetings allow for considerable savings in time and resources. Modern technologies add unproductive layers of protocol to the flow of communication between remote participants, rendering interactions far from seamless. Despite being widely adopted, traditional video conferencing approaches still uphold partial viewing or down-scaled representations of remote people that curb the sense of "being there" [60]. These conventional approaches interfere with the ability to effectively take advantage of natural nonverbal communication devices such as

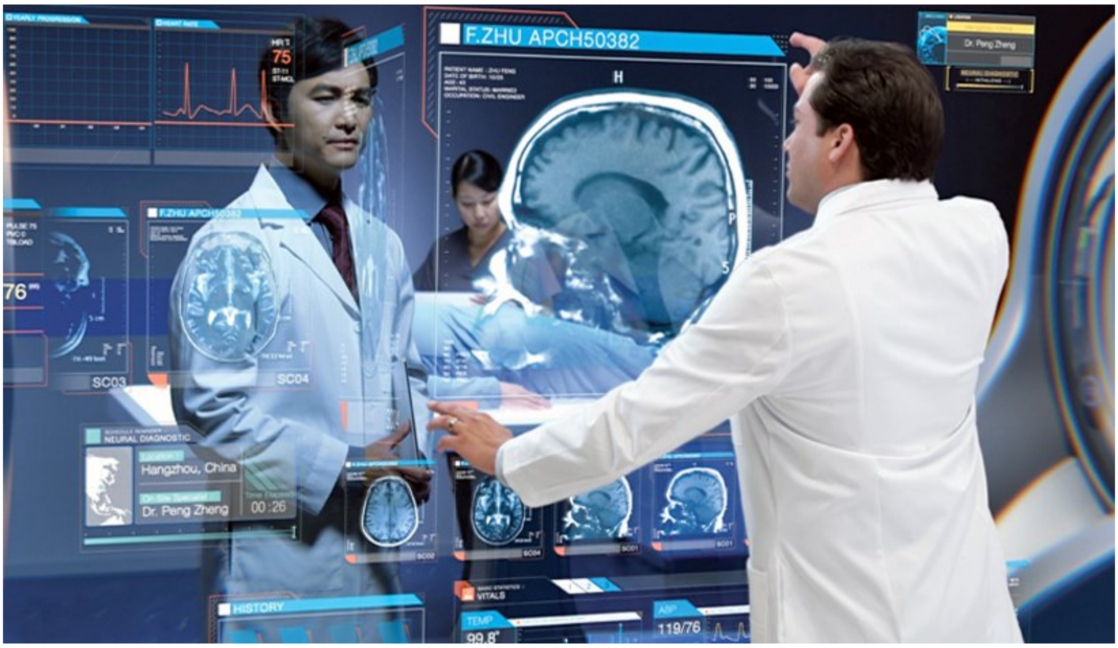


Figure 1.1: 'A day made of glass' by Corning imagines future face-to-face telepresence interactions close to a co-located experience.

gaze, body posture, and gestures that are fundamental for collaborating. Figure 1.1 shows Corning's[1] vision for future telepresence experiences.

Past research on telepresence suggests that full-body face-to-face interactions leverages the sense of co-presence and improves communication [68]. However, for remote collaborative tasks focused on virtual content, full-body face-to-face interactions over a shared workspace still entail that each participant detains a personal point-of-view of the task environment. If remote people share different perspectives of the workspace, they cannot maintain an up-to-the-minute understanding of the others' actions – *Workspace Awareness* [55] – and the collaboration turns into a coarse and disjointed experience. For this, in this dissertation, we propose new approaches focused on manipulating the individual perception of the shared environment to improve workspace awareness. And with this, improve remote collaboration between small groups of people by promoting the understanding of the task space and encourage the usage of natural face-to-face nonverbal communication.

1.1 Motivation

For a telepresence encounter to be closer to a co-located experience, the *person space*, identified by Bill Buxton [28] as "where you look when speaking to someone", should rely on full upper body portrayal of people to allow for the understanding of nonverbal visual cues in addition to normal speech currently supported by traditional videoconferencing approaches [29]. Nonverbal communicative cues include facial expressions, gaze, body posture, pointing gestures to indicate objects referred to in speech (*deictic gestures* [104]), or how people utilize the space and position themselves when communicating (*proxemics* [56]). Buxton also identified the *task space* as the "where the work appears" that can be private or shared between all remote participants, and the *reference space* as "space within which the remote party can use body language to reference the work." But, when designing for face-to-face collaboration, it is necessary to take into account how to address interactions in a shared workspace. Despite the person space being typically considered separate from the task space, Ishii et al. [68] suggest that both concepts should be integrated when considering face-to-face meet-

[1] Corning: https://www.corning.com

ings using a transparent display metaphor to maximize the perception of the other people's nonverbal cues. Indeed, with transparent displays, two videoconferencing participants can see one another and digital content, rendered between them, that can be jointly manipulated by both. In everyday face-to-face interactions mediated by displays, people have no common orientation of right or left, and therefore, a decrease in the awareness of what the other person is doing or talking about. This lack of awareness negatively affects the quality of the cooperation [166] because it constrains the ability for people to use descriptions of relative positions, either by speech or nonverbal cues, such as pointing gestures. Clearboard [68] addresses this issue by mirror-reversing the remote person's video stream, producing gaze and pointing awareness since 2D graphics and text can thus be corrected to the participant's point-of-view. This approach has been the subject of research for 2D content collaborative manipulation [164, 91, 166]. However, 3D digital content gives rise to multiple problems that affect and impair workspace awareness. Contrary points-of-view can result in different perceptions or even serious communication missteps. Participants do not share the same *forward-backwards* orientation, and occlusions can affect the understanding of where or what the remote person is pointing. Naturally, misunderstandings can lead to severe complications. For instance, in the medical community, the usage of digital imaging is commonly adopted since these virtual artifacts are easy to store, retrieve, and distribute. This issue profoundly impacts medical diagnosis, surgery planning, and provide education. 3D visualizations are essential during the analysis of the spatial relationships between anatomical elements and the surrounding structures. These spatial relationships are challenging to visualize in 2D [9]. However, in 3D, communication errors resulting from contrary points-of-view can cause disastrous consequences when health professionals collaborate to reach a delicate diagnosis or treatment plan since the analysis of 3D models is commonly used by the medical community.

1.2 Background

The activities of analysis, design, and review of 3D digital content are very appealing for engineering, architecture, and healthcare. In these fields, it is common for professionals to focus on highly detailed objects. Therefore the ability of every party to

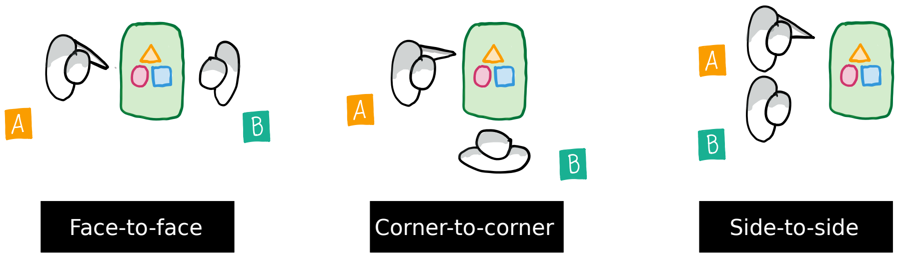


Figure 1.2: Example illustration depicting the occlusion issue present when people have opposing points-of-view.

experience a person-task space integration that preserves a non-obstructive reference space is of the utmost importance. At the same time, people should rely on natural communication, verbal and nonverbal, to convey the focus of the collaboration and pinpoint details on 3D content, as if they were physically face-to-face. Allowing people to produce and observe other people's nonverbal gestures encourages collaboration. Gestures that enhance collaboration can be *expressive*, *deictic* and *demonstrations*. *Expressive* gestures aid speech production and interpretation. The lack of these gestures impacts the sense of social presence negatively. *Deictic* gestures are natural and help refer objects or the task at hand. These gestures have an impact on task performance. Finally, *demonstrations* occur when people convey actions by performing gestures. Therefore, people need to be always aware of their colleagues' actions and the repercussions of those actions as they occur to avoid coordination errors and misunderstandings in communication.

Previous research regarding remote communication, collaboration, and awareness has addressed this issue. Yet, the challenges of face-to-face collaboration in a shared workspace for 3D digital content persist. Figure 1.2 provides a depiction on how people normally arrange themselves around a common shared task space, also known as *f-formations* [79]. While in a face-to-face formation, people can see each other at the same time they observe the 3D workspace between them. However, they are unable to share the same understanding of the workspace since they observe the space from different points-of-view and workspace artifacts can occlude one another. Oppositely, in a corner-to-corner and a side-to-side formation, the collaborators can adequately see each other and workspace. Yet, in these formations, the understanding of nonverbal gestures decreases due to the task space and personal space being separate. The continuous shift of attention between the workspace and what the other

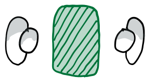
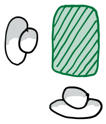
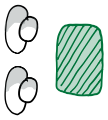
			
Properties	Face-to-face	Corner-to-corner	Side-to-side
Expressive Gestures	Yes	Limited	Limited
Deictic Gestures	Yes	Limited	Limited
Demonstrations	Yes	Limited	Limited
Gaze Awareness	Yes	Limited	Limited
Presence	Fitting	Limited	Limited
Workspace Awareness	Limited	Limited	Adequate

Table 1.1: Properties of collaborating with 3D Objects

person is gesturing can difficult nonverbal communication [68]. However, a side-to-side formation can foster workspace awareness because both participants share the same perspective of the task space, summarized in Table 1.1.

1.3 Research Statement

The main objective of this dissertation is to increase the benefits of full-body face-to-face visualization by adding the advantages of a side-to-side close encounter for remote collaboration. We argue that simulating a side-to-side behavior (sharing the same perspective) in a real-scale face-to-face collaborative environment to interact with 3D objects could improve the sense of presence and raises workspace awareness since it promotes the use of nonverbal communication. This can be advantageous to avoid communication breakdowns by making gestures and deictic idioms easier to share and understand between participants. To achieve a person-task space integration that enables such conditions we intend to broaden the perception manipulations, introduced by Ishii et al [68], by investigating manipulations of not only person-space but also, manipulations of task space, point-of-view and the positioning of remote people's representations using an "above a table" metaphor for a workspace. By perception manipulations, we mean to purposefully change the properties of the person-task space using geometric transformations, warping and repositioning devices, in a way that is not perceptible to participants or does not create an obstacle to the face-to-face collaboration.

The central problem that this dissertation address is to assure that remote collaborators are always aware of what is happening in the workspace when communicating using nonverbal cues. Therefore, the research statement addressed in this dissertation is:

> *Perception manipulation can be used to increase workspace awareness and improve face-to-face remote collaboration in shared 3D workspaces.*

Our research statement can be subdivided into three essential questions:

Question 1: Can two opposing collaborators share the same perspective of a shared 3D workspace?

While previous research has used a *WYSIWIS* [137] (what-you-see-is-what-I-see) approach in face-to-face collaboration in 2D workspaces, this approach does not scale when adding a third dimension. We believe that perception manipulation techniques can alter the characteristics of both person space and task space in a way that enables a more effective communication exchange. What is missing is a structural analysis to determine if unusually sharing the same perspective does not hinder natural communication.

Question 2: Can perception manipulations improve the understanding of nonverbal communication?

People use nonverbal cues when communicating. Also, the ability to perceive gestures and body language improves remote collaboration and enhances the feeling of being there. However, we do not know if body manipulations diminish the naturality of nonverbal communication or invalidate the readability of deictic gestures. We hypothesize that manipulating the way people are presented does not interfere with regular communication since it should not be noticeable.

Question 3: Can two opposing collaborators share the same understanding of a 3D workspace?

This dissertation aims at enabling people to collaborate using nonverbal communication in a face-to-face formation. For this, we propose the study of the integration of person space and task space that exploits perception manipulation techniques for collaborators to perceive equivalent individual reference spaces. Yet, our approach does not follow reality. And, despite allowing corrected versions of those individual reference spaces, we do not know if people can share the same understanding of the actions that can occur in the workspace.

1.4 Research Context

The context of this dissertation is illustrated in Figure 1.3. The research's background lies in the central area of *Human-Computer Interaction*, the intersection of social and computer sciences, consisting of how people and machines can work and communicate with each other. This research contributes more categorically to the HCI's branches of *Computer-supported Cooperative Work* – enabling the work of groups of people with computational technologies – and Also, we explore collaborative interactions between people in mixed reality environments with support for multiple electronic devices, while employing well established *3D User Interfaces* (3DUI) concepts and techniques. However, the theoretical background of this works lies in *Workspace Awareness*, since it is a fundamental component of effective collaborations. Workspace awareness can be defined as the understanding of other people's interaction in a shared a shared workspace [55].

The work in this dissertation builds on this background and extends remote collaboration with new concepts, insights, and approaches related to interaction design, prototyping, and user evaluations.

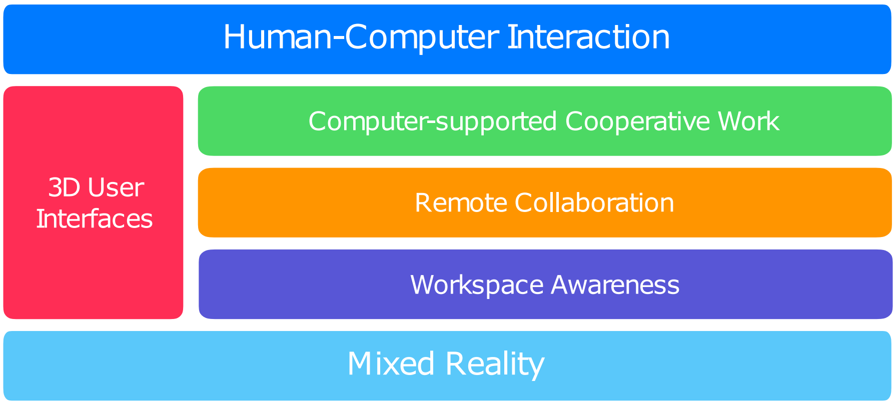


Figure 1.3: Research context overview.

1.5 Objectives and Methodology

The research present in this dissertation is targeted to answer the question of how to leverage face-to-face remote collaboration using perception manipulation. Therefore, we will focus on studying perception manipulation techniques in mixed reality environments with shared 3D workspaces. Furthermore, we will employ in all stages of this dissertation a user-centered methodology based on user studies to validate our research statement and answer the previously identified research questions.

We will meet the following objectives:

Objective 1: **We will define perception manipulations for face-to-face collaboration.**

We will identify and analyze seminal research on perception manipulation from literature in mixed reality environments, presence, and social psychology to outline how perception manipulation can be applied to remote collaboration. Accordingly, we will contribute perception manipulation techniques focused on enhancing workspace awareness and promoting the sense of "being there."

Objective 2: **We will design and implement rapid prototyping tools to make building remote interactions accessible.**

Towards this objective, we will develop the *Creepy Tracker Toolkit*. The toolkit will focus on providing the technological foundations to experiment and develop rapidly interactive experiences requiring non-invasive full-body tracking data and virtual representations of people. Furthermore, the toolkit must provide support for developing both co-located and remote interactions.

Objective 3: **We will evaluate workspace awareness using variations of the shared workspaces, individual point-of-view, and remote person's virtual representation.**

We will investigate multiple workspace conditions using different combinations of *personal space* and *task space*. The main goal is to determine the effects on people's perception of the *reference space* by manipulating the remote collaborators' point-of-view, remote embodiment, and the form the workspace should be presented.

Objective 4: **We will contribute body manipulation techniques to improve deictic gestures.**

We will develop and evaluate a body distortion approach to improve the perception of pointing gestures to communicate out-of-reach objects in virtual collaborative environments. The ability to correctly perceive pointing gestures facilitates collaboration by making the communication task more natural.

Objective 5: **We will contribute body manipulation techniques to improve close face-to-face collaboration.**

We will develop and evaluate a perception manipulation techniques to improve face-to-face collaboration in shared 3D workspaces by manipulating person space, task space and reference space in subtle manners. Furthermore, we will propose an integration of person space, task space, and reference space to minimize the need for remote collaborators to constantly switch attention between other collaborators' space and the workspace, thus improving collaboration.

1.6 Results

This research contributed original ideas, knowledge and practices to HCI, CSCW, and 3DUI. Next, we enumerate the major results of this research:

1. We contributed the theoretical grounds for using perception manipulation and illusions to improve remote collaboration in mixed reality environments. Our work differed from the state-of-the-art in the sense that instead of relying upon visual illusion to manipulate reality, our approach reshapes the way people's actions are presented to a local observer without losing the message collaborators want to express.

2. The Open Source Creepy Tracker Toolkit — a set of tools for rapid-prototyping context-aware applications, that incorporates body tracking, interactive surfaces, and point-cloud representation of people. Where we explored different scenarios that can be implemented using our toolkit and discussed practical considerations obtained from a system's performance evaluation. The Creepy Tracker toolkit served as the technological foundation for all the prototypes introduced in this dissertation.

3. An assessment of different manipulation techniques to improve workspace awareness in a face-to-face remote collaborative 3D workspace using different collaboration conditions. Resulting in the proposal of a new shared virtual workspace, called *Negative Space*, linking two remote physical spaces, while providing a sandbox for interacting with 3D content.

4. *Warping Deixis*, a novel body warping technique to improve how deictic gestures to distance targets are interpreted in mixed-reality environments. With Warping Deixis, we also contributed techniques to redirect arm poses applicable to different representations of virtual humans and a user study evaluating the impact of our approach in referent identification tasks. Results from a user study suggested that warping the pointer's arm can significantly reduce misunderstandings about the target's position.

5. *Altered Presence* as an interactive approach to integrating person space, task space, and reference space for face-to-face remote collaboration in mixed reality. Our Altered Presence approach ensures that opposing participants share the same perspective of a shared 3D workspace and distorts gestures performed by a remote person to present a corrected virtual representation of those gestures to the local participant using body warping. We also contributed implementation details on warping techniques to reshape the gestures on virtual representations of people, and user study evaluating the impact of our approach on workspace awareness. Results suggested improvements in awareness, presence, and interactions between remote collaborators in shared 3D workspaces.

1.7 Organizational Overview

This dissertation is structured into the three major parts – corresponding to the theoretical formulation of our approach, the technological infrastructures developed to carry out this research, and the experimental design used to acquire and report the created knowledge. Figure 1.4 shows a visual overview of the dissertation and the inter-relation between parts and chapters. Next, we introduce the components of this dissertation.

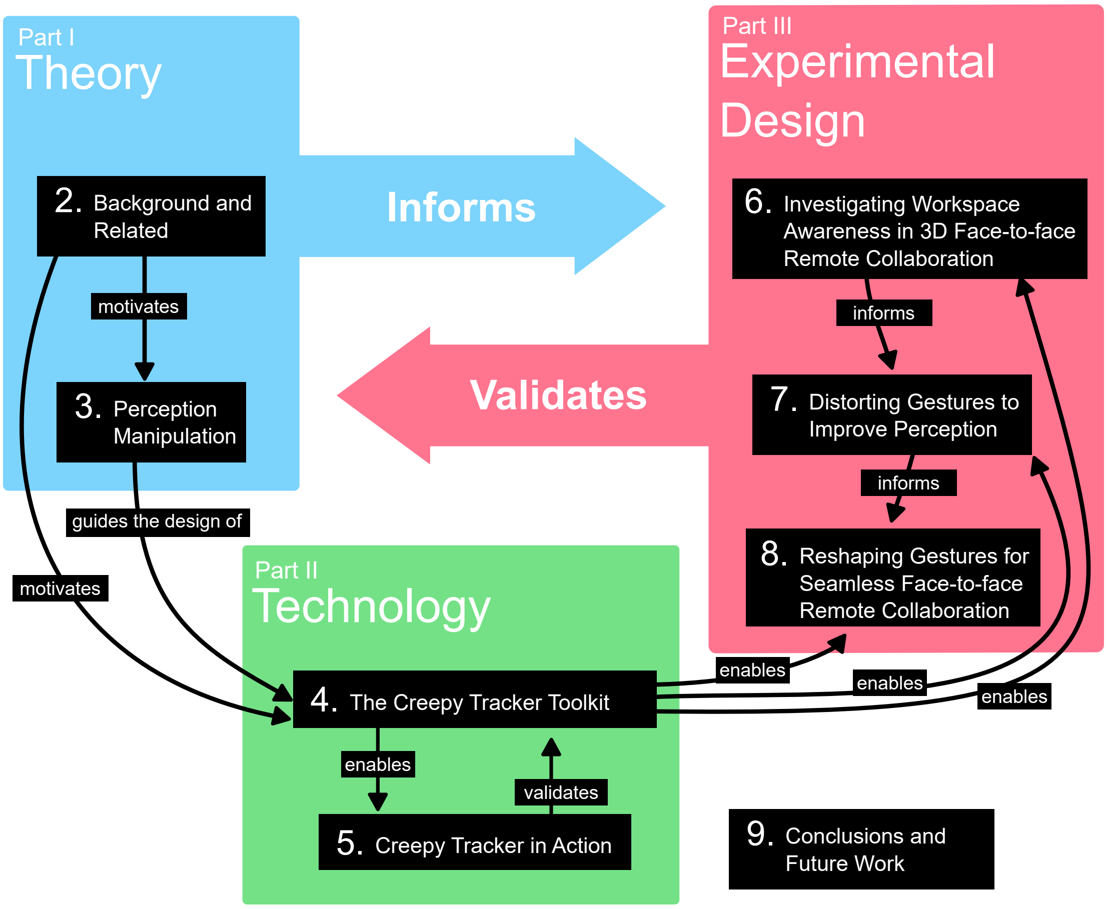

Part I
Theory
2. Background and Related
motivates
3. Perception Manipulation
Informs
Part III
Experimental Design
6. Investigating Workspace Awareness in 3D Face-to-face Remote Collaboration
informs
7. Distorting Gestures to Improve Perception
informs
8. Reshaping Gestures for Seamless Face-to-face Remote Collaboration
Validates
guides the design of
motivates
Part II
Technology
4. The Creepy Tracker Toolkit
enables
enables
enables
enables
validates
5. Creepy Tracker in Action
9. Conclusions and Future Work

Figure 1.4: Overview of the dissertation chapters and their inter-relation.

Part I: Perception Manipulation to Improve Collaboration

The first part of this dissertation covers the theoretical foundations of our approach and defines the concept of perception manipulations applied to remote collaboration to make it clear, measurable, and understandable by empirical experimentation.

In Chapter 2 we present an overview of the research field that forms the framework for this thesis. It introduces a survey of the state-of-the-art on workspace awareness, remote collaboration, interpretation of gestures, and virtual representations of people. In its discussion, we compare the presented related works to highlight the benefits and limitations that motivate our approach. In Chapter 3, we present our approach, and we hypothesize how can perception manipulation techniques be applied to remote collaboration to improve workspace awareness in face-to-face interactions.

🟥 *Part II : Prototyping Co-located and Remote User Experiences*

The second part of this dissertation details the technological tools developed to support the context-aware evaluation prototypes created to evaluate the co-located and remote user experiences. In Chapter 4, we[2] introduce the Creepy Tracker Toolkit, our approach to tracking people and capture virtual representations of people. We start by presenting the motivation, layout the toolkit's components, and describe the implementation details. Then, we disclose the results of the system's performance evaluation. Chapter 5 adds the description of how to use the Creepy Tracker Toolkit for prototyping context-aware user experiences. First, we introduce client-side developing details and then demonstrate the tracker's capabilities in five different application scenarios.

🟩 *Part III: Exploiting Perception Manipulation to Enhance Workspace Awareness*

The third and final part of this dissertation embodies the experimental approaches accomplished to validate the research statement.

In Chapter 6, we investigate how to maintain workspace awareness in a virtual volumetric workspace connecting two physical remote rooms. We present an evaluation comparing four different workspace conditions in which we varied reflections of both workspace and remote representations of people. The chapter ends with a discussion of the results. Chapter 7 introduces Warping Deixis, an approach aimed at improving the perception of deictic gestures in mixed reality collaborative scenarios using body warping. In this chapter, we detail our body warping approach, describe the evaluation procedure and methods, and discuss the results. Chapter 8 is the final chapter of Part III related to the experimental design. In this chapter, we present our perception manipulation approach to improving workspace awareness in remote face-to-face collaborative scenarios. Furthermore, we describe the user evaluation, detail the implementation of the evaluation prototype, and end with a discussion of the results.

[2] The use of the plural 'we' in Part II and Part III refers to Maurício Sousa, Joaquim Jorge, and the co-authors acknowledged at the end of each chapter.

This dissertation concludes in Chapter 9 with an overview of the presented research, a summary of the results and contributions, and pointers for future work. Finally, drawing inspiration from comic book trade paperbacks, I compiled a collection of doodles, sketches, and drawings made to assist me in solving problems and materialize ideas during this dissertation.

Part I

Perception Manipulation to Improve Collaboration

In this first part of the dissertation, we investigate the operationalization of perception manipulation for improving workspace awareness in face-to-face remote collaboration in shared 3D workspaces. First, in Chapter 2, we survey related work in remote collaboration, virtual representations of remote people, workspace awareness, and human interpretation of nonverbal communicative gestures. We also present, in Chapter 2, exploratory work that contributed with knowledge and experience to the main path of this thesis. Last, in Chapter 3, we introduce our approach and describe how to leverage perception manipulation techniques to improve awareness in face-to-face collaboration.

2
Background and Related Work

This chapter provides a part introduction of previous works related to this dissertation. The work developed in this dissertation builds on top of previous research concerning remote computer-aided collaborative work in shared workspaces. Loren Terveen [148] defines *collaboration* as the "process in which two or more agents work together to achieve shared goals". Terveen also considers that to achieve collaboration agents must: agree on the shared goal(s); allocate responsibility and coordinate with each other by determining what action should be accomplished; share a common context to continually evaluate the effects of their actions and decide whether or not they are pursuing a correct course of action; communicate with each other and observe each others' actions; and finally, adapt and learn with each other. These fundamental aspects of collaboration are not straightforward to achieve in remote settings where people are not present in the same physical space. However, there is a great body of previous research on creating remote collaborative experiences suggesting that staying aware of others is crucial to the fluidity and naturalness of collaboration [54].

We first introduce the key concepts and importance of workspace awareness in computer-aided collaborative work. Then, we cover previous research proposed in literature addressing virtual meetings and approaches to facilitate co-located encounters between remote people. We also discuss context-aware environments and sensing methods and the production and interpretation of deictic gestures in collaborative environments. The surveyed literature is followed by a discussion on the open challenges presented by the state-of-the-art and how it related to our proposed approach.

2.1 Workspace Awareness

Seminal research on collaborative writing [16, 43, 148] have identified the fundamentals of collaboration as *information sharing*, *knowledge of the group*, *individual activity* and *coordination*. Though, at the same time, Dourish and Bellotti [35] abstracted these fundamentals and propose the concept of *Awareness* as the *"understanding of the activities of others, which provides a context for your own activity"*. In simple terms, awareness is *"knowing what is going on"* [37]. More specifically, prior works suggest that awareness have these four characteristics [2, 37, 54, 109]:

- ➜ Awareness is knowledge about the state of an environment bounded in time and space.
- ➜ Environments change over time, so awareness is knowledge that must be maintained and kept up to date.
- ➜ People interact with and explore the environment, and the maintenance of awareness is accomplished through interaction.
- ➜ Awareness is a secondary goal in the task – that is, the overall goal is not simply to maintain awareness but to complete some task in the environment.

With this, Greenberg et al [48] identified four overlapping types of awareness that people naturally maintain in collaborative activities (Figure 2.1):

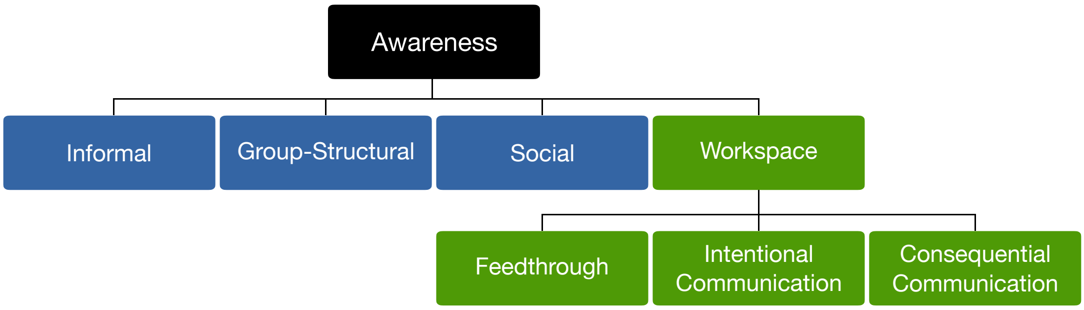


Figure 2.1: Taxonomy detailing the different types of awareness.

1. **Informal Awareness** – the knowledge of who is collaborating and what activities they are engage with;

2. **Group-structural Awareness** – the knowledge about people's roles, responsibilities, their positions on a issue, their status and group processes;

3. **Social Awareness** – the information that a person maintains about the others in a social context using nonverbal cues. This information can whether be if another person is paying attention, their level of interest or emotional state;

4. **Workspace Awareness** – the knowledge about the "who, what, where, when and why" questions that inform people about the change environment.

In Table A.1 and A.2 of the Appendix 1, we show the full list of elemental questions that inform workspace awareness. Workspace awareness relates to the contents and the immediate changes that occur in a shared workspace, it also relates to what actions people are doing. Thus, workspace awareness can be defined as the *"up-to-the-moment understanding of another person's interaction with a shared workspace"* [54]. Greenberg et al [48] suggest that workspace awareness is different from the other types of awareness because its integral role in collaboration. In a way that comprises the *identity* of those in the workspace, their *location*, their *activities* and the immediacy of *changes* with which others' activities are communicated. This real-time knowledge of another person's interactions and the effects on the workspace is essential to an effective collaboration. After all, working together causes people to undertake the additional task of maintaining the collaboration. When working alone, as depicted in Figure 2.2 (Left), people are solely focused on completing the domain tasks required to

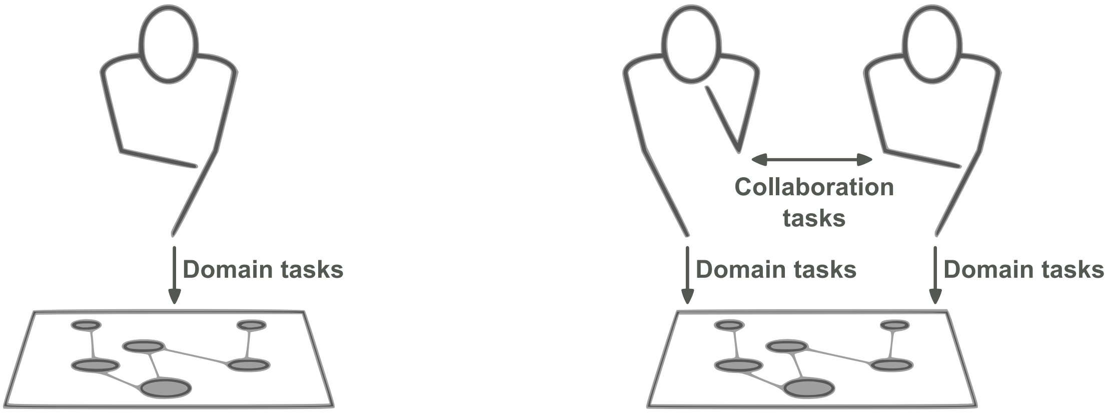


Figure 2.2: Domain and collaboration tasks (from Gutwin and Greenberg [54]).

achieve their goals. However, in a collaborative setting, meeting participants have to constantly carry out the collaboration tasks of communication and decision making, apart from their individual domain tasks, as depicted in Figure 2.2 (Right). Hence, an inadequate awareness of the workspace makes people perform more challenging and awkward collaboration tasks which, in turn, causes the domain tasks to be more laborious.

To achieve workspace awareness, Dourish and Bellotti [35] suggest that people engaged in collaboration should have at their disposal a shared feedback of the results from the others' actions. Additionally, Gutwin and Greenberg [54] suggest that people achieve awareness naturally in the everyday world using the following mechanisms (Figure 2.1): *feedthrough*, *consequential communication* and *intentional communication*. Feedthrough is the ability to perceive how the artifacts within the workspace change as they are being manipulated, if actors and artifacts are visible. Consequential communication related to the perception of where people are looking (*gaze awareness*) and if the their actions can be understood by seeing the performance of that action (*visual evidence*). Lastly, intentional communication happens when people are able to include gestures to qualify verbal references to artifacts on the workspace (*deixis* or *deictic gestures*) and use *demonstrations* to convey actions. It is also include in intentional communication, overheard knowledge from verbal communication about what people are doing or planning to do, also known as *outlouds*. As mentioned, these mechanisms can all be achieved in co-located encounters, yet their are extremely difficult to acquire in remote collaboration because of current technological limitations. Low fidelity representations of remote people, not seeing the full picture of what is happening, lack of presence and the separation of task and person space, all contribute to a decrease in workspace awareness. Adding to this, the perception of deictic gestures vary from the person performing the gesture and the observer, which deeply hinders communication during collaboration when people are physically collocated. In remote collaboration, this issue further undermine the perception of intentional communication. More specifically, failing to understand the exact target of a pointing gesture impede the capacity for people to correctly be aware of the collaborative task's context.

Despite that, workspace awareness has an decisive role in improving the following aspects of collaboration [54]:

- **Managing coupling** – Maintaining workspace awareness facilitate transitions in focus between loosely and tightly-coupled collaboration activities.

- **Simplification of communication** – The ability to use and observe nonverbal communication cues, reduces the complexity and length of the dialog.

- **Coordination of action** – Coordination is facilitated because everyone is attentive to the actions of others. Robinson [121] proposes that coordination can be achieve during explicit communication when people naturally determine what to do, the division of labor, how they assist one another and how they deal with different tasks performed simultaneously.

- **Anticipation** – Consisting of actions and decisions based on the prediction of what the others will do. For this, people can take advantage of the information that can be provided from consequential communication and outlouds.

- **Assistance** – Awareness helps people determine what assistance is required and what is appropriate to complete tasks. The necessity for assistance can be anticipated when difficulties in performing tasks are observable.

Our work follows the workspace awareness fundamentals to improve remote collaboration. We hypothesize that perception manipulation techniques can facilitate access to the knowledge of what actions remote people are performing and their impact on the artifacts present on the shared workspace in remote face-to-face encounters.

2.2 Virtual Meetings and Remote Encounters

In virtual meetings, technology plays a decisive role in providing the necessary means for people to communicate and collaborate while not sharing the same space. Wolff et al. [161] argued that systems that enable virtual meetings can be categorized into *audioconferencing*, *groupware*, *videoconferencing*, *telepresence* and *collaborative mixed reality systems*. Regarding videoconferencing and telepresence, there is research addressing the interpersonal space of the people involved in virtual meetings, by providing a broadcast of a single person to the group.

In virtual meetings the sense of presence of remote people has an important role in the capacity if people to communicate and collaborate. Previous research suggests that utilizing full or upper-body representations improves awareness [19, 29], since a richer vocabulary combining body language with speech can be used. Furthermore, having an understanding of the other person's gaze [107], communicative gestures [18, 80] and deictics [44, 141] are known to improve remote collaboration. Traditional systems do enable communication and even eye contact [68, 107] using video streams [42]. Sellen [129] suggested a system where remote people were represented by individual video and audio terminals, called Hydra. And, MAJIC [65] employed life-size projections of remote people to enable multi-party interactions. Morikawa and Maesako [106] follows a shared space approach and introduces HyperMirror to display local and remote people on the same screen, preserving each participant's interpersonal space. In *Office of the Future*, Raskar et al. [118] suggested that distant spaces can be blended for participants in a virtual meetings to collaborate while seeing each other as they were in the same place. Greenhalgh and Benford [50] presented MASSIVE, a virtual environment for life-size telepresence. Later, Gross et al. [52] utilized a projection-based CAVE approach to life-size telepresence also using a virtual environment.

Recent developments in commodity depth cameras enabled 3D representations that permit a more reliable life-size scale portrayal of remote people. Jones et al. [74] presented a telepresence approach that ensures of gaze and eye contact cues between

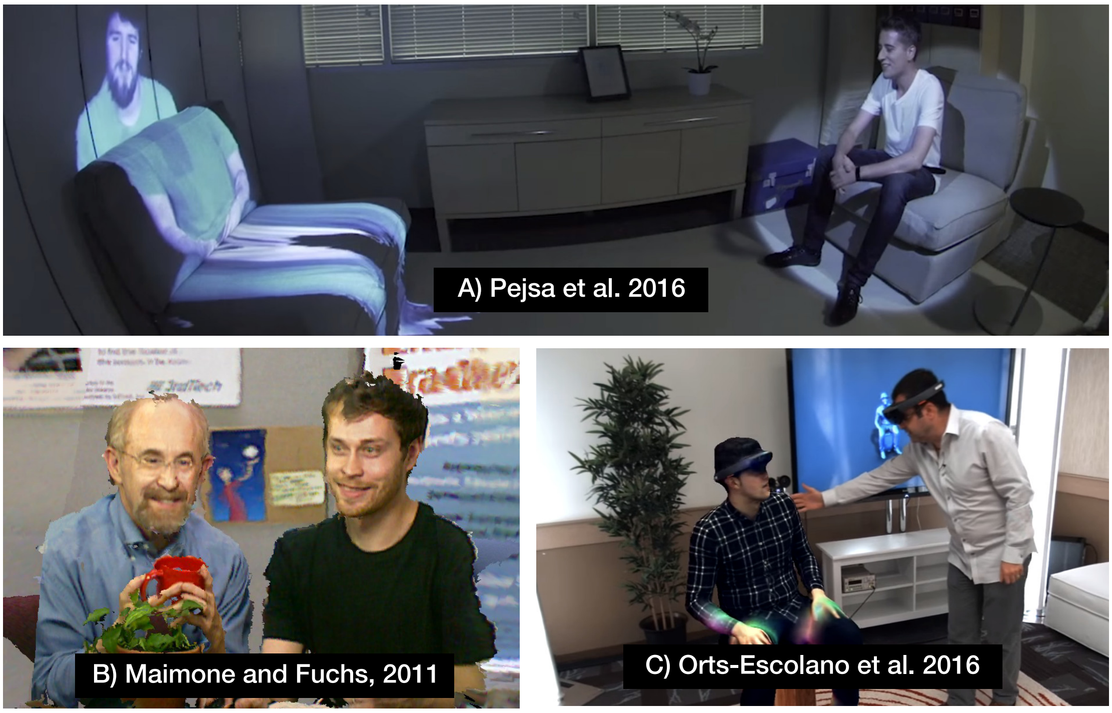


Figure 2.3: Previous research featuring full body representation of remote people.

an audience and a remote person. They resorted to the transmission of the remote person's scanned face into a 3D display. Maimone et al. [95](Figure 2.3B) presented a proof-of-concept telepresence system that enables 3D capture of a remote person and resorts to a 3D display for correct visualization. Beck et al. [17] presented an immersive telepresence system that allows distributed groups to meet and explore a virtual world. There is also previous research in utilizing mixed and augmented reality to bring remote people to the local environment. Pejsa et al. [113] employs depth cameras combined with commodity projectors to capture and render life-size representations of people creating the illusion of co-presence (Figure 2.3A). Orts-Escolano et al. [111] with Holoportation, utilizes custom depth cameras to render high-quality, real-time reconstructions of people, furniture and objects (Figure 2.3C). As well, Gotsch et al. [47] introduces TeleHuman2, a telepresence system that conveys full-body 3D video of interlocutors using a human-sized cylindrical light field display.

Preserving each participant's singular interpersonal space enable communication, however, focusing on the interpersonal space renders the user experience not appro-

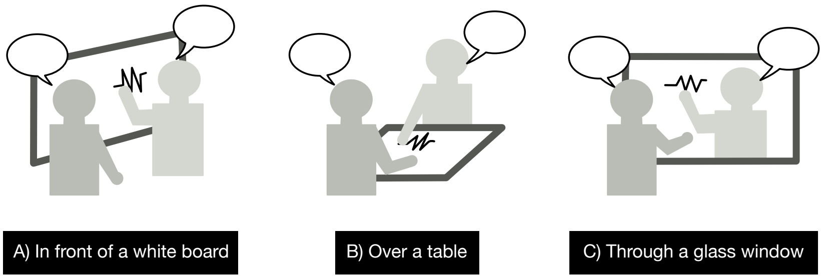


Figure 2.4: Three metaphors of seamless space for shared drawing and face-to-face conversation (from Ishii et al [68]).

priate to jointly create content. People need to meet in a shared space to perform collaborative work [30]. Thereby, Buxton [29] argued that virtual shared workspaces, enabled by technology, are required to establish a proper sense of shared presence, or telepresence when collaborating. Conventional telepresence systems rely on a separation between person space and task space, yet Buxton [28] suggested that it is also important to meet in a shared space to collaborate and identified the reference space as the shared locus were people can refer to portions of the workspace using gaze or deictic gestures. The reference space depend on the integration of person and task spaces to create a seamless and continuous space where people are able to benefit from the consequential and intentional communication mechanisms required to maintain workspace awareness. Studying the integration of person and task spaces to create a seamless space, Ishii and Kobayashi [68] identified three groupware metaphors (depicted in Figure 2.4):

- *"talking in from of a whiteboard"* or *whiteboard metaphor* – This metaphor takes advantage of the common board orientation to present to the people collaborating a side-to-side visualization of the workspace. However, since people are in a side-to-side formation, people have to be constantly shifting their attention between task space and the others' person space to perceive any of the workspace awareness mechanisms.

- *"talking over a table"* or *table metaphor* – This metaphor is familiar and derives from sitting on opposite sides of a table. Is is quite suitable for face-to-face communication because two participants can easily see each other's face. However, the orientation of the workspace becomes upside-down for one of the parties and applying a different point of view for each person renders the consequential and intentional communication mechanisms unreliable.

➜ *"talking through a glass window"* or *glass window metaphor* – This metaphor profits from the nonexistent necessity to shift focus between the task and other's person space and requires less eye-movements. However, participants do not share the same task space orientation, which obstructs the perception of the consequential and intentional communication.

Apart from the whiteboard metaphor, both table and glass window metaphors present mismatched views of the task space due to the participants' opposing points-of-view. On the table metaphor one of the participants perceive the task space to be on the wrong-side-up, and on the glass window metaphor the task space appears backwards. Indeed, this poses a challenge in awareness, however in computer-aided collaboration with digital content the task space can be rendered in a that to present exactly the same point-of-view to each participant. This task space delivery approach is known as *WYSIWIS* (What You See Is What I See - pronounced *"whizzy whiz"*). In "strict" WYSIWIS, every participant perceive the exactly same task space [137].

The whiteboard metaphor improves collaboration in the sense that having a common context imposes a shared focus, which complements our memory capabilities [138]. However, when in remote collaboration, the whiteboard metaphor hinders the capacity to recognize consequential and intentional communicative cues, since all participants are on the same side facing the task space. Tanner and Shah. [146] proposed a side-by-side approach that exploits multiple screens. One was used for content creation and another to display a side view of the remote user. Side-by-side interactions allow people to communicate and transfer their focus between watching and interacting with others and the task space [146]. Yet, as identified by Ishii and Kobayashi [68] and Gutwin and Greenberg [54], the cognitive workload of being constantly shifting focus between both task space and person space is mentally demanding and complicates further the task of maintaining the collaboration going. To deal with the visualization of remote people, Tang et al. [140] introduces the whiteboard VideoArms approach, which renders the remote

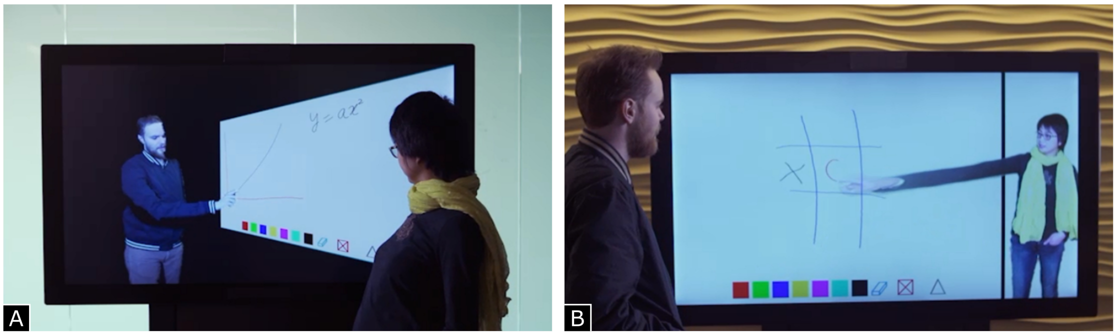


Figure 2.5: Higuchi et al. [63] whiteboard metaphor using the A) tilting whiteboard and the B) extended arm approaches.

people's arms on to the task space. The VideoArms approach provides intentional awareness but offers limited consequential awareness, since participants do not share gaze awareness. Furthering the concept of rendering person spaces on top of the task space, Kunz et al. [86], with CollaBoard, employed a life-sized video representation of remote participants on top of the shared workspace on a digital whiteboard. In the same way as the VideoArms [140] approach and despite rendering upper-body representations of remote people, CollaBoard [86] makes it difficult for participants to share gaze and, as a consequence, also offers limited consequential awareness. In a different fashion, Higuchi et al. [63] presents Immerseboard, a remote collaboration approach that employs a depth camera mounted on the side of a large touch display to acquire the front facing side of remote participants. In Immerseboard, Higuchi et al. [63] studied multiple approaches to afford workspace awareness. By tilting the virtual task space, as demonstrated in Figure 2.5A, participants are able to perceive feedthrough, consequential and intentional communicative cues. However, some usability studies suggested that the perspective introduces imprecision in perceiving the task space. Also, Higuchi et al. [63] also studied a extended arm approach (Figure 2.5B) to account for all workspace awareness mechanisms. Yet, imposing the person space on top of the task space decreases the visibility of the workspace's artifacts, causing a decrease in feedthrough.

The table metaphor exploits the concept of bringing people together face-to-face as if they were working on the same table. Nonetheless, remote collaboration presents the concern on how to depict the person space of remote participants.

Tang et al. [140] also presents a tabletop approach contained in VideoArms. Their tabletop approach follows the same principles of VideoArms on a whiteboard, rendering the remote people's arms on top of the task space. Genest et al. [45] with KinectArms proposes a low-cost toolkit that helps groupware developers build similar arm embodiments with a minimum of effort. Junuzovic et al. [78] presented the Illumishare approach that combines physical and virtual objects on arbitrary surfaces enabling participants to collaborate in a common reference space, sharing the same point-of-view. BeThere [134] resorted to depth sensors and augmented reality to render the remote participants hand enabling deictics. These approaches focus on rendering representations of remote participants' arms with little support for consequential communicative cues, despite providing for adequate feedthrough and intentional information. Though, Benko et al. [20] presents MirageTable

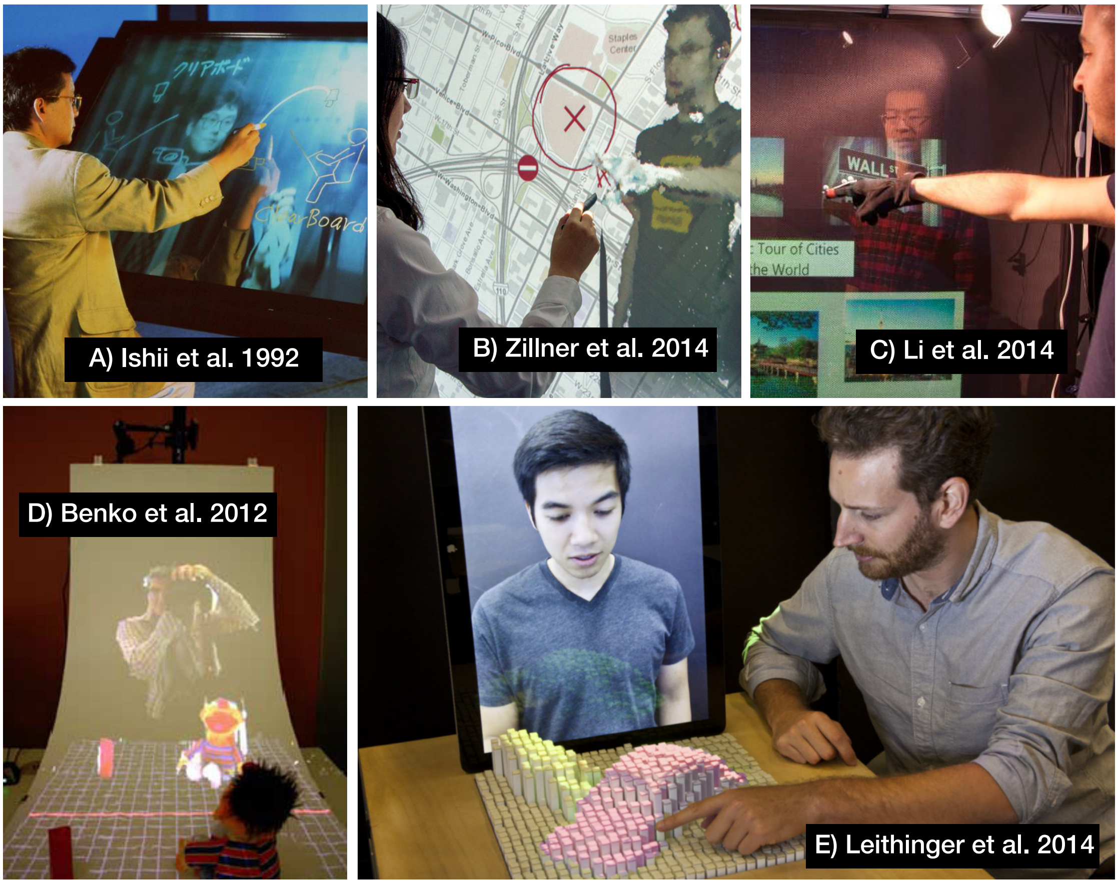


Figure 2.6: Previous research featuring remote collaboration in a shared workspace.

(Figure 2.6D), an interactive system designed to merge real and virtual worlds into a single spatially experience on top of a table. MirageTable employs a depth camera to track the user's eyes and perform a real-time capture of both the shape and the appearance of any object placed in front of the camera, including user's body and hands. By projecting a 3D mesh of the remote user, captured by depth cameras, onto a table curved upwards, a local person can interact with the virtual representation of a remote user to perform collaborative tasks. Participants share the same task space to interact with 3D physical objects, although the MirageTable remote collaboration experience is akin to sitting at the same desk opposite of one another which is prone to the occlusion issues discussed previously in §1. More recently, Leithinger et al. [90] (Figure 2.6E) proposed a physical telepresence approach based on shared workspaces with the ability to capture and remotely render the shapes of people and objects. With the Leithinger et al. [90] approach two remote participants can manipulate physical shapes in a face-to-face or corner-to-corner formation on the sides of the shared workspace. Similarly to MirageTable, Leithinger et al. [90] approach have limited support for intentional communication due to the same occlusion issues.

The glass window metaphor does not produce any coordination issues since each participant task space is isolated from their partners. Furthermore, this metaphor have integrates the interpersonal space with the shared workspace, resulting in a expedited work flow, enabling seamless integration of live communication with joint collaboration. The glass window metaphor was first used by Tang and Minneman in VideoDraw [144]. VideoDraw used video cameras to capture and super-impose remote people's hands behind the a drawing task space on horizontal displays. Following VideoDraw, Tang and Minneman introduced VideoWhiteBoard [143], where shadow representations of people were also rendered behind the task space. Both approaches were adequate in providing feedthrough, but provided no consequential information and had limited support for intentional communicative cues. From the ideas presented by Tang and Minneman [144, 143], Ishii and Kobayashi [68] introduces Clearboard (Figure 2.6A), a videoconferencing board that connects remote rooms to support informal face-to-face communication, while allowing users to draw on a shared virtual surface. In Clearboard, two participants can engage in collaborative drawing tasks while seeing each other face-to-face. To correct for the inaccurate reference system, the authors resorted to WYSIWIS approach combined with horizontally reversal of the video streams to establish the same point-of-view for both participants as if they were side-by-side. And therefore, preserving feedthrough, consequential and intentional communication in 2D tasks. Li et al. [91] (Figure 2.6C) reiterated Ishii and Kobayashi [68] findings and suggested that to maintain workspace awareness in face-to-face interactions, telepresence systems should resort to selective image reversal of text and graphics, also known as *relaxed WYSIWIS* (relaxed what-you-see-is-what-I-see) approach.

The glass window metaphor also contributes with the technological challenge of capturing the representation of remote people. Previous related work using the glass window metaphor relied on rendering remote streams of video. Using video cameras seems straightforward, however due to people's proximity to the surface displaying the workspace, video cameras need to be positioned at an angle facing the participants. This can create awkward renditions of remote people with distorted images with a different perspective from the observer. For this, Wood et al. [164] introduces the ShadowHands Technique for visualizing the remote person's hand gestures using a 3D model hand reconstruction using a depth camera. Yet, reconstructing only the remote participants' hands limits the possibility to user consequential communication. In ImmerseBoard's Mirror technique that followed the Clearboard's [68] approach, Higuchi et al. [63] concluded that using a depth camera mounted on the side of the task space, originated a rendition of remote people with low quality. Nevertheless, Zillner et al. [166], with the 3D-Board approach, concluded that multiple depth cameras can be combined to capture a 3D representation of remote people. With 3D-Board, Zillner et al. demonstrated, that a face-to-face telepresence approach using 3D reconstructions of remote people improves effectiveness when compared to a side-by-side setting (Figure 2.6B).

2.3 Context-aware Environments and Tracking Humans

Mark Weiser [155] suggested that computing technologies would move beyond devices towards being embedded in the environment, in order not to intrude on people's daily tasks. Indeed, ubiquitous environments are becoming commonplace due to the rise of interactive devices and advances on sensing technologies [120]. Therefore, by sensing the situation of the environment, digital systems can use that knowledge to infer intent to interact [127]. Context-aware interactions can thus exploit what is happening in close proximity of a person or device. Matthews et al. [103] developed a toolkit that examines context by grabbing users' attention to peripheral displays. Annett et al. [4] demonstrated a proximity-aware tabletop that determines users' presence, position and the arm which is interacting with the display. Jota et al. [101] described the interaction design space on and above horizontal interactive surfaces. Interactive vertical display interfaces can also benefit from the environmental context. Vogel et al. [152] proposes an interaction framework for interactive public displays. The authors map out the area in front of a vertical display into ambient, implicit, subtle and personal interactive spaces. When a person is transitioning between those spaces, the ambient display can display different contextual information based on the person's proximity, ranging from public details at a distance to a more private information, when in close proximity. Also, with proper sensing technologies, context-aware systems can react to the normal everyday interactions of groups of people [100]. Edward Hall introduces the Proxemic Theory [56] and observed that space, distance and orientation between people impact the way they interact with each other. Ballendat et al. [13] suggested that knowledge about the way people position themselves can be exploited in ubicomp environments to start or end interactions, establish connections, and even, automatically transfer personal files between devices. Pushing this notion further, Marquardt et al [96, 99] applied proxemics to explore the design space in ubicomp environments to mediate interactions between people and their personal and public devices, as demonstrated in Figure 2.7.

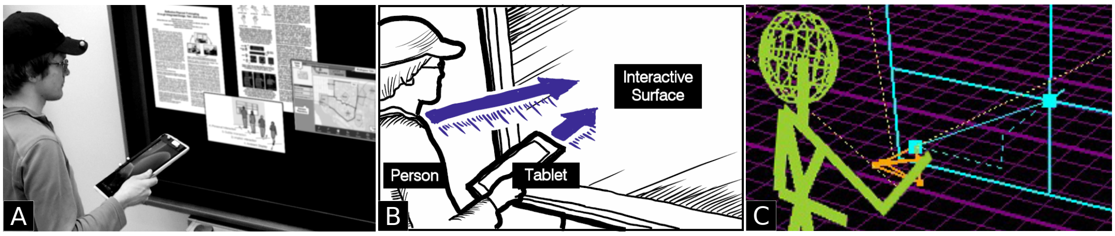


Figure 2.7: The Proximity Toolkit keeps track of the A) entities (person, device and vertical surface) present in the environment and the B) proxemic relationships between them to provide a C) computational model for client applications – Marquardt et al. [97]

Our work builds on previous research on instrumenting the environment instead of requiring users to carry physical tracking devices [66]. Antifakos and Schiele [5] demonstrated that wifi networks are able to detect proximity. Fails and Olsen [39] proposed a technique to sense hand gestures to interact with surfaces via skin color detection using RGB cameras. However, recent developments in commodity depth cameras allow people tracking [158] in expeditious manners.Depth cameras can disambiguate color images with depth information [3] and allow devices to estimate human poses [132]. Wilson and Benko [159] presented LightSpace, a prototype that combines depth cameras and projectors to provide interactions on and between multiple surfaces. People can transfer and manipulate objects in one surface, "pick" and "drop" on another device. LightSpace employs multiple depth cameras calibrated to the same coordinate system. It is also the first approach to combine depth data from multiple depth cameras. Sousa et al. [136] also used multiple depth cameras to deal with body occlusions in a remote collaborative environment for groups of people. Yet, their tracking approach did not consider users' orientation and full body tracking. The Proximity Toolkit [98] uses sensor fusion to gather data from multiple different tracking systems to gather proxemic data about people and devices. The toolkit combines skeleton data from depth cameras for body tracking with a marker-based system to track devices. Developers can create user experiences using both the tracked entities and the proxemic relationships between them. Despite having the same motivation, our approach focuses not on the relationships between entities, but rather deals with combining multiple depth sensors. In this work, we address resolving multiple skeletons into persons, choosing the optimal sensor for each person while avoiding orientation errors when tracking people from behind.

More recently, Wu et al. [165] presented EagleSense (Figure 2.8), a top-view camera-based system for tracking people's position, orientation and activities. A top-view approaches minimize body occlusions by other people, although, as reported by the authors, this approach proves to be difficult when acquiring body skeleton joints. Seyed et al. [130] introduced the SoD-Toolkit. It offers a set of tools for tracking people, interactions between them and devices using multiple sensors.

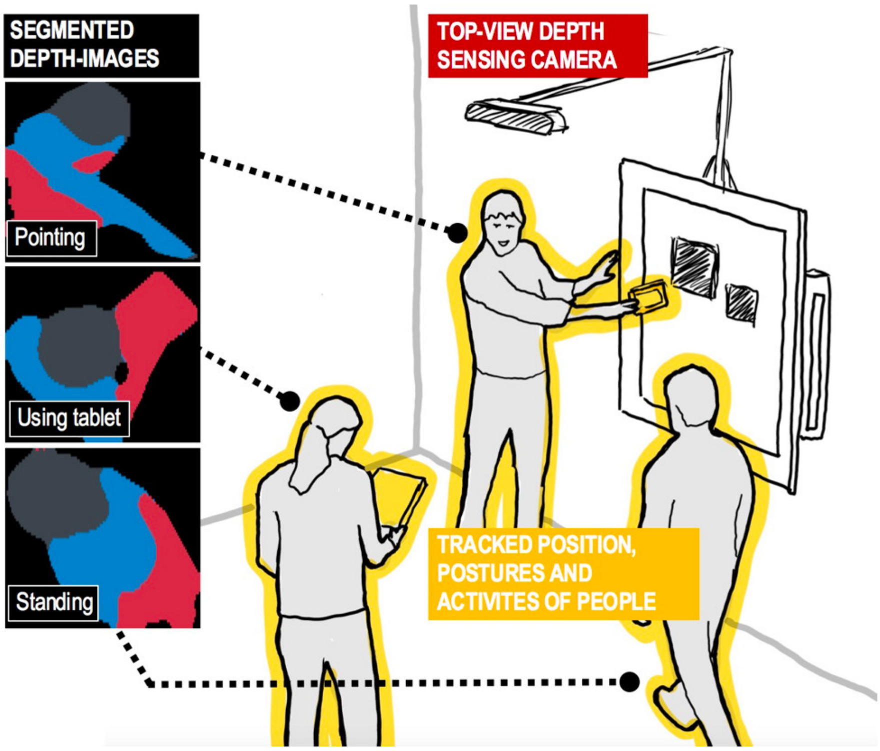

SEGMENTED DEPTH-IMAGES
Pointing
Using tablet
Standing
TOP-VIEW DEPTH SENSING CAMERA
TRACKED POSITION, POSTURES AND ACTIVITES OF PEOPLE

Figure 2.8: The EagleSense recognition system for human posture tracking and activity recognition using a single top-view depth- sensing camera – Wu et al. [165].

To achieve the goal of improving workspace awareness in face-to-face remote collaboration with 3D digital artifacts, as proposed in our research statement, this work relies on context information of the environment to infer participants' intentions to interaction with the task space. Moreover, to provide for compelling telepresence experiences for collaboration with 3D objects, contextual information of the environment and the participants' posture are required [32]. Yet, the above mentioned approaches that constitute the related work on capturing context data are not optimized to support rapid-prototyping of remote collaborative experiences. Thus, it is the intention of this research to also contribute a tracking approach focused on remote collaboration and telepresence.

2.4 Virtual Representations of People

To enable collaboration, mixed reality environments should rely on complete portrayals of people to allow for the understanding of nonverbal cues in addition to normal speech communication. Nonverbal communicative cues include facial expressions, gaze, body posture, *deixis* to indicate objects referred to in speech [104], and how people utilize the space and position themselves when communicating (*Proxemics* [56]). Being able to perceive such nonverbal cues is beneficial for the sense of co-presence and helps people to communicate naturally [28]. For this reason, such environments rely on virtual representations of people to provide the necessary awareness [55] of the collaborator's activities.

Early groupware approaches employed telepointers and cursors [49] as a mean to provide awareness of people's actions on a shared workspace. However, telepointers and cursors provide limited knowledge about people's gestures, making it impossible to anticipate their actions. Hence, dynamic representations of arms [141, 142] and hands [134, 164] have been studied to convey such nonverbal cues, yet yielding limited awareness of other people's presence. Indeed, the more realistic fully 3D rigged virtual avatars are, the better they convey the feeling of co-presence [72]. Recent developments in commodity depth cameras enabled lifelike 3D full-body reconstructions of people. For example, Maimone et al. [95] presented a telepresence approach that employed full-body reconstructions of people in a 3D display for correct visualization. In the case of Beck et al. [17], 3D reconstructions are applied to virtual worlds where distributed groups meet. There is also previous research in utilizing Mixed and Augmented Reality to bring remote people to the local environment. Pejsa et al. [113] employed commodity projectors to render life-size representations of people creating the illusion of co-presence. Furthermore, Orts-Escolano et al. [111] demonstrated high-quality reconstruction through head-mounted displays, creating an experience akin to physical presence.

The state-of-the-art suggests that full body virtual avatars can convey natural nonverbal communication cues essential for collaborating in a shared environment. Our research builds on this previous work to improve the perception of pointing gestures by manipulating the way virtual embodiments are presented to people in mixed reality environments.

2.5 Deixis in Mixed Reality Collaborative Environments

Pointing gestures are an important nonverbal communication tool to coordinate and maintain a up-to-date understanding of the context when collaborating, yet there are situations where people experience difficulties describing verbally distal referents with hard-to-describe shapes or locations [85]. In these cases, the ability to observe pointing gestures facilitates collaboration by making the communication task more natural.

Fussel et al. [41] suggested that, in collaborative distributed settings, perceiving gestures improves task performance. Indeed, deictic references increase workspace awareness by allowing people to qualify verbal references to artifacts in a shared workspace [54]. However, Wong and Gutwin [163] suggested that using deictic referencing in mixed reality environments is more demanding than in the real world due to narrow fields of view (FOV) and poor resolution of current display technology.

Previous works showed that awareness cues can effectively support communication, such as using virtual pointers [49, 36, 110] or enhancing the collaboration through highlighting visual and audio cues [21, 117]. Yet, virtual pointers can provide inadequate or conflicting augmentations of pointing and produce a direction different from the pointers arm, and highlighted objects may not match the pointing gesture. Also, target highlighting is limited to predefined objects and its discrete movement makes it harder to control. When used in a collaborative environment these can contribute to clutter [163]. Furthermore, Piumsomboon et al. [115] concluded that such enhancements could obfuscate important social cues (facial expression or body gestures). Despite that, Piumsomboon et al. [116] introduced Mini-Me, an adaptive avatar that uses redirected gaze and gestures, and found that their approach was successful in improving user's awareness of their partner in a collaborative mixed reality interface.

Our work focuses on improving the perception of deictic references, and consequently, expedite collaboration. Thus, our approach exploits the concept of gesture redirecting by warping virtual representations of people in an imperceptible way, without losing other important social cues.

2.6 Production and Interpretation of Deictic Gestures

Gestures to indicate referents typically include extending a body part (e.g., fingers, hands, eyes, head) towards an object or location [81]. Pointing at things is both a natural and ubiquitous practice during communication. Wong and Gutwin [162] suggested that people are "experts at (...) interpreting deictic gestures". Yet, people often fail to determine the exact location to which another person is pointing to.

The perceptual accuracy of discerning referents of pointing gestures depends on whether the pointing gesture is proximal or distal [128]. When indicating proximal referents (proximal pointing), pointers are able to touch the target and observers can identify referents with confidence [15]. In contrast, when pointing at distal referents, people usually align the tip of their pointing finger with their dominant eye [160]. Indeed, ray pointing techniques using the eye to index finger vector have been employed to detect pointing gestures for object selection and manipulation in Mixed Reality [24, 89, 154, 114, 153] and large scale displays [6, 76], since they offer an high level of accuracy [76]. Despite that, previous research suggested that interpretation of distal pointing is an extrapolation of the vector defined by the pointer's posture [15, 160].

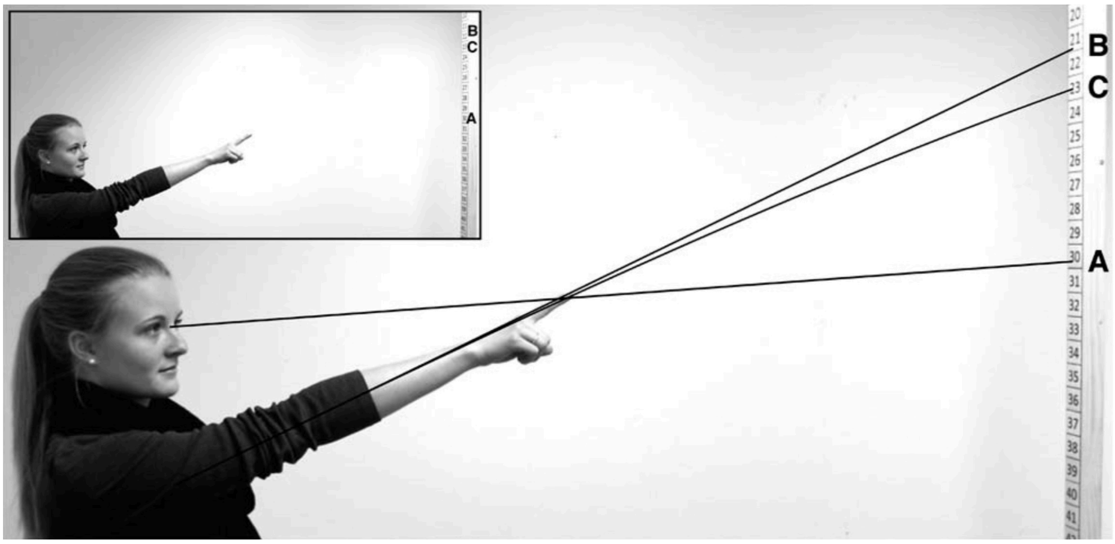


Figure 2.9: Misunderstanding pointing gestures. The Figure (from Herbort and Kunde [61]) illustrates the linear extrapolation of the eye–finger line (A), the linear extrapolation of the arm–finger line (B), and an exemplar nonlinear, human extrapolation of the arm–finger line (C). Although most observers would think that the person in the inset is pointing to C, the pointer was instructed communicate A.

Herbort and Kunde [61] proposed that this difference between production and interpretation accounts for the systematic spatial misunderstanding of pointing gestures to distant referents. Notably, people carrying out a vector extrapolation exercise often exhibit a bias toward horizontal or vertical axis [23]. Salomon [124] suggested that human attempts at vector extrapolation deviates from a geometric linear extrapolation. Insomuch as people observing the arm to index finger vector in Figure 2.9 would interpret a target position between locations B and C.

Herbort and Kunde [61] asserted that people interpret pointing gestures by using a nonlinear extrapolation of the pointer's arm-finger vector. This non-linearity characteristic of perceiving pointing gestures, can be described as a Bayesian-optimal integration of a linear extrapolation of the arm-finger vector and the observer's prior assumptions about likely referent positions. Following this insight, Herbort and Kunde [61] introduced a predictive Bayesian model of pointing gesture interpretation to estimate the position of referents. Their model is based on the assumptions that participants engage in geometric extrapolation of the arm–finger line or eye–finger line and participants integrate the geometric extrapolation and a priori information according to Bayesian theory [82, 84]. The proposed Bayesian model can be expressed as follows:

$$\hat{y}_{Bayesian} = \frac{d^{-2}(1-w)y_{geo} + wy_0}{d^{-2}(1-w) + w} \tag{2.1}$$

Equation 2.1 considers d as the horizontal distance between the plane containing referents and the pointer's shoulder, y_{geo} as the result of geometric extrapolation, and y_0 as the a priori assumed average referent position, which is set to the shoulder height of the pointer. The Bayesian model also considers the free parameter w, which relates to the variability associated with the linear extrapolations to the variability associated with the observer's unknown prior assumptions. The parameter w can assume values between 0 (participants rely exclusively on geometric extrapolation) and 1 (participants rely exclusively on the a priori assumption). The authors determined values of w individually for each participant and provided the average values for different gesture interpretation conditions. The estimation of the referent's position using the Bayesian model was evaluated in a study with participants, revealing that the nonlinear extrapolation of pointing interpretation can successfully be described by the proposed model and referent estimates changed nonlinearly as a function of distance. Moreover, the study showed that head orientation plays a marginal role in the interpretation process.

In collaborative settings, this gesture interpretation issue undermines the perception of intentional communication. More specifically, failing to understand the exact target of a pointing gesture impedes the capacity for people to correctly be aware of the collaborative task's context. In this research, we propose using perception manipulation to better match the way people interpret the location of distal referents that the pointer wants to communicate.

2.7 Discussion

In this work, we focus on the revision of previous research on remote collaboration over a common shared workspace. Furthermore, since the aim of our research is to improve workspace awareness in such collaborative settings, we classified the presented related work taking into account how workspace awareness mechanisms were afforded by each approach. The classification is presented in Table 2.1 The approaches are listed in order to group them by the collaboration metaphor employed, starting with the *"talking in from of a whiteboard"*, then *"talking over a table"* and finally *"talking through a glass window"*. We also classify in regards to the *close encounter formation*, *person space*, *reference space* and whether they support 2D or 3D artifacts on the shared workspace (*workspace artifacts*).

From the proposed classification, we can identify that approaches using the whiteboard metaphor [63, 86, 140] are successful in providing feedthrough feedback, as initially suggested by Ishii and Kobayashi [68]. However, these approaches are unable to convey effectively consequential communication, which makes it difficult for participants to infer if a colleague needs to be assisted or anticipate their partner's actions. Regarding intentional communication, Collaboard [86] and VideoArms [140] both demonstrated that the whiteboard metaphor can provide de means for participants to use deictic gestures. Still, they do not support other types of nonverbal cues such as facial expressions or *demonstrations*.

Approaches following the table metaphor [20, 45, 90, 78, 134, 140] and using an *individual point-of-view* are unable to support intentional communication. However,

Approach	*Close Encounter Formation*	*Person Space*	*Reference Space*	*Workspace Artifacts*	*Collaboration Metaphor*
Collaboard [86]	Side-to-side	Upper body	WYSIWIS	2D	Whiteboard
VideoArms [140]	Side-to-side	Remote Arms	WYSIWIS	2D	Whiteboard/Table
Immerseboard [63]	Side-to-side	Upper body	WYSIWIS	2D	Whiteboard/Glass Window
Leithinger et al. [90]	Face-to-face/Side-to-side	Upper body	Individual POV	Physical 3D	Table
MirageTable [20]	Face-to-Face	Upper body	Individual POV	Virtual + Physical 3D	Table
KinectArms [45]	Side-to-side	Remote Arms	WYSIWIS	2D	Table
Illumishare [78]	Side-to-side	Remote hands	Individual POV	2D	Table
BeThere [134]	Side-to-side	Remote hands	Individual POV	2D	Table
ShadowHands [164]	Face-to-face	Remote hands	WYSIWIS	2D	Glass Window
VideoDraw [144]	Face-to-face	Remote hands	WYSIWIS	2D	Glass Window
VideoWhiteBoard [143]	Face-to-face	Shadow	WYSIWIS	2D	Glass Window
Clearboard [68]	Face-to-face	Upper body	WYSIWIS	2D	Glass Window
Li et al. [91]	Face-to-face	Upper body	relaxed WYSIWIS	2D	Glass Window
Zillner et al. [166]	Face-to-face	Full body	WYSIWIS	2D	Glass Window

Table 2.1: Classification of the most relevant shared workspace approaches for remote collaboration (Part I).

when employing the *WYSIWIS* method, the reference space can support deictic gestures if the person space of remote people is mirror reversed. Yet, for 3D object selection and manipulation, the support for intentional communication still is limited.

Regarding approaches following the glass window metaphor [63, 68, 91, 92, 143, 144, 164, 166], person spaces comprised solely by renditions of the remote person's hands are unable to convey consequential communication. Nevertheless, to convey intentional communication cues, the glass window metaphor is highly effective. This leave us to conclude that the glass window metaphor associated with full or upper

	Workspace Awareness		
Approach	*Feedthrough*	*Consequential Communication*	*Intentional Communication*
Collaboard [86]	Available	Limited	Available
VideoArms [140]	Available	Limited	Available
Immerseboard [63]	Available	Available	
Leithinger et al. [90]	Available	Available	Limited
MirageTable [20]	Available	Available	Limited
KinectArms [45]	Available		Available
Illumishare [78]	Available		Limited
BeThere [134]	Available		Limited
ShadowHands [164]	Available		Available
VideoDraw [144]	Available		Available
VideoWhiteBoard [143]	Available		Available
Clearboard [68]	Available	Available	Available
Li et al. [91]	Available	Available	Available
Zillner et al. [166]	Available	Available	Available

Table 2.2: Classification of the most relevant shared workspace approaches for remote collaboration (Part II).

body representations of remote people can provide for a seamless and continuous interaction space for remote collaboration, as previously argued in §1. Despite this, research enabling face-to-face telepresence for collaborating in 3D virtual workspaces is still limited, mainly due to occlusions resultant from contrary points-of-view. Therefore, there is still the need for further improvements to deal with existing ambiguities when people perceive 3D objects from different perspectives.

In conclusion, there is no approach, as far as we know, for face-to-face remote collaboration with 3D content that offers the devices to fully provide participants with the adequate workspace awareness for tasks in 3D space.

2.8 Chapter Summary

In this chapter, we surveyed related work in the research areas of remote collaboration and awareness of workspace and remote people. We also reviewed previous research on virtual representations of people. Furthermore, a detailed discussion of the state-of-the-art is also presented.

3
Perception Manipulation

One recent trend to improve the overall user experience in mixed reality settings is to capitalize on the dominance of vision when our sense conflict by manipulating human perception. Our vision is that perception manipulation techniques can also be applied to improve remote collaborators having the same perception of what is happening in a shared workspace. In this chapter, we detail the theoretical foundations of our approach. We start by introducing our integration of person-, task-, and reference space and then proceed to lay out our concept for using perception manipulation in remote face-to-face interactions.

3.1 Integrating Person-, Task-, and Reference Space

We aim at contributing an interactive space where remote face-to-face collaborative interactions with 3D virtual content can occur in a continuous area supporting nonverbal cues. While a naive approach to reach a capable design space could draw from a

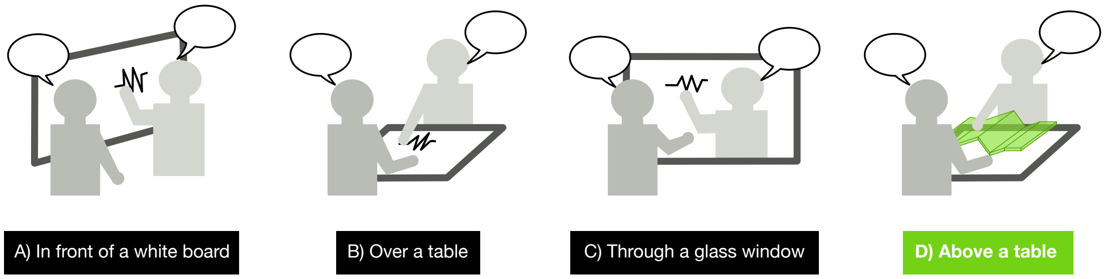


Figure 3.1: We introduce "above a table" metaphor, an approach to remote collaboration.

"through a glass window" metaphor (Figure 3.1C), the spatial perception of 3D models is inadequate due to the medium's lack of ability to provide an acceptable comprehension of depth. The best option could be, drawing from a "over a table" metaphor (Figure 3.1B) that can easily afford spatial interactions with adequate depth perception. However, in a simple "over a table" approach, 3D objects between two remote people imply that no one shares the same understanding of the shared workspace. In short, both approaches successfully hold adequate space for people to communicate, but at the expense of workspace awareness naturally. And, as suggested by Gutwin and Greenberg [54], an insufficient workspace awareness hugely interfere with a productive collaborative experience.

Hence, we argue in this thesis that existing metaphors are not capable of being extended to face-to-face interactions with 3D content. For this reason, we propose adding a new metaphor to the set identified by Ishii et al [70]. We propose an "above a table" metaphor for interacting with 3D digital content for face-to-face remote collaboration, as depicted in Figure 3.1D.

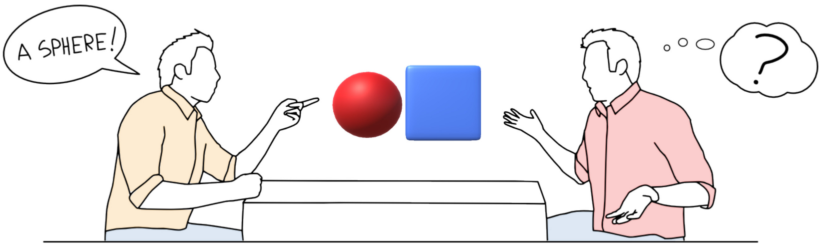


Figure 3.2: Occlusions in face-to-face interactions.

The "above a table" concept merges "over a table" with "glass window" configurations by exploiting perception manipulation techniques to overcome occlusions and guarantee that everyone has a shared understanding of the workspace. Figure 3.2 provides a simple depiction of the occlusions issue when people collaboration face-to-face. This concept provides the three-dimensionality to populate the space between collaborators with 3D objects. And geographically distant people can relate to objects without added effort when shifting their attention between workspace and the other person. Indeed by positioning the virtual content between people, participants can profit from regular face-to-face interactions as if they were physically co-located. Next, we introduce perception manipulations and how to utilize it to promote workspace awareness.

3.2 Perception Manipulation and Illusions

Perception manipulation utilizes the notion that when senses conflict, vision often dominates (visual dominance). That is, subtle illusions can give people a different or improved understanding of a specific reality. In HCI, visual dominance can be used to overcome the limitations of the physical world. Recent research in mixed reality suggested that manipulating the environment or the person can change people's judgment of the surroundings to enable more effective use of the physical space or enhance one's ability to interact with the environment. Current mixed reality technologies rely heavily on head-mounted displays for people to experience virtual content. Therefore, perception manipulation can be achieved by changing, reshaping, and manipulating to various degrees the way people see the world or their virtual body, as summarized in Figure 3.3.

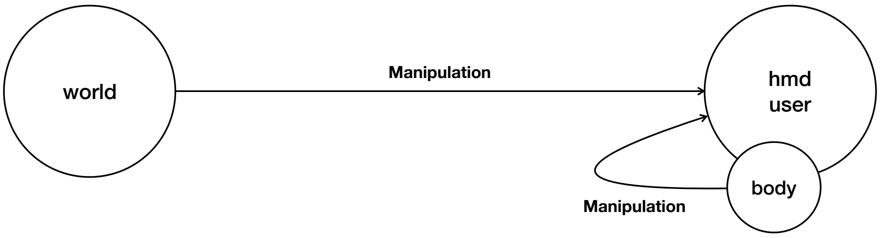


Figure 3.3: Perception manipulation concepts.

A classic example of this type of approach is the *redirected walking*, firstly introduced by Razzaque et al. [119]. In this technique, the virtual environment is transformed in such a way that the person can use a natural walking to cover more considerable distances in a limited physical space. Transformations are applied to the environment in the form of unnoticeable rotations during gait or when the user is blinking [87]. Illusions can also be used to for haptics since visual dominance studies show that people often perceive the visible shape instead of the tactual shape [123, 46]. This method is especially useful when interacting with virtual objects. *Pseudo-haptic feedback* [94] can be used to give the illusion of weight by introducing an offset between the real and rendered position of the hand while pushing against a virtual object conveys the illusion of stiffness [88]. Samad et al. [125] visually manipulated the control/display ratio of the hand movements to change the perception of weight when manipulating objects using physical proxies. Visual illusions have also be used for retargeting people's gestures, to overcome tracking issues [1] or for reprising physical proxy objects [12]. Azmandian et al. [11] theorized that perception manipulations can be classified into *body warping* – "manipulating the virtual representation of the person's body", and *world warping* – "manipulating the virtual world's coordinate system to align virtual and physical objects". Yet, they proposed *hybrid warping* – "A dynamic combination of *body* and *world warping*", and demonstrated that this technique is effective in diminishing the effects noticeability of *body* and *world warping*. Therefore, previous approaches suggest that, in interactive systems, perception manipulation can be used to warp reality before being presented to users, contributing to the people's willing suspension of disbelief. Furthermore, Congdon et al. [33] suggest that shared virtual spaces can be presented differently to two collaborators. The authors propose merging the individual virtual environments for peoples' movements to be are dynamically mapped into their collaborator's environment while creating the impression of sharing the same space.

Taking inspiration from these previous works, in this dissertation, we hypothesize that by manipulating the reality of what people expect in a remote face-to-face encounter, participants can perceive the same contextual knowledge and therefore overcome the limitations of the "above a table" approach.

3.3 Perception Manipulation Applied to Face-to-face Remote Collaboration

In this dissertation, we intend using perception manipulation in a way that differs from the state-of-the-art. As described in the previous section, current perception manipulation approaches manipulate people's understanding of their environment or their bodies. However, to overcome the limitations of the above a table metaphor, we propose that virtual meeting collaborators should perceive manipulated versions of their counterpart's person-task space that matches their own reference space (Figure 3.4). In other words, we propose translating remote people's actions into equivalent behaviors from the perspective of a local observer. Through warping transformations, we suggest spatial interactions in a remote reference space can be altered into different actions in a local reference space without losing the message that collaborators want to communicate. For this, we identify three types of transformations:

1. **Shared Perspective** – Collaborators are at opposing ends of the workspace, yet they have the same overview of the workspace, as depicted in Figure 3.5. This concept clusters transformations to the environment that enables opposing collaborators to share the same point-of-view of the shared workspace. This approach differs from simulating a side-to-side behavior because sharing the same perspective considers that participants perceive the same view of the workspace without them assigning to each other left or right stance.

2. **Workspace Warping** – This concept corresponds to any manipulation that changes the properties or the representation of the workspace to translate

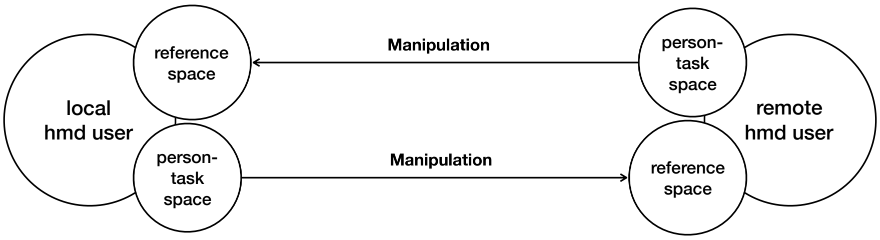


Figure 3.4: Our perception manipulation approach.

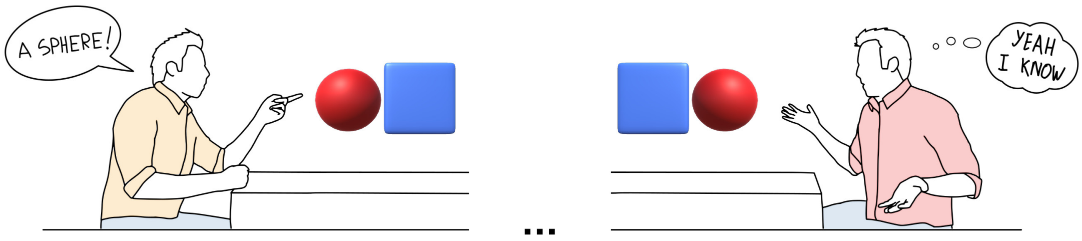


Figure 3.5: Remote collaborators are at opposing ends of the workspace, yet they share the same understanding of the workspace.

remote people's actions to the local reference space. Including geometrical transformations such as mirroring or dimensional inversions.

3. **Body Warping –** It is corresponding to any transformation that entirely or partially changes the properties or appearance of virtual representations of people. Individually, we consider being body warping when virtual representations of people are reshaped to improve perception but do not faithfully portray the actual likeness of the people they are representing.

Summarizing, we propose an ”above a table” approach for collaborators to interact face-to-face while the same understanding of the workspace results from the manipulations above identified. Our approach consists of investigating these perception manipulation techniques to study their impact on workspace awareness.

3.4 Chapter Summary

In this chapter, we introduced our vision to improve face-to-face remote collaboration. Our approach combines an “above a table” metaphor with perception manipulation techniques designed to establish a shared understanding of a common workspace. This chapter also concludes the first part of this dissertation. In the following parts, we detail the technological infrastructure developed to support the prototyping of co-located and remote interactions and introduce the empirical steps we took to validate our research statement.

Part II

Prototyping Co-located and Remote User Experiences

Virtual body representation for telepresence depends on the ability of the systems to capture, broadcast, and render people in different extended reality environments, as well as ascertain the relationship between people and their surroundings. Past research has been proposing new approaches and technologies for sensing and capturing people (Sections §3 and §4 of Chapter 2), yet none provide the tools necessary to easily incorporate body tracking, interactive surfaces, and point-cloud representation of people. In this part of the dissertation, we introduce the Creepy Tracker Toolkit that was developed to facilitate the access to real-time tracking information from multiple sources without the need to tinker with low-level data whenever it is necessary to tests new ideas.

4 The Creepy Tracker Toolkit

In this chapter, we introduce an open-source toolkit to ease prototyping context-aware interactive approaches with multiple commodity depth cameras. Human-Computer Interaction researchers strive to understand the context to anticipate user requirements when designing new experiences. Context-aware computing relates to interactive systems that leverage different sensing methods to gather understanding about their surroundings [127]. Moreover, context-awareness is an essential foundation of ubiquitous [156] and pervasive systems [7]. In this chapter, we introduce a set of tools for tracking people that were fundamental to developing and evaluating the research described in this dissertation.

4.1 Motivation

Recently, the interaction space and the physical relationships between people and interactive devices have been the focus of much research. Indeed, recent develop-

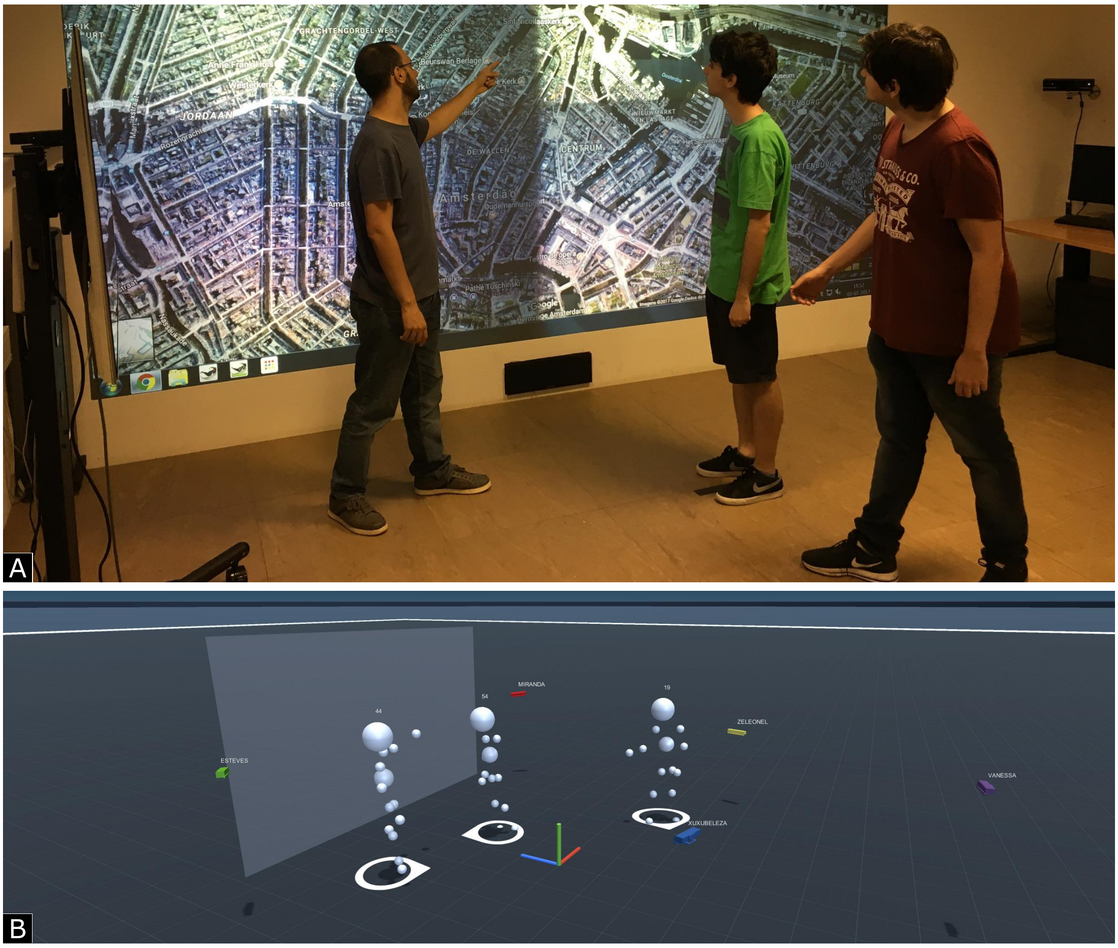
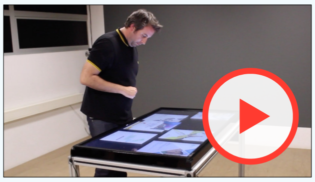


Figure 4.1: *Creepy Tracker* is an open-source toolkit that provides spatial information about people and interactive surfaces. To do this, it resorts to multiple depth-sensing cameras to A) capture what is happening in the physical world and B) maintains a dynamic data structure, available to software clients.

ments using spatially-aware ubiquitous environments can infer interactions and even people's intentions to interact using fused sensor data [73, 75]. The emergence of commodity depth sensors, such as the Microsoft Kinect, contributed out-of-the-box



Video Figure 1. Creepy Tracker Toolkit Overview.
http://web.ist.utl.pt/~antonio.sousa/videos/sousa2017-acmiss-video-figure.mp4
(File size: 16.9 MB)

tracking approaches to focus ubicomp researchers' attention on the interaction design and user experience. However, depth cameras' limitations can create barriers to the design and evaluation of new interaction approaches and techniques. In fact, single depth cameras cannot handle occlusions. While multiple cameras can mitigate this problem, they generate large volumes of network traffic for real-time data streaming. Furthermore, combining data from multiple coordinate systems (one per sensor) require additional processing and calibration methods. These issues must be addressed before any attempt to designing novel user experiences.

Because of the reasons presented above, we present the *Creepy Tracker* Toolkit, a set of open-source[1] software tools to aid rapid-prototyping of context-aware interactive systems using multiple depth cameras. The proposed tools allow seamless installation and offer a backbone for developing such systems, thus concealing the complexity inherent to current approaches, as demonstrated in Figure 4.1 and in Video Figure 1. We followed a conceptual line parallel to the work from Seyed et al. [130]. Our tracker consists of a network server that combines data from multiple depth sensors to provide full-body positional tracking of people within a room-sized volume. The toolkit manages the spatial locations of interactive surfaces and can easily infer spatial relationships to the people surrounding them. Also, it supports flexible full-body point-cloud representations of people. For the purpose of this work, we consider the definition of context provided by Dey et al. [34], which is the knowledge of "location, identity and state of people, groups, and computational and physical objects". To this end, the *Creepy Tracker* toolkit combines multiple depth sensors to provide full-body positional tracking of people and tools to acquire precise locations of interactive surfaces while providing a networked stream of context data front-end applications.

Seyed et al. [130] tackled a similar challenge to ours. They introduced the SoD-Toolkit. It offers a set of tools for tracking people, interactions between them and devices using multiple sensors. Our *Creepy Tracker* follows some of the concepts from SoD-Toolkit, further exploring them. Indeed, we focus on continuous 3D spatial context and full body tracking of multiple people. Moreover, we allow for explicit and accurate surfaces' calibration, as well as point-cloud user virtual representation.

The contributions of this research are threefold:

[1] Github: `https://github.com/vimmi3D/CreepyTracker`

1. a set of tools for rapid-prototyping context-aware applications, that incorporates body tracking, interactive surfaces and point-cloud representation of people;

2. practical considerations when using *Creepy Tracker*, obtained from a system's performance evaluation;

3. and, different scenarios that can be implemented using our toolkit (detailed in Chapter 5);

In addition to disclosing an open-source toolkit, we also detail how we overcame every major technical obstacle while detailing the toolkit's design and implementation. Furthermore, we demonstrate the scalability and performance of our toolkit using five Microsoft Kinect depth cameras in a evaluation regarding latency and accuracy against a maker-based optical system, developed specifically for motion capture and computer generated imagery. We also provide a discussion of the evaluation results.

4.2 Approach

The *Creepy Tracker* toolkit uses a network of distributed sensor units connected to a central hub. Each sensor unit is composed of a Microsoft Kinect depth camera and standalone C# application running on a single computer. The number of sensors is directly related to the area required by the interaction being designed. Interactions with a single typical (up to 4×2m) vertical surface may require one or two units, while interactions around a tabletop most commonly need several (up to 5) sensors surrounding that surface. Each sensor unit provides a continuous data stream. These converge on the tracker's central hub, which is responsible for synchronization, processing and merging the data, as depicted in Figure 4.2. The central hub also broadcasts the state of the tracked environment to client applications. Moreover, for the virtual model of the tracked people and surfaces to be precisely aligned with the physical topology of the room, the sensors' position and orientation must be first calibrated. Adding surfaces requires an active calibration for each new surface by defining the surface plane using 3D depth data of one sensor. After calibration, as people move in the tracked area, the virtual model gets updated in real-time, while broadcasting the updated data.

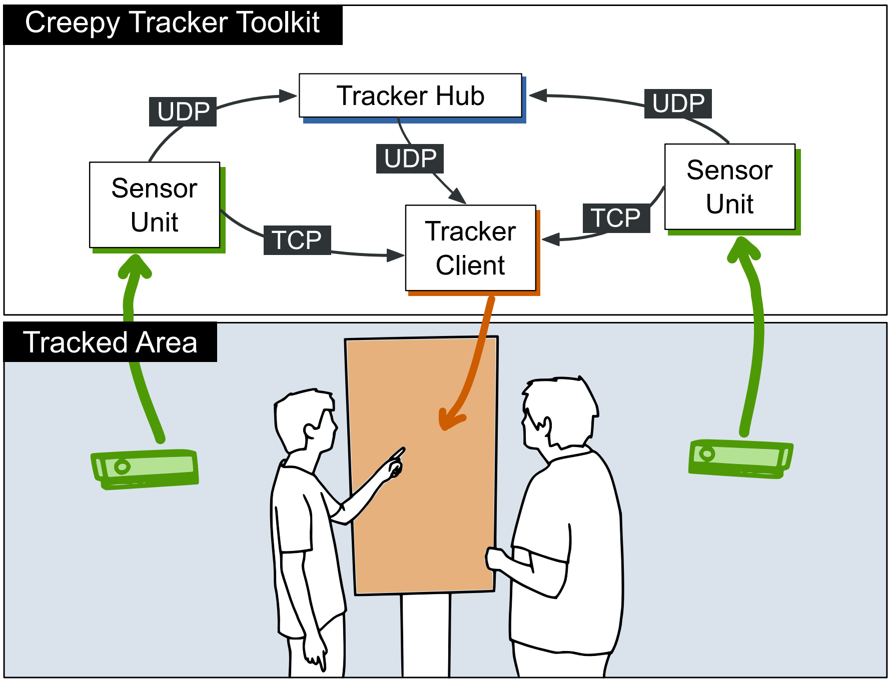


Figure 4.2: Overall system's architecture.

In this section, we provide a detailed overview of the system's components, describing their implementation and application.

4.2.1 *Sensor Unit*

All sensor units, each of them connected to an individual depth sensor, capture color, depth data and the body tracking model of every observed person in the tracked area. Each body model is associated with a numerical factor to represent the estimated degree of confidence about the quality of the tracked person, which is sent together with the body model data. The confidence factor is calculated by adding all tracked body joints' weight while discarding inferred ones. The weight of each joint can be customizable so that the tracker can favor specific joints, useful for different scenarios. For instance, pointing tasks require far more importance given to hands' than feet' joints. Figure 4.3A shows an individual sensor client tracking two people, the person closest to the camera has a lower degree of confidence because half of the lower limbs' joints cannot be seen. Tracked people with confidence factors below a configurable threshold are ignored. This method is highly effective to deal with the recurring oc-

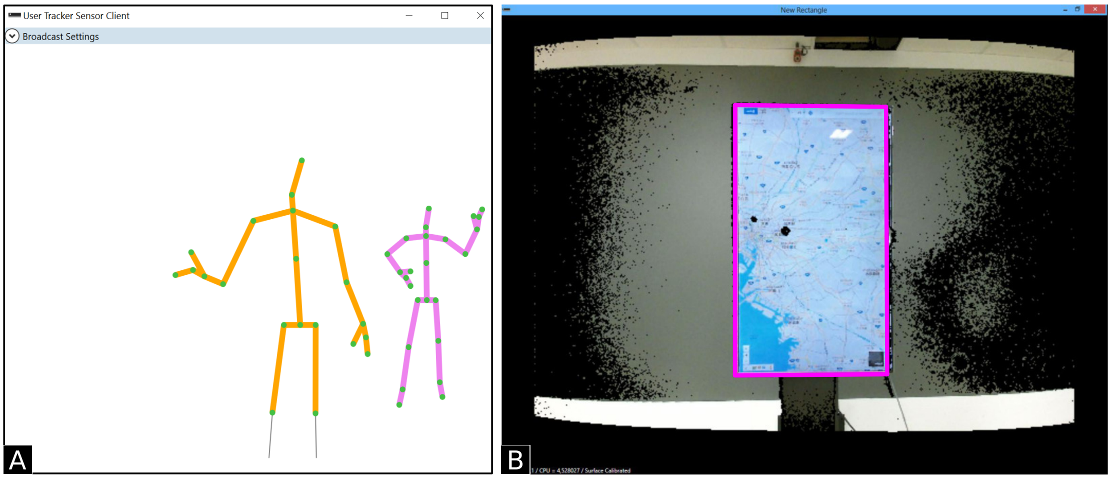

User Tracker Sensor Client
Broadcast Settings
A
New Rectangle
B
1 / CPU = 4,528027 / Surface Calibrated

Figure 4.3: *Creepy Tracker*'s (A) sensor unit and (B) surface calibration application.

currences of false positives common in body tracking models provided by commodity depth cameras. Color and depth data are processed for the point-cloud representation of each person. The body tracking model is broadcast to the Tracker Hub using a UDP stream, while point-clouds are available via a concurrent TCP connection.

4.2.2 Tracker Hub

The Tracker Hub component handles the unified model of the tracked area by combining the data streams from all sensor units. To create a reliable model, the Tracker Hub requires a calibration process to transform all received data into a single coordinate system. Figure 4.3A shows three calibrated sensors with both position and orientation matching the physical cameras. Data received from each of those sensor units will be spatially correct in the unified model's coordinate system. Analogously, a surface calibrated on one depth camera's coordinate system also is transformed to match the area of the physical one, as shown in Figure 4.1A. Since a surface is a collection of four fixed 3D points, the setting up and calibration process needs to occur only once. The Tracker Hub is a Unity3D application that acts as a broadcast server of the unified model to application clients.

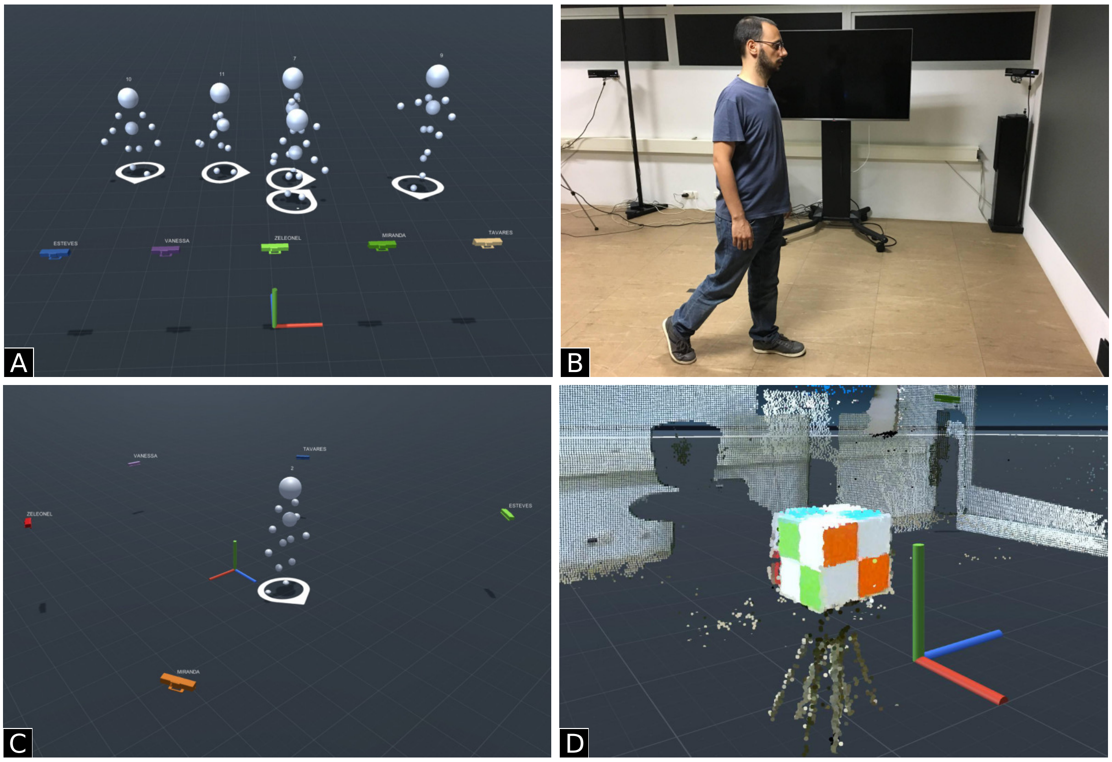


Figure 4.4: Calibration process: (A) center; (B) step forward; (C) result; and (D) calibration cube for manual adjustments.

4.2.3 Calibration Method

A calibration process is required to unify all data streams into a single coordinate system. For this, *Creepy Tracker* relies on the body tracking model of a person from each sensor unit to calculate the new global coordinate system and all cameras' position and orientation. The calibration process requires one standing person to be seen by all sensor units, in two discrete steps. Figure 4.4A shows five uncalibrated sensor units before the calibration process.

Creepy Tracker requires calibration parameters from body tracking models at two distinct locations separated by the distance of a step to calculate the origin and forward and up vectors of the new calibrated coordinate system. In the first step, the position of the person is used to define the origin. The up vector, defined by the spine base and spine shoulder joints of the body model, is also stored, as well as the position of both feet. The second calibration step can be performed after the person moves a step forward (Figure 4.4B). This new position is used in conjunction with the first to define the coordinate system forward vector. The second up vector is averaged with the first to minimize the impact of incorrect poses when calculating the final up vector. Finally, the minimum height according to the up vector of the four feet positions is used to define floor's position. Figure 4.4C shows five calibrated sensors around the coordinate system's origin.

This calibration is usually enough for most interactive scenarios. However, we reckon that a more precise calibration might be needed for more demanding cases. For such situations, we created an additional calibration step. It consists of capturing a depth data frame of each sensor and displaying them using point-clouds, with a simple object placed in the middle of the tracked area. For this, we resort to a cardboard cube with a colored checkerboard in each face (Figure 4.4D). Then, it is possible to manually adjust the position and rotation parameters of each sensor, so that the point clouds match as well as possible.

A new calibration process is required when the setup undergo any adjustment or modification in sensor units.

4.2.4 Adding Surfaces

The *Creepy Tracker Toolkit* also contributes with a standalone C# surface calibration tool (Figure 4.3B). We consider surfaces with three common standard aspect ratios: 4:3, 16:9 and 9:16, for surfaces in portrait mode. Our calibration tool treats a surface as a subspace of an infinite geometric plane restricted by the size and aspect ratio of the physical surface. The surface calibration method requires the mouse click input (on the calibration application) of two corner points from the same edge (surface's bottom left and bottom right) to define size and position, and a third inner point to calculate the surface's normal vector. Only three points spare the need for a depth sensor to have the entirety of the surface in its field-of-view. Figure 4.3B shows a calibrated 9:16 vertical surface with the manually selected points. By knowing the position and orientation of the calibrated sensor from which the surface's points originated, the surface is then transformed to the tracker's coordinate system.

4.2.5 Tracking People

The Tracker Hub formalizes a body tracking model from the sensor unit into a *Body* entity and a person into an instance of *Human*. Overlapping individual body tracking models from different sensor units map into a single person. Consequently, a *Human* preserves a set of *Bodies*, one for each seen by a sensor unit.

When information regarding a new *Body* arrives, the Tracker Hub will try to fit that body in a *Human* within a parameterizable distance threshold. This threshold is set by default to 30 cm, to account for different sensors' perspectives, as it is impossible for two people to have their Spine Base joints closer than this threshold without intimate space violations. The distance is calculated according to Spine Base joints of both *Bodies*. If there is no suited *Human*, a new one is created. When a *Body* is no longer seen by its sensor, it is dissociated from the corresponding *Human*. If a *Human* has no more associated *Bodies*, it will enter a waiting period of 1 second. During that period, if a new *Body* appears within the distance threshold from the *Human*'s last position, it is associated to that *Human*, which exits from the waiting period. Otherwise, if no *Body* is associated with the *Human* until the waiting period expires, the *Human* is removed from the tracker.

Each *Human* entity is constantly choosing the most appropriate *Body* by selecting the one with the highest confidence value. It is not always easy to acknowledge for sure where people are turned to, as some sensors may be facing each other and perceiving different body models for the same person. To overcome this, we follow two approaches. Firstly, we consider a disambiguation pose consisting of having at least one forearm approximately parallel to the floor. The direction one is pointing at, can be used to define that person's forward vector, as it would be both unnatural and very difficult to accomplish such a pose with the arm pointing backward. When this vector's direction is opposite from the current *Human*'s forward, we automatically mirror *Body*'s left and right data. Secondly, as front and back switching occurs mainly when a *Body* from a different sensor is chosen, we also mirror the *Body* when the *Human* is detected to rotate faster (approximately 180 degrees in two consecutive frames) than it is humanly possible.

To deal with the known noisy skeleton information from the Microsoft Kinect, we implemented a double exponential smoothing filter [10]. The filter's parameters can be configured to achieve a compromise between smoothness and added latency. This filter is applied to *Human*'s joints, and helps when dealing with sensor switching in setups with coarse calibrations.

4.2.6 Point-cloud Representations

Using real-time body representations can enhance interactive scenarios for collaborative telepresence [113], computer-assisted rehabilitation [145] and immersive virtual reality [133]. Our approach relies on processing separate streams of point clouds. Each individual sensor unit first captures the skeletal body model data for each person in its field-of-view. We create a point-cloud by combination depth and color values captured by the camera. Then the person's relevant points are segmented from the background. Figure 4.5 shows the resultant streamed body representation. When interacting with applications using the body representation, different cameras will be sending very similar information. However, due to network, processing, and rendering constraints, integration of different streams or redundancy resolution is not performed. We implement a task-oriented decimation, taking into account what body parts are more relevant. To this end, we attribute different priorities to each joint of the available *Body* information according to user-defined parameters. To wit, in

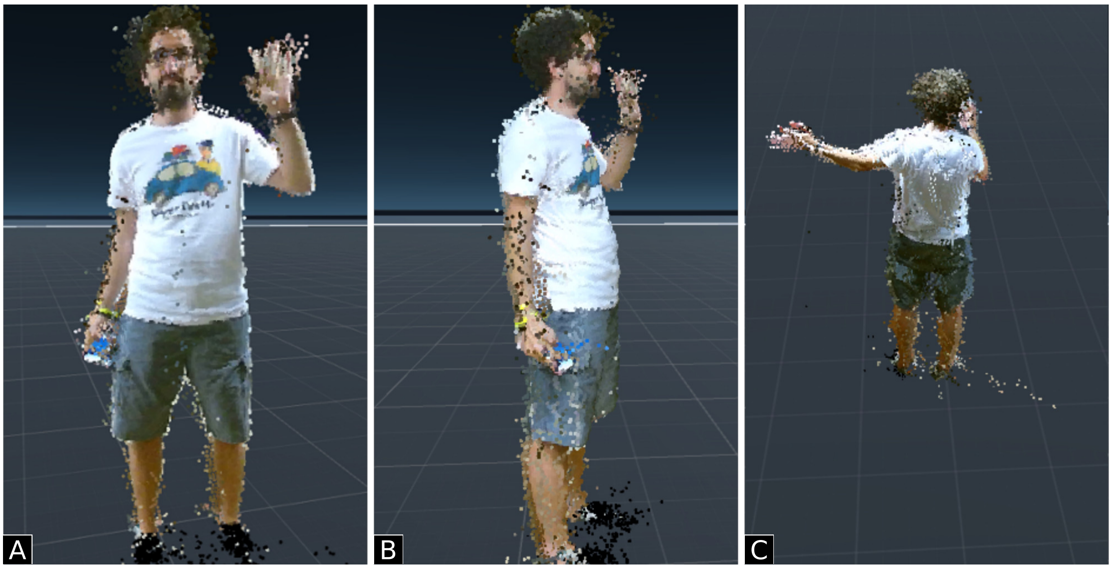


Figure 4.5: Point-cloud body representation: (A) front, (B) side and (C) top views.

collaborative telepresence scenarios, head, face and hand non-verbal communication cues are more relevant than capturing other body parts. Whereas in first-person virtual reality scenarios, detailed hands are more valuable and head information can be totally discarded.

Thus, for each point in the segmented cloud, we calculate its Euclidean distance to the body joints marked as relevant on the sensor units. If this distance is smaller than a threshold value (proportional to the user's body height), this point is marked as a high-quality point for transmission. Points not marked as high-quality, are sampled at a lower frequency to reduce data redundancy on less relevant areas. For each point, sensor units transmit 3D point coordinates, color, and a data bit to indicate high or low quality. This last bit is used to adjust the rendering parameters in the application client. Each coordinate in a 3D point is multiplied by 1000 and rounded to the closest integer, to assure millimeter precision, and packed into two bytes. Data read through the network is then parsed and rendered in the environment using surface-aligned splats at 30 frames per second. While higher quality points are more tightly spaced, requiring smaller splat sizes, we use larger splats in under-sampled regions to create closed surfaces.

4.3 Evaluation

We carried out a performance evaluation against a marker-based infrared tracking system as a baseline. The goals of our evaluation were to determine how the Creepy Tracker differs from a highly accurate marker-based system and determine body tracking consistency using a stress test. Besides, this evaluation serves to inform design decisions when developing future context-aware interfaces.

4.3.1 Design

We devised three different evaluation tasks, depicted in Figure 4.6. For this, a logger application was developed to combine data from the two tracking streams by matching both coordinate systems. The logger was able to record, into a file, sessions containing timestamps per frame, the spatial position from a person's right hand from

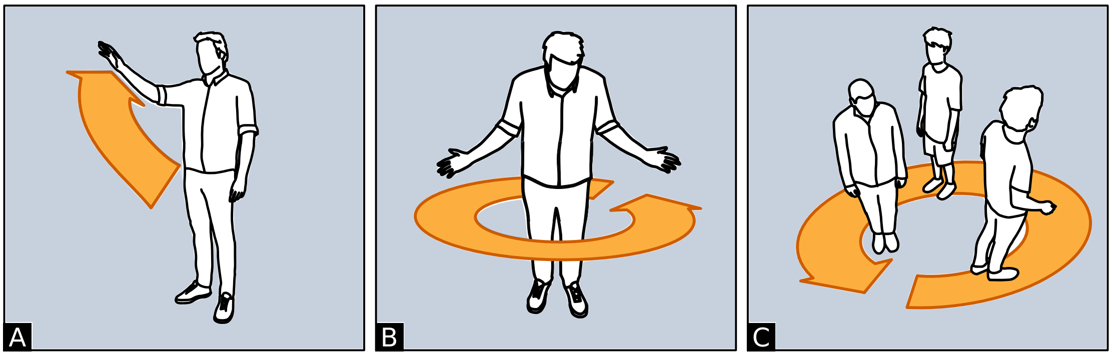


Figure 4.6: Evaluation tasks: (A) hand raise, (B) spin about oneself, and a (C) circular path with multiple people.

our tracker and the position of a rigid-body marker attached to that person's right hand from the baseline tracker.

Therefore, the evaluation consisted of a latency task, and two tasks to measure the accuracy and consistency of our tracker by observing a single person holding an infrared marker. To calculate the average latency we recorded a raising gesture of the right hand five times. Figure 4.6A depicts the first task. The second task, depicted in Figure 4.6B, required a tracked person to fully spin about oneself to test for accuracy, with the purpose of forcing our tracker to switch to different sensors when choosing the adequate body model. Finally, the last task consisted of walking a circular path with two meters of diameter (Figure 4.6B). This required multiple sessions and incrementally adding a different person until the tracked area was congested to the point that the observed person's tracking data was erratic due to body occlusions.

4.3.2 Setup

The evaluation was conducted in a controlled laboratory environment with four by five square meters of tracked area. We employed five Microsoft Kinect version 2 depth cameras, evenly distributed around the room roughly forming a geometric pentagon, as shown in Figure 4.7. For the marker-based tracking system, we resorted to 12 Optitrack Flex 3 infrared cameras evenly placed around the room at 2.30 meters above the floor, providing a tracking volume surrounding the Kinect sensors' setup. To minimize the effects of network communication, both tracking systems and logger applications were running in the same desktop computer. Also, to obtain accurate po-

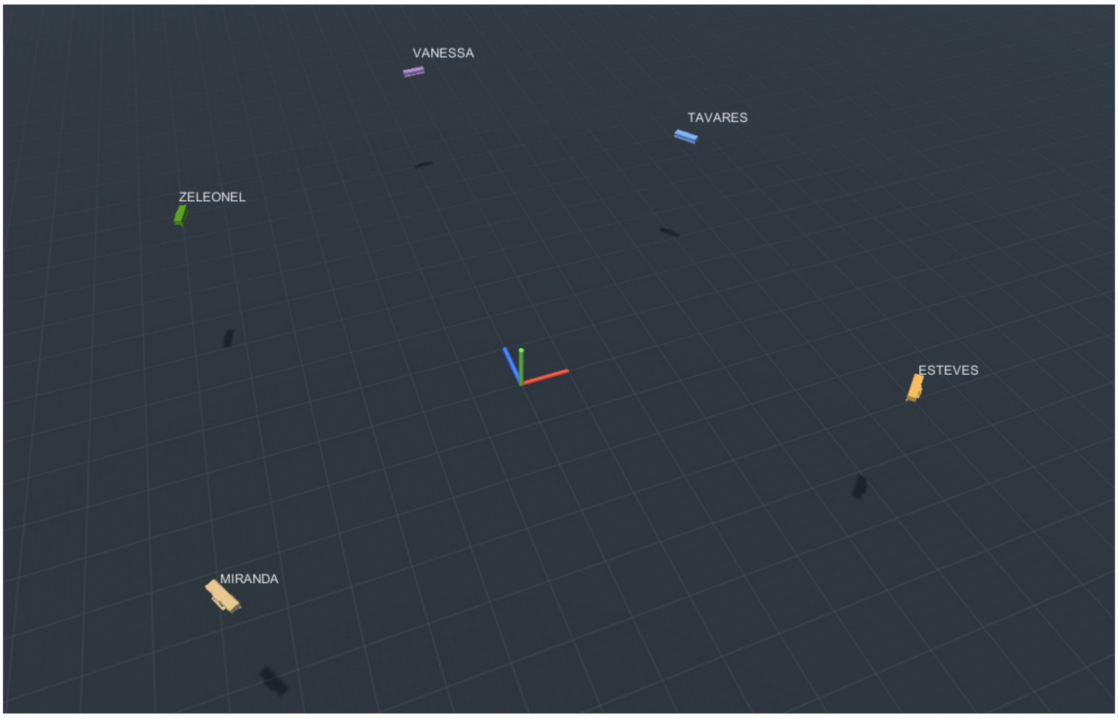


Figure 4.7: Five depth sensors were used in a pentagon formation to capture data for the system's evaluation.

sitional data, all smoothing filters were disabled. Still, all sensor units were remotely streaming data within the same local wired network. We limited Optitrack's send rate to 100 frames per second and *Creepy Tracker* was set to 20 frames per second due to both Kinect and local network limitations.

4.3.3 Results

Using the data thus collected we calculated the latency and position error of *Creepy Tracker* as measured against Optitrack. These allow us to assess how best to explore the flexibility vs accuracy tradeoffs.

Latency

To measure latency we used the logs from the raising hand task, which are depicted in Figure 4.8. Logs had data recorded at an average of 170 frames per second. The difference in send rates justifies the less smooth data of our tracker, as visible in the

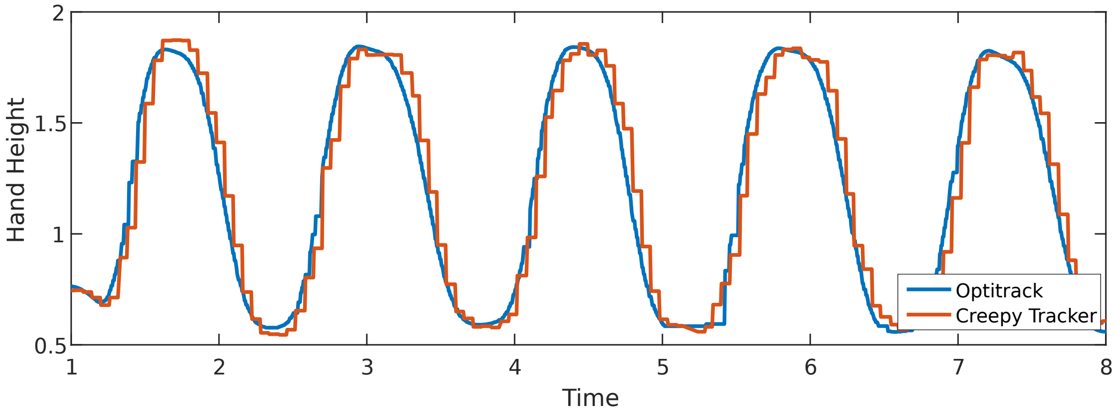


Figure 4.8: Latency comparison between Optitrack and *Creepy Tracker*.

chart. We calculated the difference between local maxima and minima of both trackers. *Creepy Tracker* had a delay averaging 76 milliseconds in comparison to Optitrack.

Although this latency alone might not be enough for real-time performance in virtual reality, combining *Creepy Tracker*'s positional data with the orientation provided by current head-mounted displays appears sufficient to fulfill the illusion of being there.

Accuracy

We compared the hand position from each system, in five different conditions, illustrated in the plots of Figure 4.9. Although the plots are 2D we stored full 3D data. We averaged the Euclidean distance between both trackers' data in each frame. As evidenced in the spinning task (Figure 4.9A), *Creepy Tracker* is capable of accurate results, where switching between sensors is not noticeable. While the average error was of 62mm, we estimate that it is in part due to latency. However, in more demanding scenarios, inaccuracies may arise. In all circular path tasks (Figures 4.9B-E), spikes occasionally occurred. These correspond to sensor switching, where the new sensor perceives the tracked person on his back. While we try to deal with this by mirroring skeleton's information when suited, our approach is not perfect, resulting in right and left side swapping, including hands. This persists until a disambiguating pose is detected with certainty. These spikes are more recurring when tracked people's density increases because sensors with non-optimal point-of-views are used to

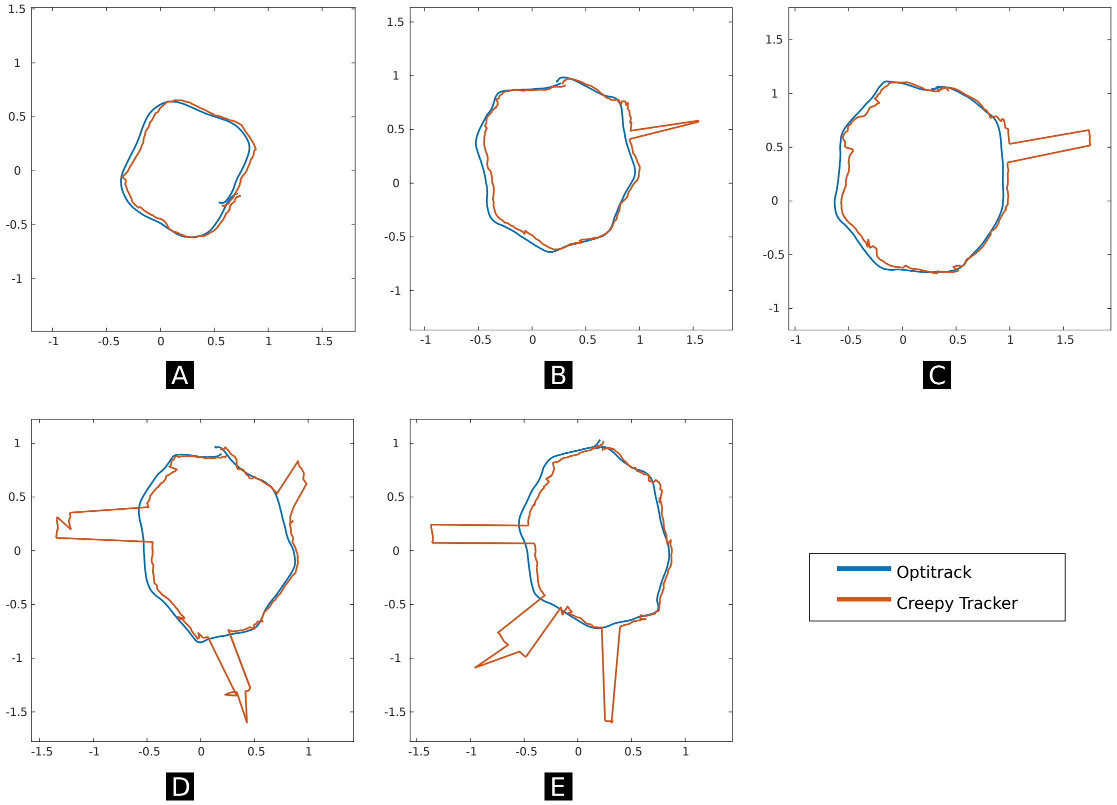


Figure 4.9: Hand tracking data for each condition, and average position error: A) Spinning about oneself (avg error = 62mm); B) Circular path with 1 person (avg error = 80mm); C) Circular path with 2 people (avg error = 93mm); D) Circular path with 3 people (avg error = 134mm); and E) Circular path with 4 people (avg error = 127mm).

circumvent occlusions. When testing with more than four people in the same conditions, we noticed that *Creepy Tracker* sporadically lost the main subject for a brief moment. This originated a new identifier and invalidated the trial session.

4.4 Discussion and Design Considerations

The obtained measurements are sufficient to delineate guidelines and design considerations for applications using the *Creepy Tracker*. Indeed, results of the latency task suggest that our tracker is adequate for context-aware scenarios and for more traditional input modalities, such as pointing techniques. Although, for virtual reality applications, latency is definitely just above the minimum threshold for virtual reality applications. Furthermore, results also show that for context and proximity-aware interactions the accuracy is satisfactory, despite the Kinect's relatively low resolution and noise. Except when using exclusively tracker data for selection tasks, targets should have a radius of at least 15cm. Finally, as demonstrated by the results, increasing the density of people results in an increasingly more inaccurate tracking and sensor switching. To minimize this issue, tracking areas should be calibrated with the sensors evenly arranged in a circular fashion.

4.5 Limitations

Creepy Tracker offers markerless full-body tracking of human and static surfaces in room-scale applications. However, certain limitations need to be taken into account when developing context-aware experiences. Currently, our toolkit models interactive surfaces as planar rectangular static objects, but does not yet support handheld devices. A possible solution is to associate the surface position and orientation to a specific part of a tracked person (e.g., hand). Also, a person's orientation is not always consistent in crowded settings. This is because the skeleton provided by the toolkit will depend on the confidence values reported by each depth camera. As a person moves, different cameras become selected, creating unnatural transitions between frames. This also occurs when multiple people are inside the tracking space, which also leads to inconsistencies when occlusions between users are first detected.

4.6 Conclusions

We developed the *Creepy Tracker* to facilitate researchers designing new interaction techniques for context-aware environments. Our toolkit offers real-time marker-less tracking of people while providing the means of defining the exact position and orientation of interactive surfaces. Enabling a stream of data that can be used to infer people's intent to interact with each other or with interactive surfaces. Performance evaluation against a highly accurate marker-based system shows that our tracker is adequate for building and evaluating context-aware interactions. Despite that, interaction designers need to accommodate tracking errors when considering scenarios that require high accuracy.

4.7 Chapter Summary

In this chapter, we presented the *Creepy Tracker Toolkit* for rapid-prototyping context-aware interfaces. We started by presenting our motivation for investing in this work and proceed to detail how our toolkit operates and the main algorithms used. Our performance evaluation showed that, although slightly less precise than marker-based optical systems, *Creepy Tracker* provides reliable multi-joint tracking without any wearable markers or individual devices. The next chapter features representative scenarios implemented to show that Creepy Tracker is well suited for deploying spatial and context-aware interactive experiences.

Corresponding Publication

Part of the contents of this chapter previously appeared in the following publication:

[P3] Maurício Sousa, Daniel Mendes, Rafael Kuffner Dos Anjos, Daniel Medeiros, Alfredo Ferreira, Alberto Raposo, João Madeiras Pereira, and Joaquim Jorge. **Creepy Tracker Toolkit for Context-aware Interfaces.** In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17). ACM, New York, NY, USA, 191-200. DOI: https://doi.org/10.1145/3132272.3134113

5
Creepy Tracker Toolkit in Action

In this chapter, we focus on the process of developing context-aware interactive scenarios using *Creepy Tracker*. In the following, we describe the tracking data available to construct prototypes and detail setup configurations for such scenarios. Therefore, we first explore the design space of context-aware applications built on top of toolkit's provided data by demonstrating five fundamental application scenarios.

5.1 Using Creepy Tracker

Creepy Tracker offers a client-side C# API with a layer to render network communication transparent and provide updated encapsulated abstractions of tracking data. Figure 5.1 shows the data model put up by the API.

An independent tracker client, upon connection, continuously receives a list of *Humans* and can request at any time a list of available *Surfaces*. A *Human* is a repre-

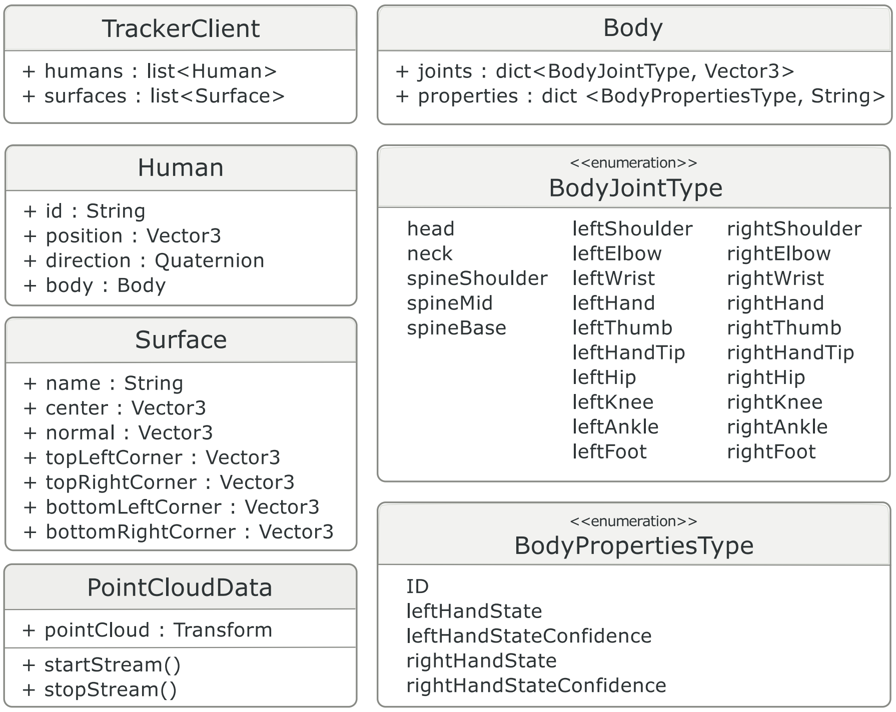


Figure 5.1: *Creepy Tracker* client-side API classes + properties.

sentation of a real person in tracked area. It holds an unique identification provided by the tracker, a point in space correspondent to the person's position, a client-side calculation of the person's direction and a list of all body joints. This is specially useful a application scenario requires the spatial distribution and proximity relationships between people. We defined a surface as a geometric plane. The *Surface* entity is a combination of a center point, a normal vector and the four vertexes to define the surface's edges. Making it easy to calculate if a person is near, approaching or facing the surface's front side.

5.2 Sample Scenarios

To demonstrate how to exploit the tracker's API and resources, we develop five different sample scenarios. The Video Figure 1 features the developed scenarios. Next, we introduce each scenario and describe the development process from interaction design to setup using our toolkit.

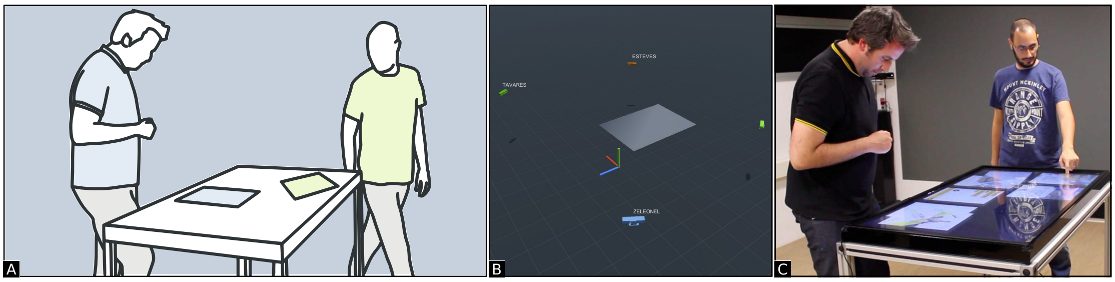


Figure 5.2: Tabletop context-aware example: A) idealized interaction design, B) calibration setup and C) the final prototype.

5.2.1 Tabletops

Multitouch tabletops with large screens enable simultaneous interactions from multiple people and foster collaboration by allowing individual interactions with common content [58, 102]. To explore this scenario, we devised an interactive tabletop application that enables people to take temporary ownership of digital content. Figure 5.2A depicts the interaction design rational for a tabletop application that utilizes context-aware information provided by *Creepy Tracker*. Users can get hold of digital content by touching it and the selected content starts following the person around until another touch breaks the temporary lock. For this, four depth cameras were distributed around the room with a calibrated surface right in the center of the tracked area, as shown in Figure 5.2B. Since a tabletop can provide multitouch inputs, associating content to a person during a selection requires the distance of all the users' hands spatial position to the point of touch. The person with the nearest hand takes ownership of the digital content. From here, the digital content follows the user around, until another touch gesture is detected.

5.2.2 Wall-sized Displays

Bolt's Put-that-there [22] multimodal interface is a canonical approach to interact with objects in large scale displays. Put-that-there combines pointing gestures with speech recognition to select objects and define target locations. We designed an adaptation of Bolt's seminal system using body pose information to create and move objects with different shapes and colors, as depicted in Figure 5.3A. To avoid the occlusions in front of the large display, we opted for placing depth cameras on both

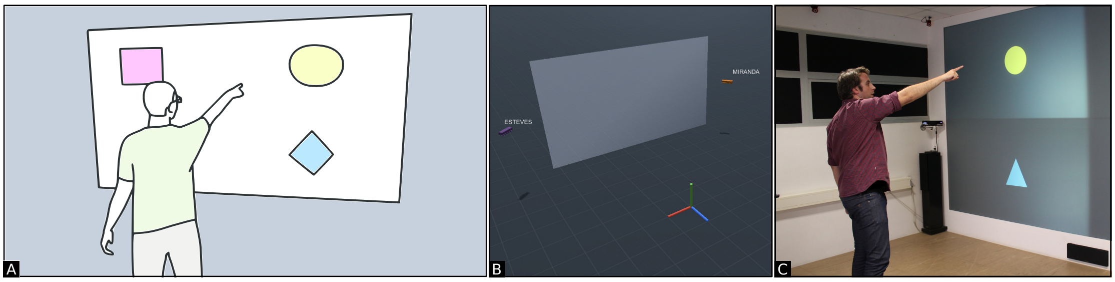


Figure 5.3: Bolt's Put-That-There with the *Creepy Tracker*: A) idealized interaction design, B) calibration setup and C) the final prototype.

sides of the screen, as demonstrated in Figure 5.3B. Instead of a laser, we utilized the image-plane pointing technique [114]. Therefore, when the tracker client detects a relevant utterance (”that” or ”there”), intersecting the vector, that starts in the user's head through the raised hand, with the surface plane, the target position can be determined.

5.2.3 Floor Projected Surface

Using large floor projected interfaces provide sufficient space to promote interactive visualizations [8, 151] and shared social user experiences [51, 53]. We design a playful projection-based collaborative game based on the classic arcade space shooter *Asteroids*, released by ATARI, Inc in 1979. A player can use their own body position and orientation to destroy asteroids, as depicted in Figure 5.4A. While the projected floor surface serves to display the game view and players can control a spaceship placed exactly below them, automatically firing space bullets in the users' forwards direction at

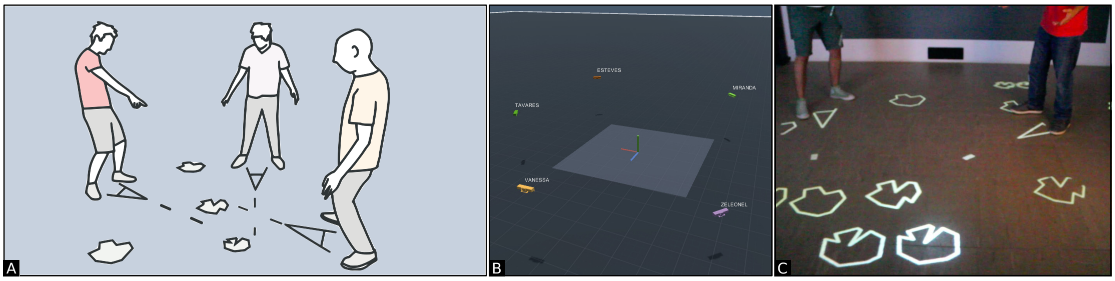


Figure 5.4: *Asteroids* game on the floor: A) idealized interaction design, B) calibration setup and C) the final prototype.

a fixed rate. Figure 5.4B shows the tracker's setup with five cameras and a calibrated 4:3 surface on the floor. For this experience, we resorted to *Human*'s position and orientation for steering the ship and utilized the *Surface* to map the players' positions.

5.2.4 Telepresence Portals

Taking inspiration from the Office of the Future by Raskar et al. [118], we designed a telepresence experience to connect two remote locations. Indeed, depth sensors have already been used to create highly realistic avatars in remote collaboration scenarios [111], but with custom hardware and cluster-based rendering.

Creepy Tracker can be used to show real-time realistic user representations using commodity hardware, easing the development of approaches similar to the work of Beck et al. [17]. This approach allows for verbal and non-verbal communication, and at the same time, creates a seamless visual continuity from the local to the remote location, as depicted in Figure 5.5A.

The portals implicitly establish or cease the link between them by allowing a person to transition between a non-interactive space to a explicit interactive space, similarly to Vogel et al. [152]. Therefore, to initiate link between two portals, someone simply needs to walk into close proximity of the surface, triggering a presence notification on the other side. Analogously, the receiver of the notification can proceed by walking towards the portal to accept the connection. For this, sensors were placed on each side of the surface to capture a point-cloud of the user, combining information from both sides of the body, as shown in Figure 5.5. Then, the setup was replicated in another remote location. The distance between *Humans* and *Surfaces* were used to

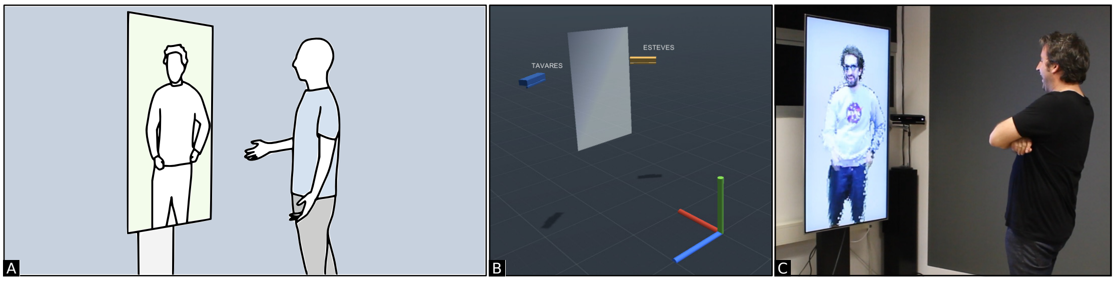


Figure 5.5: Telepresence scenario: A) idealized interaction design, B) calibration setup and C) the final prototype.

established and destroy the communication link. While *PointCloudData* was used to start streaming remote people. Furthermore, the setup snd prototype used in Chapter 6 was constructed upon this telepresence scenario.

5.2.5 Virtual Reality Interactions

Multiple depth sensors setups have been deployed to capture full body data to animate generic avatars and explore novel interaction techniques in virtual reality [105]. Using several sensors allow users to freely navigate the interactive space. In such immersive virtual environments, realistic self representation enhance the perception of being there [133]. We designed a virtual reality gaming experience that takes advantage of full body point-cloud representations aided by body joints' positions. Therefore, we idealized a gaming zone where people have to catch basketballs thrown at them by three surrounding cannons. Figure 5.6A depicts a person trying to catch a basketball. We placed five cameras around a room for maximum body coverage as demonstrated in Figure 5.6B. *PointCloudData* was used to render the user's body and the tracker was set up to ignore the user's head when streaming. Human's hand joints were used to distinguish touching on a basketball. Still, all joints were used to build the person's bounding volume for the basketballs to collide and bounce back. Also, a *Surface* was used to define the gaming area.

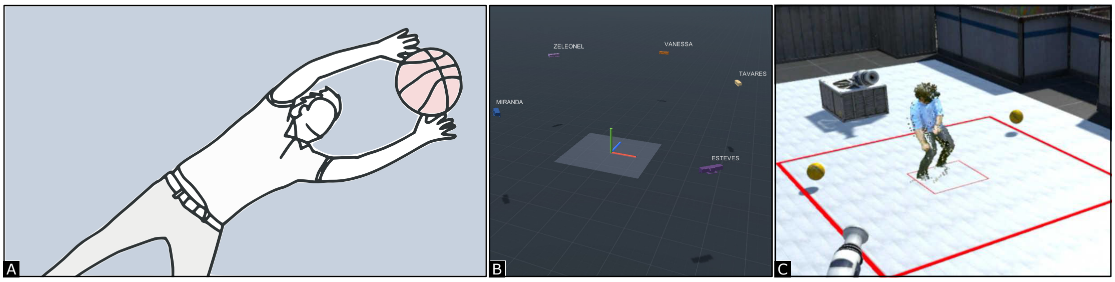


Figure 5.6: Virtual Reality game: A) idealized interaction design, B) calibration setup and C) the final prototype.

5.3 Chapter Summary

In this chapter, we centered on demonstrating the Creepy Tracker's capabilities and detailed how-to prototype new interactive experiences. For this, we presented representative scenarios show that Creepy Tracker is well suited for deploying spatial and context-aware interactive experiences. This chapter concludes Part II. In the next part of this dissertation, we present the steps taken to investigate our vision for face-to-face remote collaboration in shared 3D spaces. The following evaluations heavily exploited the Creepy Tracker's aptitudes to support remote embodied experiences.

Corresponding Publication

Part of the contents of this chapter previously appeared in the following publication:

[P3] Maurício Sousa, Daniel Mendes, Rafael Kuffner Dos Anjos, Daniel Medeiros, Alfredo Ferreira, Alberto Raposo, João Madeiras Pereira, and Joaquim Jorge. **Creepy Tracker Toolkit for Context-aware Interfaces.** In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17). ACM, New York, NY, USA, 191-200. DOI: https://doi.org/10.1145/3132272.3134113

Part III

Exploiting Perception Manipulation to Enhance Workspace Awareness

The third part of this dissertation aggregates the methods and experiments completed to validate the concept of using multiple perception manipulation techniques to enable face-to-face remote collaboration in shared 3D workspaces. First, in Chapter 6, we investigate the impact of using different workspace conditions on workspace awareness. Next, in Chapter 7, we propose a novel approach to improve the perception of deictic gestures using body warping. Last, in Chapter 8, we evaluate the combination of body and environment warping to improve workspace awareness in a face-to-face collaborative telepresence scenario.

6

Investigating Workspace Awareness in 3D Face-to-Face Remote Collaboration

In this chapter, we investigate how people and content should be presented for discussing 3D renderings within face-to-face collaborative sessions. To this end, we present a user evaluation to comparing different manipulations of workspace and remote people representation. Furthermore, we present a new design space, the Negative Space, for remote face-to-face collaboration focusing on 3D content.

6.1 Motivation

Videoconferencing and telepresence allow for virtual encounters to take place and expedite the communication between geographically separated people. Modern technologies add unproductive layers of protocol to the flow of communication between remote participants, rendering interactions far from seamless. Despite being widely adopted, traditional videoconferencing approaches still uphold partial viewing or down-scaled representations of remote people that limits the sense of co-presence [60].

Videoconferencing systems using real size portrayal of people become closer to a co-located experience. In fact, studies suggest that full-body face-to-face communication improves task completion time, presence, and efficiency of communication [166, 113], while enabling nonverbal visual cues including posture, proxemics and deictic gestures in addition to the usual speech and facial expressions currently supported by commercial approaches. Hence, people should rely on natural communication, verbal and nonverbal, to convey the focus of the collaboration and pinpoint details on shared content as if they were physically co-located.

When designing for face-to-face collaboration, it is necessary to take into account how to address interactions in shared task spaces. Despite being typically considered separate from the person space, Ishii et al. [68] suggest that an integration of both task and person spaces should be employed when considering face-to-face meetings using a transparent display metaphor. Indeed, with transparent displays, two participants can see one another and share digital content, rendered between them, that both can jointly manipulate In everyday face-to-face interactions mediated by screens, people have no standard orientation of right or left. Clearboard [68] addresses this issue by mirror-reversing the remote person's video stream, producing gaze and pointing awareness since 2D graphics and text can be corrected to the participant's point-of-view. This approach has been the subject of research for 2D content collaborative manipulation [164, 91, 166]. However, 3D digital content gives rise to detracting issues that affect and impair workspace awareness. Participants do not share the same *forward-backwards* orientation, and occlusions can affect the understanding of where or what the remote person is pointing. Also, contrary points-of-view can result in different perceptions or even serious communication missteps, as illustrated in Figure 3.2 in Chapter 3.

This work focuses on assessing workspace awareness using variations of the shared workspace settings, individual point-of-view and remote user's representation. For this purpose, we conducted an evaluation comparing task performance and user preferences under four different conditions. We employed an evaluation environment inspired by the *"portal to a distant office"* concept from Wen et al. [157] creating a virtual space between two real spaces. Unlike Wen et al. [157] our collaborative approach provides the three-dimensionality to populate the space between remote rooms with 3D digital content. From the results, we conceptualize the *Negative Space*, an approach to face-to-face remote collaboration, creating a shared virtual workspace linking two physical remote spaces, while providing a sandbox for interacting with 3D content. Therefore, the contributions of our work are:

1. the proposal of *Negative Space* as a new shared space concept where remote 3D interactions can occur;
2. and an assessment of workspace awareness in face-to-face remote collaboration using different collaboration conditions.

6.2 Negative Space

Previous research on remote face-to-face collaboration has successfully contributed full-body telepresence approaches with integrated person-task spaces, which demonstrated considerable improvements on presence and cooperative task performance.

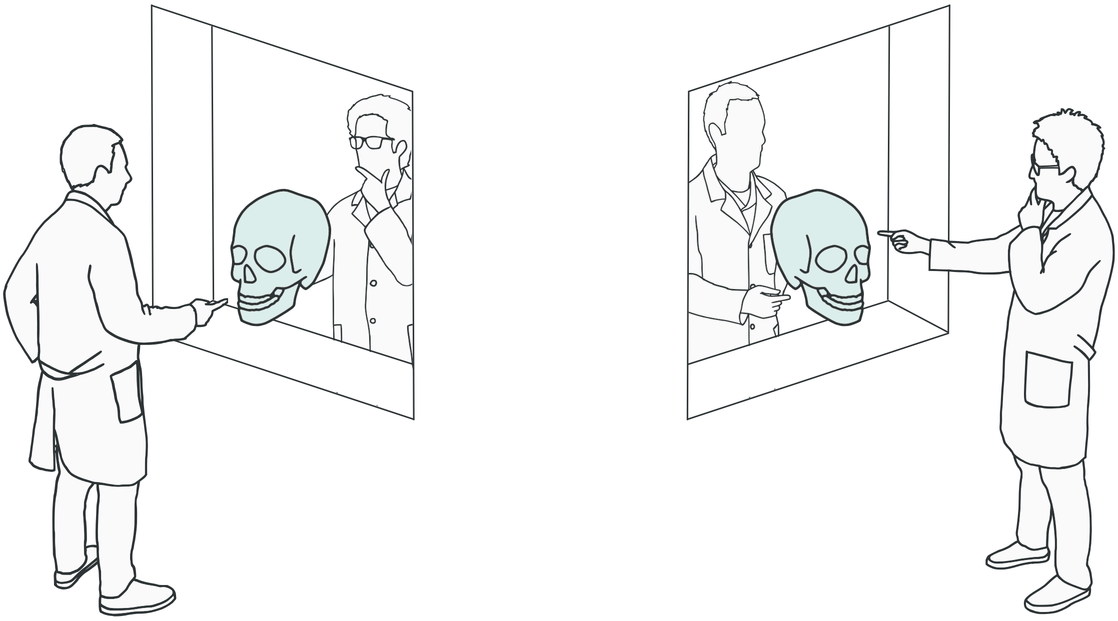


Figure 6.1: Conceptual vision of Negative Space featuring two remote locations. Participants have identical points-of-view over exact copies of the workspace.

Most focus on cooperative interactions with 2D content, although collaboration in design and review of 3D virtual models is crucial in several domains.

With this in mind, we introduce *Negative Space* as a conceptual platform with a set of rules for future works on remote collaboration. It is characterized as a virtual space that serves as a gateway between two physical rooms where collaborative 3D interactions can occur, as depicted in Figure 6.1. From the evaluation results presented in the previous section, we devised *Negative Space* as a medium to support discussions on shared views over the 3D content and, as such, it shall offer participants *Identical* points-of-view over *Exact* copies of the workspace. We also enforce the use of real-time 3D reconstructions of remote people for improved perception. Our approach can be advantageous in avoiding communication breakdowns by making many gestures and deictic idioms easier to share and understand between participants.

Negative Space exploits the benefits of a see-through display. By positioning the virtual content between two people, participants are able to profit from normal face-to-

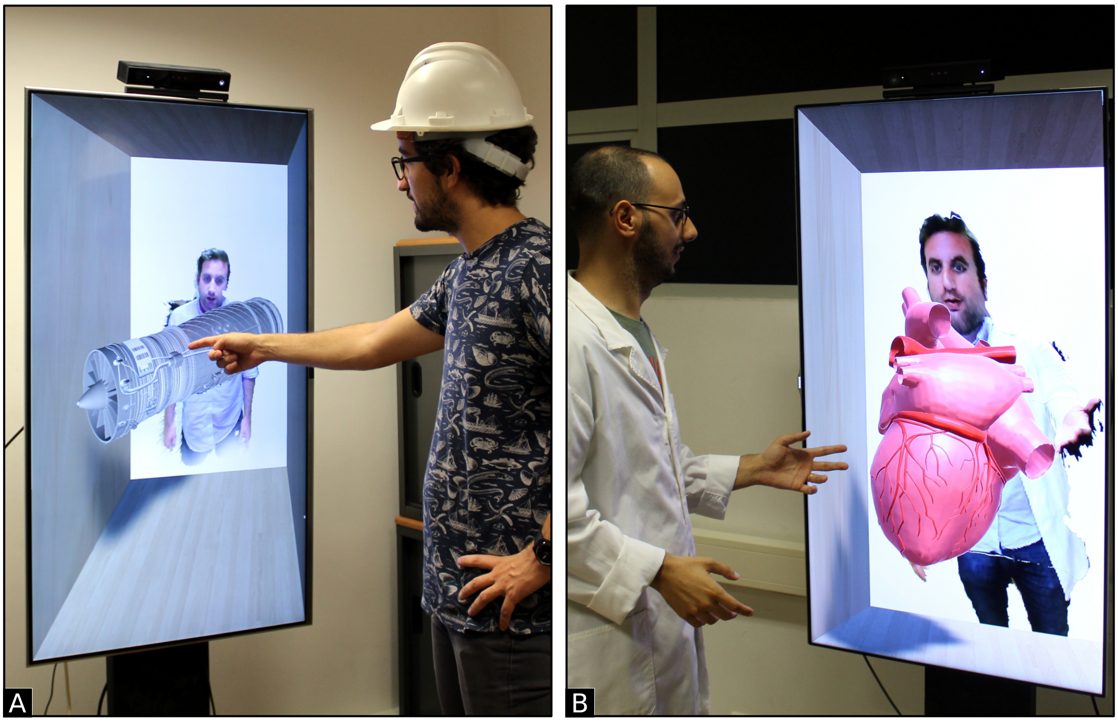


Figure 6.2: The *Negative Space* concept can be applied in multiple usage scenarios requiring visualization, design and review of virtual 3D models. Notable examples are A) engineering industries and the B) health-care.

face interactions as if they were physically co-located. Our approach differs from Ishii and Kobayashi [68] because of the implementation of a "Above a table" metaphor to allow for interactions with 3D content. This contributes to the overall workspace and situational awareness since participants are able to observe the other person's gaze direction, deictic gestures and actions, while performing selection and manipulation tasks related to multiple occupational fields, such as engineering, industrial, architecture and medical, as demonstrated in Figure 6.2.

6.3 Evaluation

We set out to assess if different manipulations of the person-task space can enhance workspace awareness and the way people communicate, when collaborating in a face-to-face setting with 3D content. We developed a full body telepresence prototype and implemented four different workspace conditions. This evaluation was based on both task performance and user preferences. For this, we designed a collaborative 3D assembly task where an *Instructor* guides a remote *Assembler* to reach the correct solution of a toy problem using cubes.

6.3.1 Evaluation Conditions

To investigate workspace awareness we devised four different evaluation conditions. We performed several combinations of person space and task space. Each combination affects differently the perception of the reference space. Our goal was to study the participants' *point-of-view*, remote participant's *embodiment* and *workspace* rendering, as depicted in Figure 6.3.

For *point-of-view* we considered that participants could observe workspace in usual opposing points-of-view or simulating an identical viewing experience. Also, similarly to Ishii et al. [70], *embodiment* and *workspace* variables could both be horizontally inverted or not. Handling the permutations of three variables should produce eight different conditions. Yet, for this evaluation, we argue beforehand that any combination with an opposing point-of-view other than the real world case, suffers absolutely from worse awareness issues with no improving benefits. Naturally, even in the real world scenario, opposing point-of-views can create communication issues,

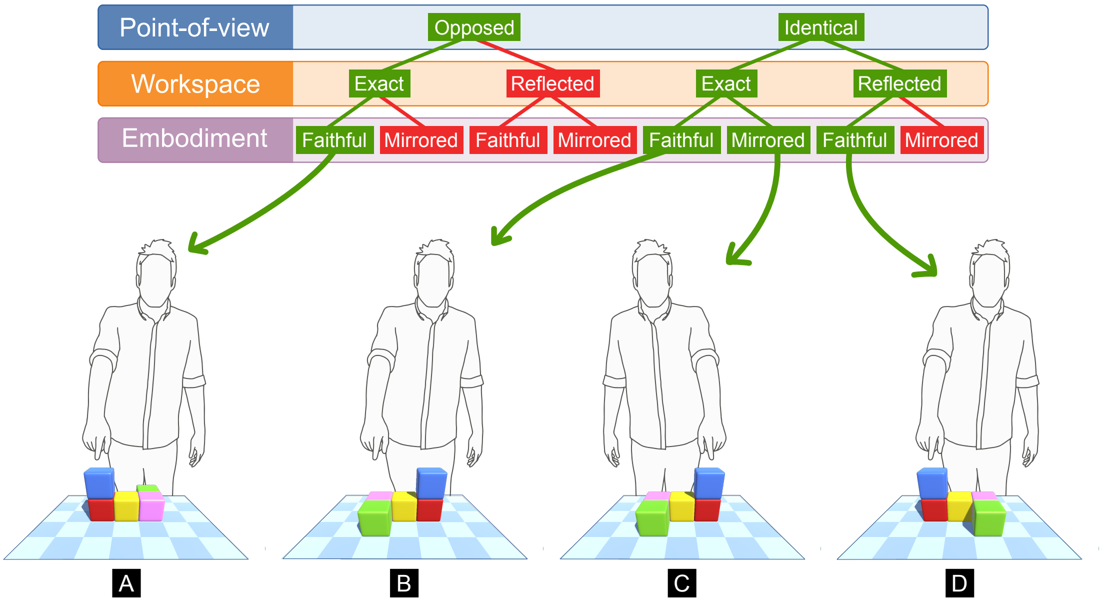


Figure 6.3: Workspace conditions from the Assembler's perspective: A) Real life face-to-face (*RL*), B) Simulated side-by-side (*SS*), C) Mirrored Person (*MP*), and D) Mirrored workspace (*MW*). In all cases, the Instructor is pointing with his right index finger to the blue cube.

but we decided to maintain it in our evaluation to act as a baseline. We also did not consider the reflected workspace with mirrored embodiment condition, since neither verbal nor deictic gestures match any real reference frame. Therefore, our evaluation followed a within subjects design with four conditions:

1. **Real Life Face-to-face (*RL*)** – Derived from the real world face-to-face scenario, both participants can see each other and the workspace as if they were in opposite ends. As such, the reference space should be natural for the participants since this condition match everyday face-to-face interactions. However, the participants have contrary points-of-view and cannot observe the workspace opposite side, as demonstrated in Figure 6.3.A.

2. **Simulated Side-by-side (*SS*)** – While remaining face-to-face in regard to the embodied representation, participants share the same point-of-view of the workspace, in a way that simulates a side-by-side approach. Participants can perceive the workspace from the same side and use verbal relative directions, but pointing gestures from the instructor do not match the reference space (Figure 6.3.B).

3. **Mirrored Person (*MP*)** – Participants share the same point-of-view, yet the instructor's embodied representation is horizontally inverted to match the reference space. Despite the assembler perceives a mirror embodiment of the instructor, both deictic gestures and verbal relative directions match, as depicted in Figure 6.3.C.

4. **Mirrored Workspace (*MW*)** – With an identical point-of-view, participants also share faithful face-to-face embodiment representations of each other, although the assembler's workspace is horizontally inverted. Thus, deictic gestures can be used to reference a point. Yet, participants have to accommodate the fact that any verbal relative direction is in reverse (Figure 6.3.D).

6.3.2 Method

Participants were grouped in pairs and were asked to perform a set of four tasks, one with each condition, where one participant played the role of the Instructor and the other the Assembler. After completing the four tasks, they were asked to switch roles for another four tasks. All sessions followed the same structure and lasted approximately 50 minutes in total: 10 for the introductory briefing, 20 minutes for the first set of four tasks, and, finally, more 20 minutes for the last set of tasks after participants had switch roles.

We started by introducing the experiment procedure to each pair of participants, followed by a brief description of the interaction technique. Each participant was then randomly assigned his initial role, and was conducted to each individual room. Afterwards, participants jointly executed the evaluation tasks, described in the next session, each task with a different workspace condition.

All tasks were preceded by a training session to familiarize participants with the current condition, and where the assembler had the opportunity to learn how to select and move the checkerboard's blocks. Also, to avoid biased results, in all evaluation sessions the order of the workspace conditions and tasks were performed in an alternated order. Moreover, we devised eight different puzzles to assure that no pair of participants would experience the same puzzle twice. At the end of each task, both participants were asked to fill up a user preferences questionnaire. The evaluation session concluded with profiling questionnaires.

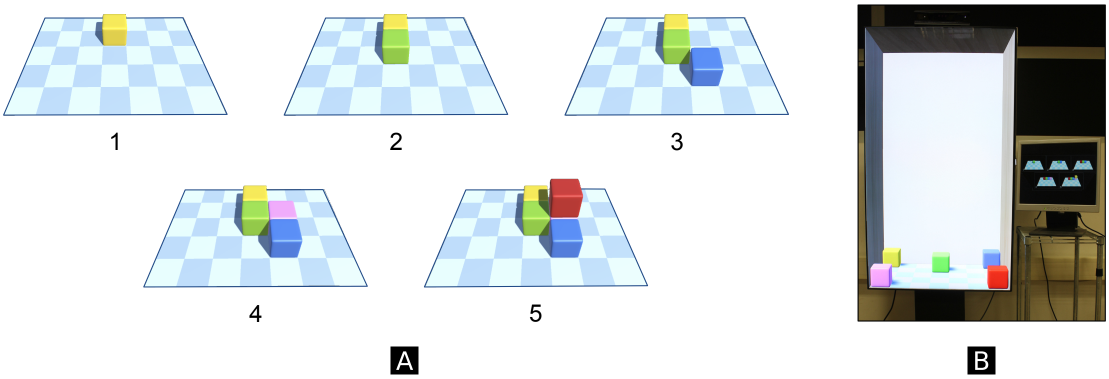


Figure 6.4: *Instructors* had access to A) step-by-step solutions for each task in B) a separate display.

6.3.3 Tasks

All tasks consisted in solving a block-based puzzle with five colored cubes on top of a checkerboard, where the instructor helps the assembler completing the puzzle using verbal and non-verbal communication cues. That is, instructor's actions and gestures were not augmented by technology and he was not allowed to interact with the virtual task space. For this, a step-by-step description on how to reach the puzzle's solution was provided. Figure 6.4 shows the multiple steps to complete a puzzle and the position of the instructions' screen. Also, only the instructor could see the colors of the cubes, while for the assembler all were rendered in a neutral gray color, as shown in Figure 6.5.

All tasks started with the first cube already placed in the correct initial position, while the remaining were randomly placed on the corners of the checkerboard. The instructor's duty was to make it clear to the assembler which cube to pick up next and where to place it, using speech and gestures. The interaction between both participants concluded when all cubes were correctly positioned according to the puzzle's

Video Figure 2. Workspace conditions evaluation task example.
http://web.ist.utl.pt/~antonio.sousa/videos/ch4-task-example.mp4
(File size: 6.0 MB)

solution. The screen was purposely faded to black signaling the successful ending of a task. To ensure the same complexity between all tasks, the designed puzzles' solutions followed the same set of rules. This was also applied to the training task.

6.3.4 Setup and Prototype

The evaluation environment consisted of two identical setups replicated in physically separated rooms. Each setup was comprised of an interactive surface (a 55 inch display in portrait mode), two Microsoft Kinect v2, one mounted on top of the surface facing the participant and another one for calibration purposes, and the instructor had a Samsung S7 smartphone for input, as shown in Figure 6.5 and in Video Figure 2.

We employed a non-intrusive open-source user tracker [135] to combine the interactive surface and the user's body positional data into the same coordinate system. The user's virtual representation was composed of a 3D mesh using color and depth values obtained from the front facing sensor. The prototype was developed in Unity3D and both separated setups were connected in using a network server using TCP connections for the remote user's representation and workspace synchronization. Also, UDP was used to stream the smartphones orientation and button click events over Wi-Fi.

The virtual workspace was constructed inside the closed geometric volume connecting both interactive surfaces to simulate a tunnel between them. We employed a depth of 50 centimeters to accommodate the evaluation's checkerboard. To preserve the illusion of a tunnel between both spaces, the local and remote coordinate spaces were combined, so distances between instructor and assembler were spatially correlated to real distances. We also employed motion parallax by combining a perspective projection [83] with the user's head position retrieved from the user tracker. Using a single shared coordinate system to preserve the real distances and motion parallax promotes the necessary depth cues to convey meaning for the participants' non-verbal cues and deictic gestures. We used the Kinect's own array of microphones and the display's built-in audio speakers to establish audio communication.

The checkerboard was placed on top of the lower plane of the workspace, as shown in Figure 6.5. To move the board's blocks, we resorted to a *pick* and *drop* approach using a cursor on the screen plane. The assembler could only select one cube at a

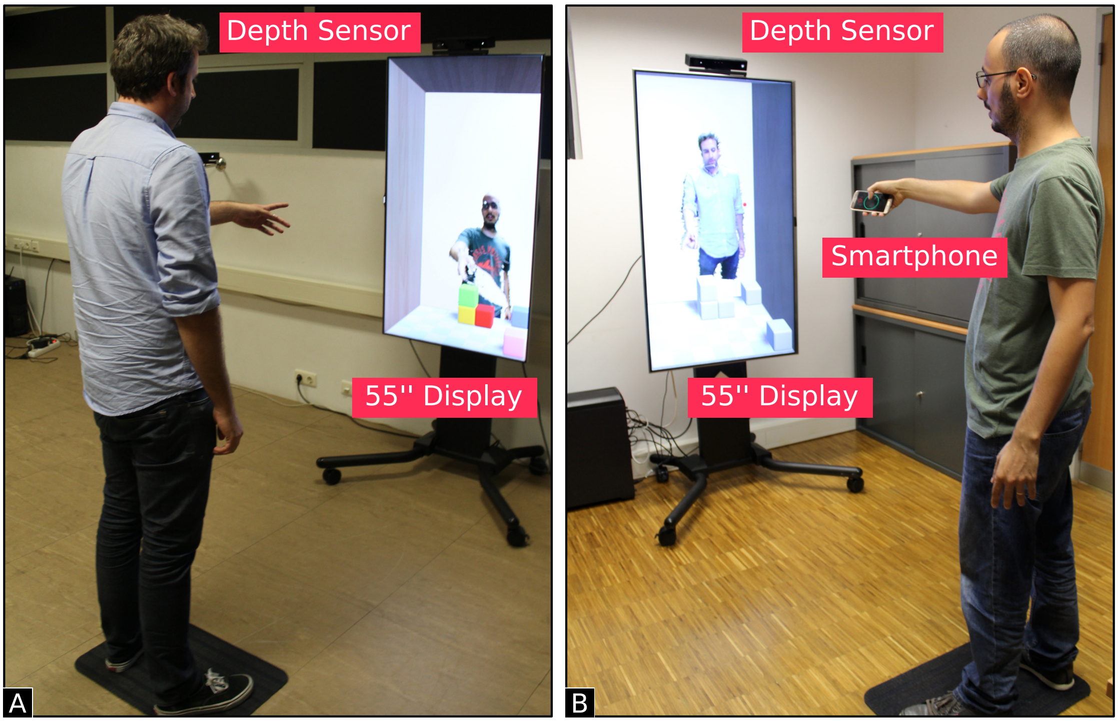


Figure 6.5: For the evaluation trials, A) *Instructor* and B) *Assembler* shared similar setups replicated in two separate rooms.

time by pointing at it. Highlights were activated when the assembler was selecting an object. Selection highlights were shared between both participants. For pointing, we employed the Laser technique [77] using the spatial position of the participant's hand given, combined with the smartphone's orientation from the built-in gyroscope sensor. Then using a ray cast approach, the system was able to determine were the participant was pointing at. Also, the cursor on the screen utilized the participant head to determine the position on the screen plane, to appear on top of the intended location in the workspace.

6.3.5 Apparatus and Participants

The evaluation trials were performed in two separate rooms in a controlled laboratory environment. Each participant was accompanied by an evaluation moderator. They were instructed to start all tasks on top of a floor mat (Figure 6.5), positioned at one meter from the display, to preserve the fidelity quality of the participant's embodiment. Despite that, they could freely move around when executing the evaluation tasks.

Our evaluation counted with 16 participants, divided into eight pairs of people. From which, 12 participants were male and four female, and with the great majority (approximately 94%) were between 18 and 35 years old. Most had at least a BSc degree (89%). All participants did not exhibit any color vision deficiency after performing a standard Ishihara test [67] with nine different plates.

6.4 Results and Discussion

During the evaluation trials we collected *Task Performance* data through logging, and gathered *User Preferences* from questionnaires filled up after the execution of each task. To perform the statistical analysis, we firstly used Shapiro-Wilk test to assess data normality. For the evaluation conditions, we ran the repeated measures ANOVA test to find significant differences in normal distributed data, and Friedman non-parametric test with Wilcoxon-Signed Ranks post-hoc test otherwise. To test for influence of puzzle complexity in tasks' performance, we employed a One-Way ANOVA. In all cases, post-hoc tests used a Bonferroni correction.

6.4.1 Task Performance Overview

We logged completion times, number of wrong cube selections and wrong cube placements. To certify that all eight puzzles presented similar complexity, we tested their times with a One-Way ANOVA, which showed that they were indeed no significantly different (F(7,53)=1.426, p=.215), meaning that different puzzles did not affect in any way the task performance. Figure 6.6 shows tasks' completion times for each condition. Although it appears to be a tendency for lower times with the *MP* condition, no statistically significant differences between the four workspace conditions were found (F(2.218,28.831)=1.981, p=.152). Wrong cube selections and placements are reported in Table 6.1. Again, no statistically significant differences were found for either wrong selections (χ^2(3)=1.719, p=.633) or placements (χ^2(3)=2.038, p=.565).

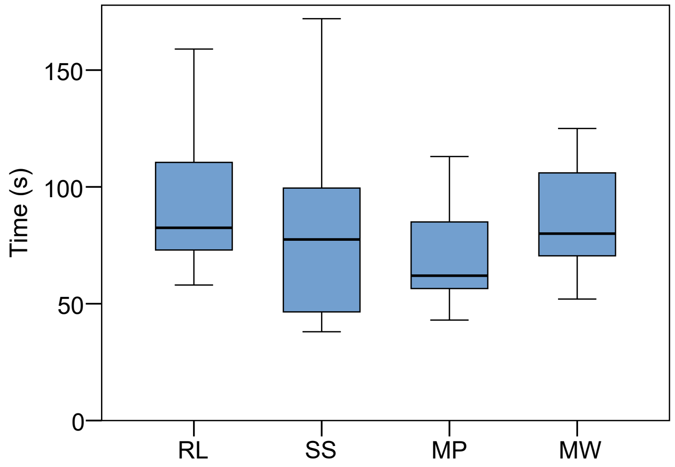


Figure 6.6: Tasks' completion times for each condition.

6.4.2 User Preferences Overview

After the completion of each task, participants were asked to fill up a preferences questionnaire related to the condition they just experimented. Table 6.2 shows prompted questions on the user preferences questionnaire with results for both assembler and instructor in all four conditions. Statistical significant differences were found on three questions for the instructor (Q1:($\chi^2(3)$=10.892, p=.012); Q3: ($\chi^2(3)$=11.598, p=.009); Q4: ($\chi^2(3)$=10.102, p=.018)). The post-hoc test revealed that participants in the instructor's role strongly agreed that *MW* was more difficult than *SS* (Z=-2.743, p=.006) and *MP* (Z=-2.722, p=.006). Instructors agreed that in the *MP* condition, explaining the row of the cube to select, is easier than *MW* (Z=-2.967, p=.003). It was also easier for instructors to explain the column of the next cube to be selected in the *MP* condition than *MW* (Z=-2.675, p=.007). We

	Wrong Selections	Wrong Placements
RL	0.5 (1.25)	2 (1.5)
SS	0 (0.75)	1 (2)
MP	0 (0.25)	1 (1.25)
MW	0 (1.25)	1 (2)

* indicates statistical significance

Table 6.1: Tasks' number of wrong selections and placements (Median, Inter-quartile Range).

	Instructor				*Assembler*			
It was easy to...	*RL*	*SS*	*MP*	*MW*	*RL*	*SS*	*MP*	*MW*
[Q1]...complete the task.*	5.5 (1)	6 (1)	6 (1)	5 (1)	5 (1.25)	6 (1)	5 (0.25)	5 (2)
[Q2]...explain/understand which cube to select.	6 (1)	6 (1)	5.5 (1)	5 (1)	5 (0.25)	6 (1)	5.5 (1)	5 (2)
[Q3]...explain/understand the row of the cube to select.*	6 (1)	6 (1.25)	6 (0.25)	5 (0.25)	5 (0.25)	5.5 (1)	5 (1)	5 (2.25)
[Q4]...explain/understand the column of the cube to select.*	5.5 (1.25)	6 (1)	6 (0.25)	5 (2.25)	5 (0.25)	6 (1)	5.5 (1)	5 (1)
[Q5]...explain/understand where to position the cube.	5 (1.25)	6 (1)	5.5 (1)	5 (1.25)	5 (1.25)	5.5 (1)	5 (1.25)	5 (1)
[Q6]...explain/understand the row to position the cube.	5 (1)	6 (1.25)	6 (1.25)	5 (1.25)	5 (2)	5.5 (1)	5 (0.5)	5 (1.25)
[Q7]...explain/understand the column to position the cube.	5 (2)	6 (1)	6 (1)	4.5 (2.25)	5 (2)	5.5 (1)	5 (1.25)	4.5 (1.25)

* indicates statistical significance

Table 6.2: Results of the user preferences questionnaires (Median, Inter-quartile Range).

did not find any significant statistical difference after participants experienced tasks in the role of assembler.

6.4.3 Observations

During the execution of each task, we observed the participants' behavior regarding their communication style and usage of gestures. Although personal and cultural differences can influence this result, we were able to identify certain trends in each condition. Throughout all conditions, verbal communication was predominant using combined spatial and temporal references (e.g. *"left to the cube you have previously moved."*). We observed that participants developed an informal shared protocol to better understand how to complete the task. This was achieved by the instructor asking several questions to the assembler. More specifically, instructors inquired if the assembler could raise a arm and/or select a cube on a specific corner of the workspace. Henceforth, instructors would communicate the commands already in the assemblers' reference frame, which justifies the existence of significant differences in the questionnaires only for instructors.

Participants that started with *RL* condition used indicative gestures much more naturally and frequently, until experiencing the *SS* where these were ambiguous. At that point, the mentioned communication style would be established, overpowering deictics, which would be only applied as a last resort. Even so, involuntary non-verbal cues such as gaze, subtle hand, finger gestures accompanying speech, or leaning the body to a certain direction was frequently picked up by assemblers, who would try to predict the next instruction according to these visual cues. Explicit line and column indications had seldom use and had a negative impact in all of its occurrences. Indications such as *"third row, second column"* were harder to disambiguate than temporal references.

The usage of non-verbal communication varied widely according to the workspace condition. In *RL*, gestures were used to disambiguate depth, given that it was the only condition where this mapping was accurate. Also, *RL* was the only condition where we had some users use non-verbal cues as their main communication method. In *SS*, all attempts of using hand gestures resulted in errors by the assemblers. *MP* allowed users to use gestures naturally as a complement to clear verbal instructions. Finally, in *MW*, gestures were used by majority of participants, but less accurately than in *RL*, due to the fact that there was not a direct mapping between pointing and verbal directions.

6.4.4 Discussion

Results show an absence of significant differences in task performance. Yet, time data appears to reveal a tendency for *MP* to allow faster task completion. Regarding user preferences, the instructors' answers showed statistically significant differences. This happened because it was mostly the instructor who did the calculations regarding reference frames, which rendered all conditions alike to the assembler.

Although participants established the informal shared protocol to calibrate reference frames and achieved similar performance in all conditions, a *Reflected* workspace was slightly more complicated than the *Exact* representation, however not to the point of showing a statistically significant difference. We argue that the cognitive workload of being constantly converting coordinates between both frames is mentally demanding.

In complex scenarios, where it is imperative for both participants to observe the same details, the *RL* condition is unfit. This and the cognitive cost associated to the *MW* condition, leads us to suggest that, for these scenarios, having an *Exact* workspace with an *Identical* point-of-view is highly desirable. The choice between *SS* or *MP* will be dependent on whether the accuracy of the remote person's representation is more relevant than the consistency between the person and task spaces, respectively.

6.5 Conclusions

In this work we presented an evaluation of several combinations of different points-of-view, and workspace and embodiment characteristics to study remote face-to-face collaborative work on 3D shared content, with the objective of achieving a consistent and seamless reference space between participants while promoting workspace awareness. For this, we devised instructor-assembler trials, where participants jointly solved a puzzle in four conditions. Results show that participants can successfully collaborate in a shared 3D workspace face-to-face, and suggest that having an identical point-of-view is essential. Also, having an exact task space is highly desirable to avoid the cognitive cost of collaborating when remote people cooperate with different views of the shared 3D content.

As a consequence of the results' analysis, we conceptualize *Negative Space*, a telepresence approach that enables full-body face-to-face communication and creates a virtual task space between two remote spaces, where interactions with 3D objects can occur. Our proposed face-to-face approach considers that remote participants share an identical view of the same exact task space and are able to perceive one another as if they were in same physical space. We believe that the *Negative Space* can serve as a bootstrap template for future studies and developments on face-to-face collaborative work with 3D objects.

6.6 Chapter Summary

In this chapter, we investigated how people and content should be presented for discussing 3D renderings within face-to-face collaborative sessions and introduced a new design space to support such interactions. We detailed a user evaluation to compare four different conditions, in which we varied reflections of both workspace and remote people representation. Results suggest potentially more benefits to remote collaboration from workspace consistency rather than people's representation fidelity. Next, we focus on the perception of pointing gestures to distant targets in mixed reality collaborative environments.

Corresponding Publication

Part of the contents of this chapter previously appeared in the following publication:

[P6] Maurício Sousa, Daniel Mendes, Rafael Kuffner dos Anjos, Daniel Simões Lopes, and Joaquim Jorge. **Negative Space: Investigating Workspace Awareness in 3D Face-to-face Remote Collaboration.** ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI), 2019

[P4] Maurício Sousa, Daniel Mendes, Rafael Kuffner dos Anjos, Daniel Simões Lopes, and Joaquim Jorge. **Investigating Workspace Awareness in 3D Face-to-Face Remote Collaboration.** In International Conference on Graphics and Interaction (ICGI 2018), Lisbon.

7
Distorting Gestures to Improve Perception

In this chapter, we present a novel technique to improve communication via pointing gestures in mixed reality When communicating, people often use deictic references (*Deixis*) – designating the referent by pointing at it [81, 26, 150]. Pointing gestures are widely adopted nonverbal cues used to indicate distal artifacts and locations to others forgoing lengthy verbal descriptions [112]. *Deixis* is key to facilitating collaboration, since it simplifies information sharing [48, 54].

7.1 Motivation

In mixed reality collaborative environments, showing full or upper-body representations of people is known to improve awareness [19, 29], since natural body language combined with speech can be used. Current mixed reality technologies allow both local or remote users to be immersed in collaborative virtual environments, making it possible for people to see each other either through realistic virtual avatars [116] or 3D-scanned representations [111]. Deixis fosters collaboration via natural gestures that indicate virtual objects in a 3D environment, since both pointing and task objects are visible.

Collaboration improves when people are able to accurately perceive others' pointing gestures. Indeed, these have a considerable impact both on efficiency and task performance when referencing objects or locations that are in close proximity [54]. For this reason, current computer-supported cooperative work approaches resort to simple proxies of deictic pointing – telepointers [49], virtual rays and highlighted targets [163] – to reference workspace artifacts. However, these proxies afford limited control, create visual clutter, and exhibit unclear ownership. Furthermore, these methods fall short when communicating areas, paths, and directions [163].

While it is desirable to improve people's ability to execute and perceive natural gestures in collaborative virtual environments similarly to what they do in the real world, people observing pointing gestures are often unable to precisely determine where another person is pointing to [61, 15, 27, 128], causing people to engage in lengthy verbal descriptions to single out the location of interest [54]. There are other ways to resolve referent ambiguity besides natural language. However, our approach handles this in a natural and transparent manner.

Figure 7.1 demonstrates a typical pointing gesture where a person is pointing at a specific target (P_a). To this end, they typically align the tip of their index finger with the referent appearing in their field of view [15, 147, 160]. That is, the target location is intercepted by the vector from the eye to the tip of the index finger [61]. Yet, in contrast to referrer's gesture, observers use the direction of the pointer's arm and index finger to extrapolate its target [160, 61]. Thus, as exemplified in Figure 7.1,

Video Figure 3. Warping Deixis Overview.
http://web.ist.utl.pt/~antonio.sousa/videos/sousa2019-acmchi-video-figure.mp4
(File size: 8.9 MB)

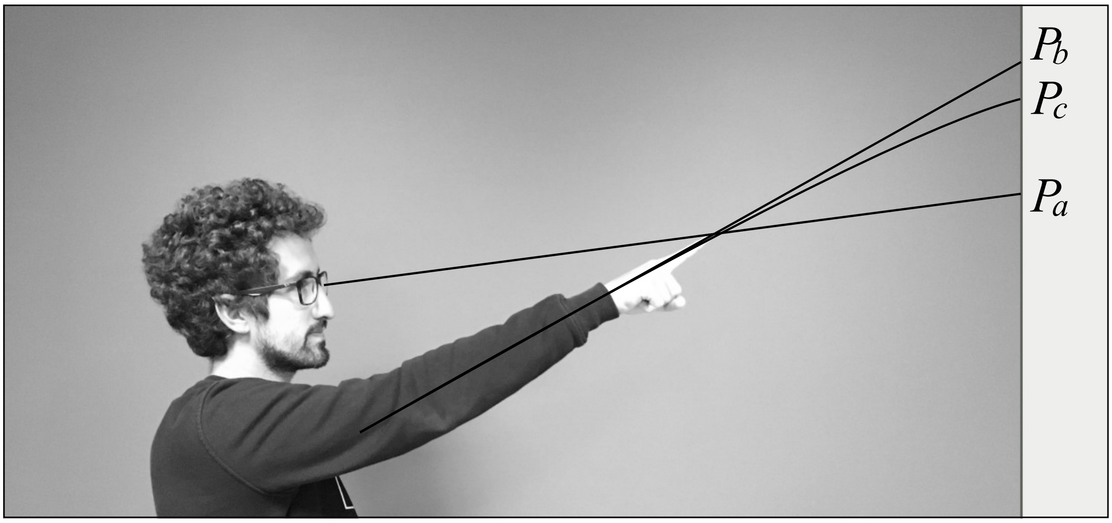


Figure 7.1: When pointing to a distal referent (Pa), people usually put their index finger between their eyes and target. Yet, observers often rely on a extrapolation of the arm-finger line to find the target of the gesture (perceiving something between Pb and Pc).

a linear extrapolation following the arm to index finger vector leads to a perceived target (P_b) that is perceived as lying above the spot designate by the person performing the gesture. Furthermore, that extrapolation is nonlinear and most observers would judge that the person in Figure 7.1A is pointing at the vicinity of P_c.

Herbort and Kunde [61] suggested that human attempts to judge the target's position using a linear extrapolation, which differs from a pure geometric linear extrapolation, can be accurately described by a Bayesian model. Therefore, observers consistently fail to interpret the target of another person's pointing gesture (by perceiving referents between P_b and P_c instead of P_a), forcing the pointer to engage in lengthy verbal descriptions to single out the location of interest.

In this work, we propose *Warping Deixis*, to improve the perception of deictic gestures in mixed reality collaborative virtual environments. Our approach manipulates the pointer's avatar to rectify the pose of the pointing arm in real-time, for the representation to match the way people perceive the deictic gesture. We do this by dynamically relocating the arm on the pointer's virtual representation to create the illusion of gesturing towards another location, thus improving the perception by an observing collaborator, as demonstrated in Figure 7.2A (before warping) and Figure 7.2B (after warping). The pointer's intended target is determined using the eye-index fin-

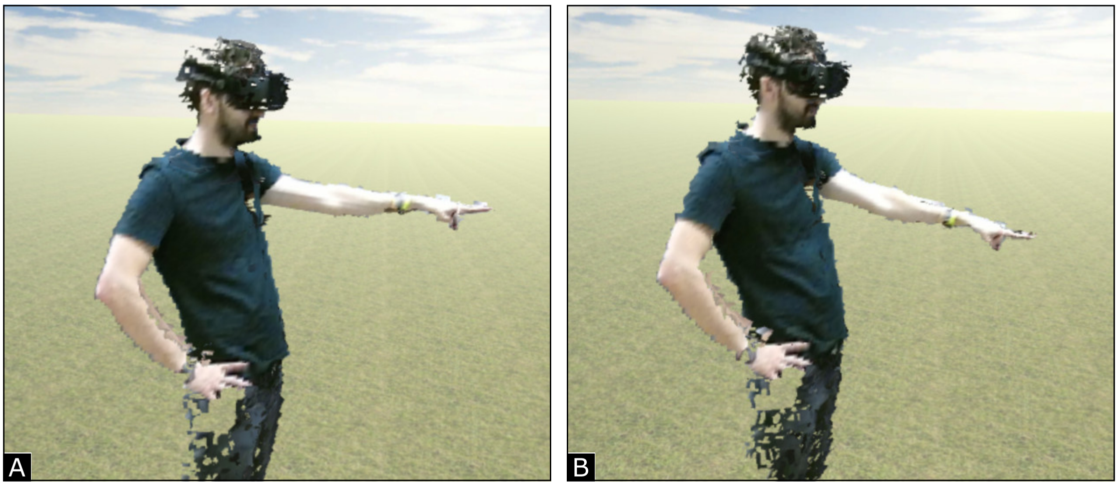


Figure 7.2: We present Warping Deixis, an approach to reducing misinterpretation of deictic gestures using body warping. Given a A) body representation of a pointing person, our approach B) changes rendering of the avatar's arm to reduce the ambiguity of the gesture.

ger vector and the pointing arm is warped around the pointer's shoulder so that the arm-index vector could match the pointer's intended target. An overview of *Warping Deixis* can be seen in Video Figure 3.

The main contributions of this research include:

1. Warping Deixis, a novel body warping technique to improve how deictic gestures are interpreted in collaborative mixed reality;
2. techniques to redirect arm poses applicable to different representations of virtual humans;
3. a user study, evaluating the impact of our approach in referent identification tasks;
4. and design considerations for future collaborative scenarios.

The user study validated the assumption that warping the pointer's arm can significantly reduce misunderstandings of the referent and that people were not aware of the avatar distortion, showing that our technique does not impair communication in shared virtual environments as compared to the non-distorted representation.

7.2 Warping Deixis

We propose Warping Deixis, an approach to reshape people's pointing poses in real-time, to improve human perception of deictic gestures in collaborative settings. We define Warping Deixis as any adjustment to the avatar of a person performing a pointing gesture in order to make distal referents both more explicit and easier to be identified. These adjustments should be plausible in order not to shift other people's attention away from the collaboration proper due to abrupt arm movements. We followed an approach analogous to the body warping technique by Azmandian et al. [12]. We also target pointing gestures towards distal referents, commonly executed with an almost fully extended arm [160], as depicted in Figure 7.1.

In this work, we focus on mixed reality environments where people collaborate with each other through virtual representations that can be manipulated whenever someone performs a pointing gesture. Therefore, our virtual representation manipulation approach incorporates two separate stages; 1) applying a Bayesian model to determine where people should be pointing when performing a gesture, 2) contributing a warping technique to suitably change virtual representations of people.

7.2.1 Bayesian-based Pointing Correction Model

As previously mentioned, when interpreting a pointing gesture, observers try to identify distal referents using a linear extrapolation of the vector that follows the pointer's arm (resulting in P_b when the pointer's intended target was P_a, in Figure 7.3A). However, experimental results from Herbort and Kunde [61], suggest that human attempts at linear extrapolation systematically deviate from a perfect geometric linear extrapolation and the observer's perceived position for the referent is usually located slightly further up, between P_b and P_c depending on the pointer's distance to the referent. Still, the observer's interpreted distal target location is disparate from the location intended by the person pointing (P_a). Accordingly, our approach follows these pointing production and gesture interpretation fundamentals to determine the optimal pointer's arm pose that would cause the deictic arm vector to appear to be pointing exactly above the intended target (P_d), as depicted in Figure 7.3. This enables the natural nonlinear human attempts of linear extrapolation to induce the

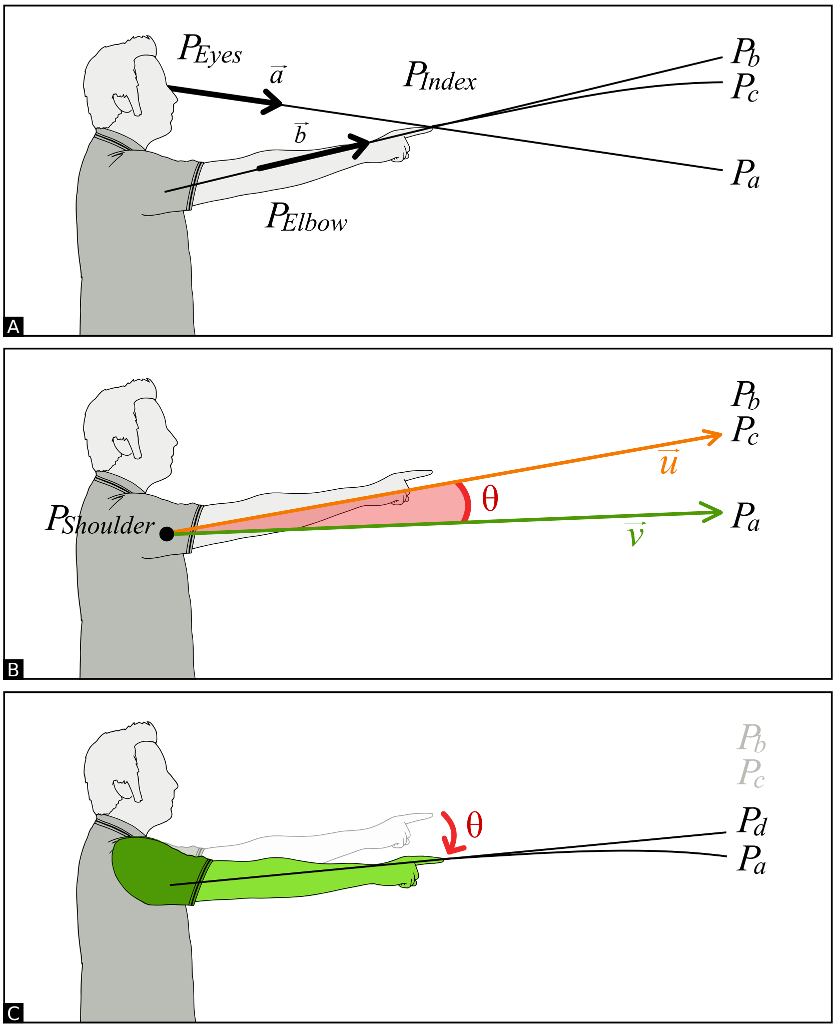


Figure 7.3: Our approach uses A) a Bayesian model to predict the referent's location interpreted by the observer, Pc, and B) calculates the necessary arm displacement to a C) distal location Pd such that extrapolations by observers would result in their correctly interpreting the pointer's intended target Pa.

observer to perceive the correct intended distal referent. Next we detail the steps necessary to calculate what the pointer's arm position that will induce the desired effect.

First, to realize the intended target ($\mathrm{P_a}$), it is necessary to calculate the pointer's deictic vector and examine its direction. So $\vec{a}$ can be located by following the vector that starts from the pointer's eyes toward the index finger, $\vec{a} = P_{Index} - P_{Eyes}$. We define the vector representing the linear extrapolation of the pointer's arm as $\vec{b} = P_{Index} - P_{Elbow}$. We established the elbow as the vector's starting point considering that pointing gestures towards distal referents are not always executed with a fully extended arm [160] and, therefore, the arm segment between shoulder and elbow are usually not considered by observers when extrapolating the arm's pointing direction. However, any transformation of the pointer's arm should use the shoulder as a rotation pivot point to exclude awkward and unnatural arm postures. Moreover, when in a pointing stance, the shoulder offers more rotation freedom in contrast to the elbow.

Our approach relies on the Bayesian extrapolation model defined by Equation 2.1 to predict the referent's position estimated by the observer ($\mathrm{P_c}$), thus $P_c = \hat{P}_b$. Given the pointer's shoulder as a rotation pivot, it is possible to determine the angular transformation needed to position the arm in the location where it should be. So, as depicted in Figure 7.3B, the rotation required is the angular distance between ($\mathrm{P_c}$) and ($\mathrm{P_a}$). Given the vectors from shoulder ($\mathrm{P_{Shoulder}}$) to the perceived target, $\vec{u} = P_c - P_{Shoulder}$, and to the pointer's intended target, $\vec{v} = P_a - P_{Shoulder}$, it is possible do determine the rotation axis $\vec{r} = \vec{u} \times \vec{v}$ and the rotation angle $\vartheta = \angle(\vec{u}, \vec{v})$.

The value of ϑ can be applied to the pointing arm of any body representation of a pointing person, since ϑ is the angular distance necessary for the arm to be pointing to a distal location ($\mathrm{P_d}$) that should result in an interpretation of ($\mathrm{P_a}$) through a non-linear human extrapolation of the pointing gesture, as demonstrated in Figure 7.3C.

7.2.2 *Warping People's Virtual Representations*

Different methods have been used to create a virtual representation of people in mixed reality environments (e.g. avatar model, point clouds, virtual hands). In all these representations, arms are usually defined by a set of 3D points representing either joints or surface points. Therefore, any warping operation will consist of transforming a set of points according to an estimated matrix. For an avatar model, skeleton transformations would bring a rigged mesh to the right location, yet, for point cloud-based representations, the transformation must be applied to each individual point comprising the full arm. Given a virtual representation $\mathcal{V}$, consisting of a set of 3D points p, warping the pointing arm to another location can be achieved considering that we have the position of the pivot point (which in our case is $P_{Shoulder}$) and that the point set representing the arm, $\mathcal{A} \in \mathcal{V}$, can be estimated. Thereby, the rotation matrix R representing the angular rotation ϑ about the axis defined by $\vec{r}$, and can be calculated by:

$$R = (\cos\vartheta)I + (\sin\vartheta)[\vec{r}]_{\times} + (1 - \cos\vartheta)(\vec{r} \otimes \vec{r}) \tag{7.1}$$

Where $[\vec{r}]_{\times}$ is the cross product matrix of $\vec{r}$, $\otimes$ is the tensor product, and I is the identity matrix. Then, our warping matrix W is:

$$W = T_{P_{Shoulder}} R T_{-P_{Shoulder}} \tag{7.2}$$

Which represents a translation of the representation $\mathcal{A}$ to the origin so it is centered around the pivot point $P_{Shoulder}$, followed by the rotation R, and translating $\mathcal{A}$ back to its original position. Finally, we apply the warping matrix to each 3D point in the virtual representation of the arm:

$$\vec{p}_{warped} = W\vec{p}\,,\ \forall\, p \in \mathcal{A} \tag{7.3}$$

Figure 7.3C describes this process visually, highlighting in green the points that were affected by the warping transformation. In the next section, we introduce the user study, describe the evaluation prototype and discuss implementation details.

7.3 Evaluation

To assess whether our approach improves the perception of pointing gestures in collaborative settings, we conducted a user study using pairs of participants. During the evaluation, participants were asked to alternate between the roles of pointer and observer. The main goal was to check how warping the pointer's arm would benefit the observer's attempts at extrapolating the target location. We also evaluated whether our warping technique was perceptible to the participants.

For this, we employed a fully-immersive virtual environment to accommodate participants in a side-by-side formation (at a distance of 2m from each other), facing the location were targets would appear. We followed the arrangement of participants and location of targets previously utilized by Herbort and Kunde [62]. However, in our evaluation, participants were immersed in a virtual environment, yet they could see each other's 3D avatars in real-time. Accordingly, we compared task performance and gathered user preferences in two conditions: (1) with Warping Deixis and (2) without Warping Deixis (baseline).

7.3.1 Procedure

Participants were asked to perform a set of three tasks for each of the two conditions. The order of conditions was counterbalanced between sessions to avoid biased results. All sessions followed the same structure: 1) an introductory briefing; 2) filling in a consent form and a profile questionnaire; 3) executing the tasks with the first condition; 4) filling a questionnaire for the first condition; 5) executing the tasks with the second condition; 6) filling the final questionnaire for the second condition. This took approximately 30 minutes in total.

We started by introducing the user study procedure to each pair of participants, followed by a description of the evaluation's main objective without revealing our body warping technique. Participants were only informed that the evaluation was a study on perception of pointing gestures. Each participant was then randomly assigned to their location, left or right in a side-by-side formation. Afterwards, participants jointly executed both sets of tasks.

Task execution for each condition was made up of two stages. In the first stage, the participant on the right initially assumed the role of observer, while the left participant was given referents to perform pointing gestures. Then, participants followed a set of three number identification tasks on a vertical pole at three different distances, similarly to Herbort and Kunde [62]. The second stage consisted of repeating the first stage, but with the roles of pointer and observer reversed. In the following, we detail the evaluation tasks.

7.3.2 Tasks

For the purpose of this research, we reduced the need for supplementary verbal or contextual information as much as possible, since our objective relates to the accuracy of the information conveyed by the pointing gesture alone. Therefore, tasks were designed to not allow participants to use verbal descriptions to convey the location of

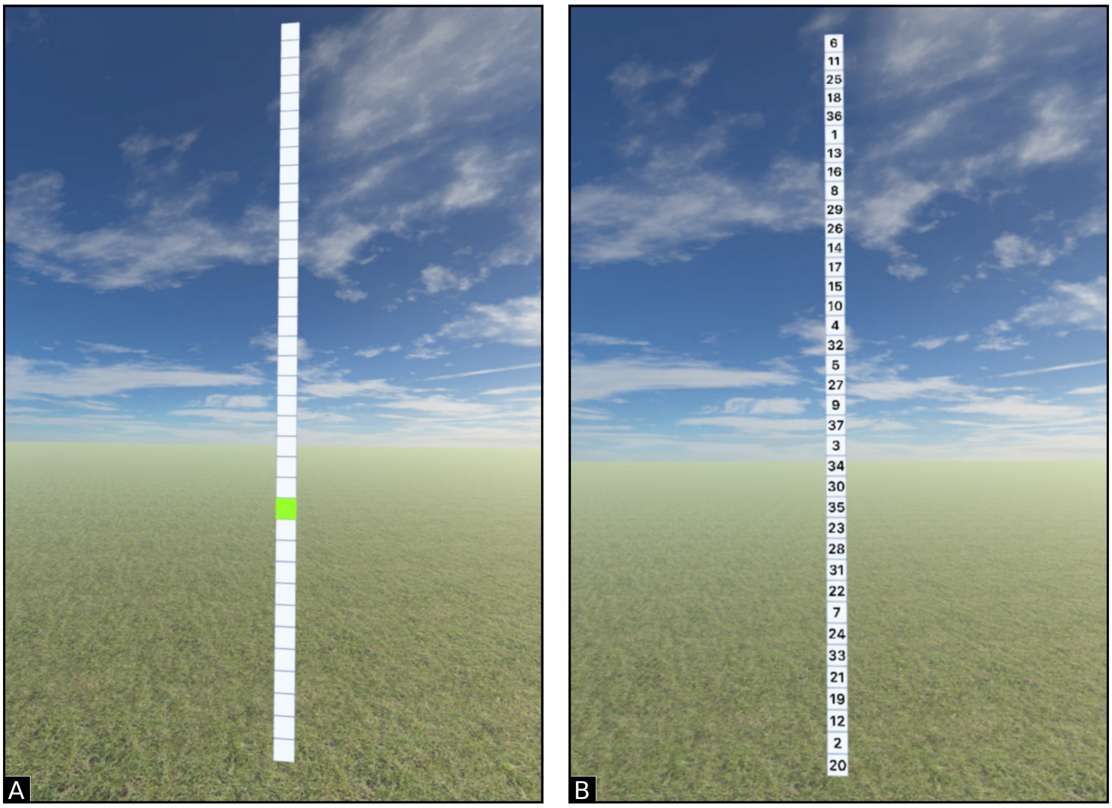


Figure 7.4: Pole task: A) from the point-of-view of the pointer, the pole featured blank squares and the target was highlighted in green; B) on the other hand, observers were unaware of the green target and the squares were numbered.

the target referents, also, the participants were encouraged to not use speech communication and just perform a pointing gesture.

We replicated the numbered pole experiment from Herbort and Kunde [62] in a virtual environment. In all tasks, different information was presented individually to the pointer and the observer, as shown in Figure 7.4. When participants assume the roles of pointer and observer, they were asked to perform three tasks using a vertical numbered pole at different distances to the pointer: one, two and three meters (Figure 7.5). While the participant in a pointing role was presented with a highlighted target and no numbers, the observer was unaware of the target's location but could see the numbers. The observer was asked to report the referent's exact location based on how they interpret other participant's pointing gesture.

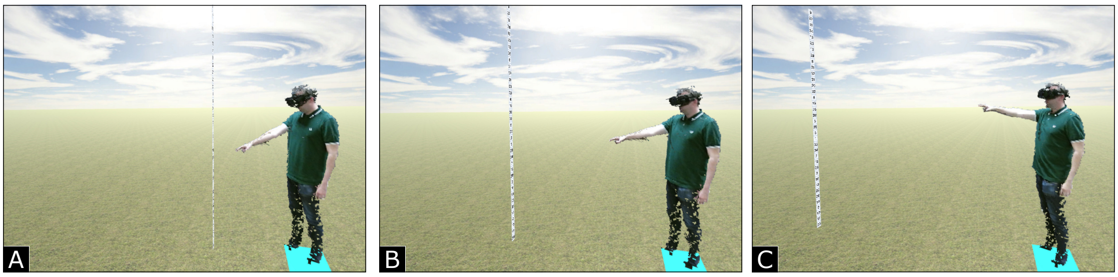


Figure 7.5: Participants were asked to perform three referent identification tasks in a vertical pole positioned at A) one meter, B) two meters and C) three meters.

The pole consisted of a vertical numbered line with 37 white squares with black borders (8cm x 8cm), starting from the floor to 296cm of height. Thus, the vertical distance between the center of adjacent squares was 8cm. We doubled the square size used by [62] to improve the readability in Virtual Reality head-mounted displays, and the pole was positioned in front of the pointer. As shown in Figure 7.4A, the pointer's view of the pole consisted of blank squares with the referent highlighted in green. Pointers were instructed to point at the green square. On the other hand, the observer's view showed numbered white squares (Figure 7.4B). The numbers on the square labels were previously assigned to each square randomly and were used by the observer to report where the pointer was pointing to. Each pole task displayed numbered squares in a different order, and the top and bottom squares were excluded as referents.

7.3.3 Setup and Prototype

We configured the evaluation environment for both participants side by side in the same room. Each setup consisted of a desktop computer connected to an Oculus Rift headset as depicted in Figure 7.6. We used a non-intrusive open source toolkit [135] for body tracking, to combine skeleton information with the 3D representations of people in the same coordinate system. Our capture setup included two Microsoft Kinect v2 sensors mounted on tripods, 2m above the floor, facing down to ensure that pointing arms were always unobstructed during capture.

We developed a prototype in Unity3D, and both setups were connected through a LAN evaluation server using TCP connections for both user's representation and the synchronization of the evaluation environment, as depicted in Figure 7.6. In the

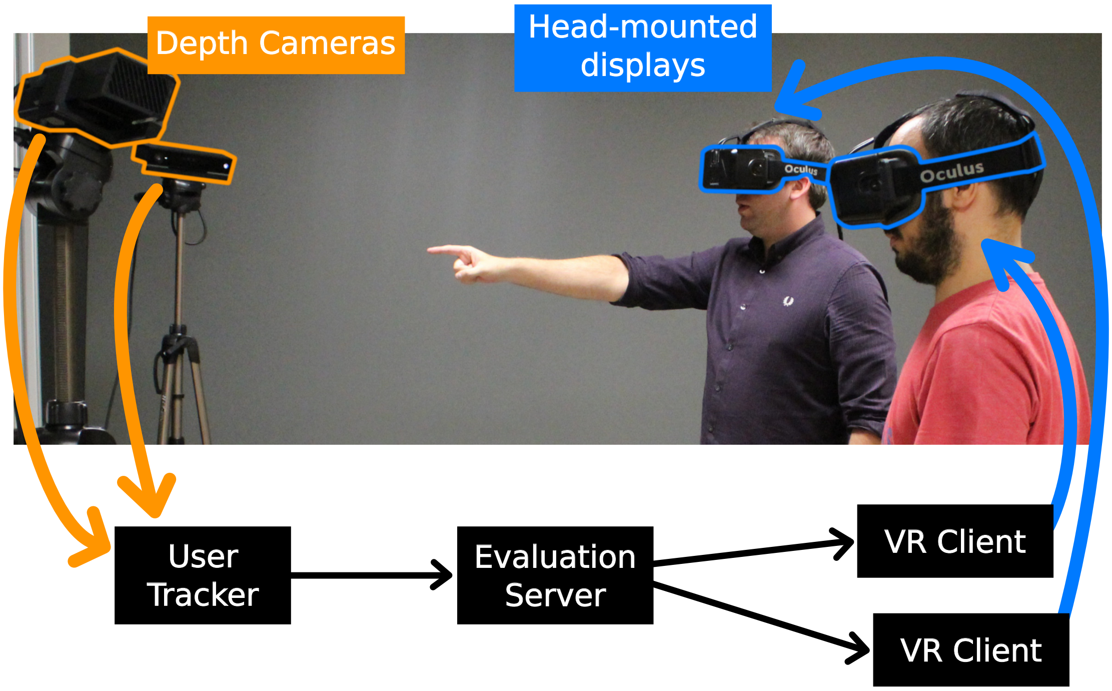


Figure 7.6: Evaluation setup with two user study participants and the prototype's architecture design.

virtual environment, participants could see their own body representations and their partners standing to their side. The virtual representations in the virtual environment matched the real world position of the participants. The virtual environment also included visual indicators of the participants assigned positions, matching physical floor mats providing passive haptic feedback. A separate controlling application was used by the evaluation moderator to advance the tasks in both environments, instantiating targets and indicators accordingly, and setting the given answers during the pole tasks.

The participants' virtual representations were drawn as a 3D polygon mesh using color and depth values obtained from the depth cameras. Body warping was implemented in a vertex shader, applying Equation 7.3 to each point belonging to the arm. To predict the observer's interpreted location of referents, we employed the average values of w for Equation 2.1 provided by Herbort and Kunde [61] for side view gesture interpretation for each referent distance.

Warping can be triggered when someone is pointing to a target location or virtual object. In this case, smooth transitions can be applied to avoid gross discontinuities

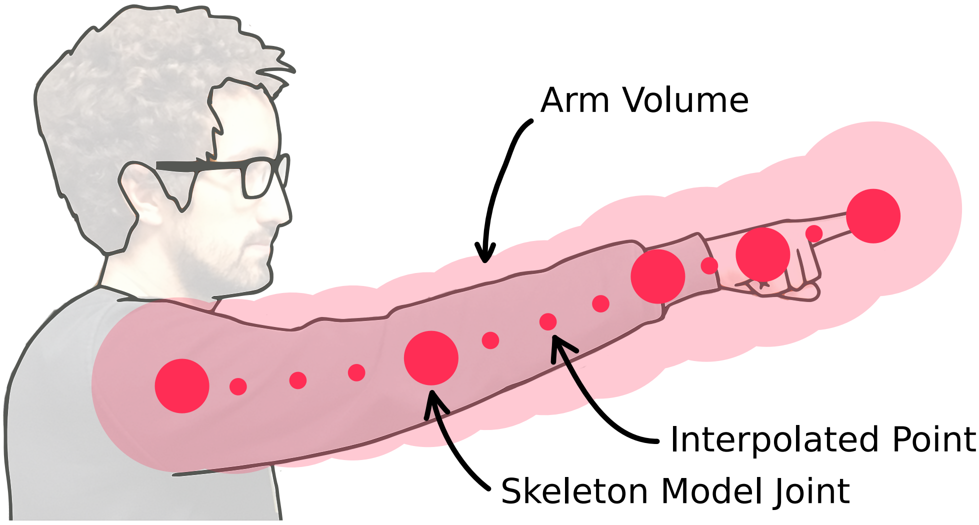


Figure 7.7: To identify the pointing arm 3D points to warp, we consider all points within a volume defined by a set of spheres centered across the arm skeleton model joints and other interpolated points between those joints.

in arm movements. For the purpose of this evaluation, and since the only target consisted of one pole, we employed a collider much larger than the pole (four meters height and a width of two meters), which triggered the warping as soon as the participant raised an arm. This triggering approach allowed for the warping to start earlier and gradually. However, this strategy would need to be refined for virtual environments with multiple targets.

In regards to warping virtual representations, whenever a participant pointed to the target area, the shader would be updated with the relevant skeleton joint positions. To determine what point-cloud elements would be warped, our approach selected the points that were contained within a bounding volume, representing the person's arm. To determine that volume, we considered all space at the distance of 15cm from the center of each Kinect skeleton model joints and interpolated bone positions calculated from increments of 5cm, as demonstrated in Figure 7.7.

7.3.4 *Apparatus*

The evaluation trials were performed in a controlled laboratory environment (Figure 7.6). All trials featured two moderators. One managed the evaluation server and guided the experiment, while another took notes and observed whether participants experienced any difficulty or discomfort. The server fired each trial and collected targets perceived by the observer (manually introduced by the first moderator). To collect targets, the second moderator used a scripted dialogue that required the observer to report and confirm the perceived target's number.

Each participant was instructed to stand on top of the floor mats positioned to match the positional indicators in the virtual environment. Participants were also instructed not to move around freely and keep to their assigned positions during each session.

7.3.5 Participants

Our subject group included 18 people (11 male, 7 female), organized in pairs. While participants' ages ranged from 18 to 44 years, most (14) were between 18 and 25 years old. All reported having previous usage experience in Virtual Environments.

7.4 Results and Discussion

During the evaluation sessions we collected *Task Performance* data through logging, and *User Preferences* from questionnaires completed after finishing each set of tasks under both conditions.

7.4.1 Task Performance

We measured participants' task performance using the distance between the task's target, as indicated by the pointer, and the perceived target reported by the observing participant. Similarly to Herbort and Kunde [62], we measured distances between the centers of task targets and perceived target squares on the virtual pole, and then converted these to meters. Figure 7.8 shows the logged mean error distances of the observers for each task under both evaluation conditions.

We performed a two-way repeated measures ANOVA to assess how the independent variables, pole distance and technique, affected the perceived distance to the target. Pole distance included three levels (1, 2 and 3 meters) and technique consisted of two levels (baseline and Warping Deixis). All effects were statistically significant at the .05 significance level. The main effect for distance yielded an F ratio of $F(2,34) = 60.325, p < .0005, \eta_p^2 = .780$. Post-hoc Paired T-Tests revealed significant differences between 1m ($M = .094, SD = .009$), and 2m ($M = .215, SD = .016, t(35) = -6.726, p < .0005, d = -1.121$), between 1 and 3m ($M = .317, SD = .023, t(35) = -8.972, p < .0005, d = -1.495$), and between 2m and 3m ($t(35) = -7.122, p < .0005, d = -1.187$). The main effect for technique yielded an F ratio of $F(1,17) = 5.753, p = .025, \eta_p^2 = .253$, indicating a significant difference between baseline ($M = .240, SD = .018$) and Warping Deixis ($M = .178, SD = .017$). The interaction effect was significant, $F(2,34) = 16.747, \eta_p^2 = .496$.

Post-hoc tests, using a Paired T-Test with Holm-Bonferroni correction (Table 7.1), revealed no statistically significant difference between our approach and the baseline condition in the first task (1m to pole), whereas for the other tasks (2m and 3m to pole), Warping Deixis (1m: $M = .107m, SD = .067$; 2m: $M = .174m, SD = .089$; 3m: $M = .252m, SD = .135$) successfully improved the observers' perception in comparison to the baseline (1m: $M = .085m, SD = .043$; 2m: $M = .255m, SD = .094$; 3m:

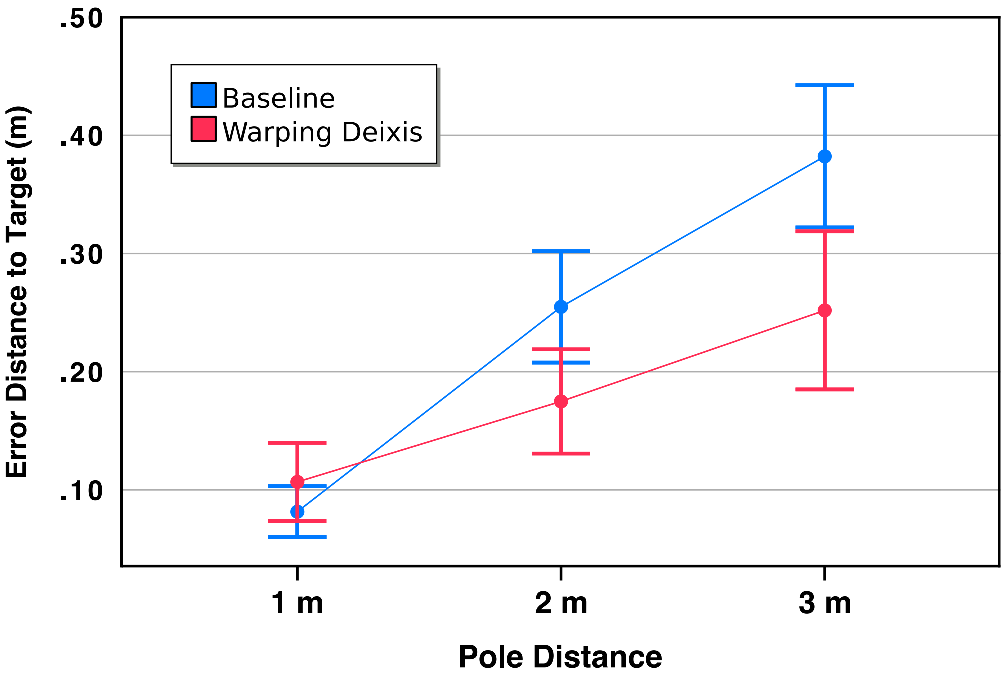


Figure 7.8: Task performance results for each condition: mean and 95% confidence interval error bars.

Comparison	*t*	*df*	*p*	*d*	*α*
BL 1m - WD 1m	-1.327	17	.202	-.312	.05
BL 2m - WD 2m	2.671	17	.016 *	.629	.017
BL 3m - WD 3m	3.386	17	.004 *	.798	.01
BL 1m - BL 2m	-9.331	17	<.0005 *	-2.199	.006
BL 1m - BL 3m	-11.212	17	<.0005 *	-2.642	.006
BL 2m - BL 3m	-8.57	17	<.0005 *	-2.019	.007
WD 1m - WD 2m	-2.66	17	.016 *	-.627	.025
WD 1m - WD 3m	-4.354	17	<.0005 *	-1.026	.008
WD 2m - WD 3m	-3.277	17	.004 *	-.772	.013

Table 7.1: Statistical tests reported at $p = .05$ significance levels (BL: baseline, WD: Warping Deixis). $*$ denotes statistical significance compared to the Holm-Bonferroni corrected α value.

$M = .382m, SD = .121$). At one meter, the pointer's index finger is so close to the referent that our approach yields no significant gain. This result agrees with the findings at one meter reported by Herbort and Kunde [62]. However, for either technique, longer distances to the pole significantly increase the error to the perceived target as shown on the last three lines of Table 7.1. For these, Warping Deixis shows significant advantage.

7.4.2 User Preferences

After completing each set of tasks, participants were asked to fill in a preferences questionnaire related to the condition they had just experienced. One of our key goals was to assess whether warping was perceived by the observers. The questionnaire included statements scored on a 6-point Likert Scale where a value of 1 meant that users did not agree at all with a statement and 6 meant that they fully agreed with it. Table 7.2 shows posed questions and corresponding results for both conditions.

For all questions, the Wilcoxon-Signed Ranks test revealed no statistically significant differences between the baseline and Warping Deixis conditions. This suggests that our approach warped the arm in a convincing manner, since participants did not seemingly distinguish any morphological changes in the body representations of their companions.

In addition, the questionnaire also featured the open question: "Did you find anything strange about your partner's body representation in the Virtual Environment?

Question	*Warping Deixis*	*Baseline*
Q1. I felt present in the Virtual Environment.	5 (1.5)	5 (1.5)
Q2. I felt that my colleague was present in the Virtual Environment.	5 (1.75)	5 (1)
Q3. I felt that I was pointing to were I wanted to point.	4.5 (1)	5 (.75)
Q4. It was easy to understand where my colleague was pointing to.	4 (0)	4 (1.75)

Table 7.2: Results for the user preference questionnaires (Median, Inter-quartile Range).

If so, please state what.". Seven participants reported the somewhat noisy representations caused by the depth sensor for both conditions. However, they did not report anything specifically relatable to the Warping Deixis condition, reinforcing that avatar distortion was not noticeable.

7.5 Discussion

From the findings, as revealed by the evaluation, we conclude that Warping Deixis demonstrates a significant improvement in the interpretation of deictic gestures to distal referents in a Virtual Reality environment. The tasks presented show that observers benefit from our warping technique when interpreting the referents located two and three meters in front of the person performing the pointing gesture. Still, our approach has some limitations.

The employed Bayesian model only considers the vertical axis to extrapolate the observers' interpretations of pointing gestures. In this research we focused on improving the accuracy of the vertical component, because misunderstandings occur consistently due to the elevation of the arm [15, 61, 160]. Furthermore, arm elevation is not only relevant to indicate referents in a vertical plane, but also is useful to refer to objects at different depths/distances. Yet, previous research suggests that human vector extrapolation is often biased toward both the vertical and horizontal axis [23]. Further research is necessary to assess the benefits of using a Bayesian correction approach to horizontally distributed referents.

In our evaluation prototype, we employed a virtual representation of the participants based on point cloud data converted to a textured mesh, using data from commodity depth cameras. Our approach showed some noisy contours, especially in parts of the participants' body that were not facing the depth cameras. Some participants reported this, although none suggested that the issue affected the experience. In future research, more accurate representations of people should be used to assess body warping techniques. One might argue that camera noise had the positive effect of masking distortions induced by warping limbs during deictic gestures. A more accurate representation might require more work to make geometric distortions imperceptible.

Finally, our approach provides the means to reduce the ambiguity of deictic gestures but does not allow for precise identification of referents. Indeed, if the evaluation participants were able to use verbal communication to resolve target misunderstandings, tasks would require considerably longer periods of time to be accomplished and the identification of referents would be more exact. Yet, pointing gestures also function as a complement to speech, when verbal communication combined with deictic references is difficult [61]. Furthermore, pointing gestures to ambiguous referents require longer verbal descriptions than unambiguous ones [14], allowing people in collaborative environments to become more focused on domain tasks and less involved in the tasks of maintaining the collaboration [54].

7.6 Conclusions

In this chapter we introduced Warping Deixis, a body distortion approach to improve the perception of pointing gestures in virtual collaborative environments. The effectiveness of the technique, is backed by an experimental evaluation as compared to a baseline condition. To this end, we compared our warping method with not applying it at all in a series of tasks to identify referents on a numbered pole. We devised a virtual environment where two participants alternately assumed roles of pointer and observer. Results suggest that Warping Deixis is successful at reducing the ambiguity of pointing gestures. Furthermore, people failed to notice the effects of our body warping approach when interpreting pointing gestures and arm motions.

Our results, beyond suggesting that Warping Deixis can improve collaboration in mixed reality scenarios, also indicate that retargeting pointing gestures could benefit future human-technology approaches. Environments that exploit avatar-like or real-time 3D reconstructions of people are not the only systems that would benefit from retargeting the direction of pointing gestures, since any setting relying on the interpretation of *deixis* to interact with humans, currently suffers from the misunderstandings and miscommunication previously described. Thus, improvements in retargeting pointing gestures should enhance the effectiveness of virtual humanoid companions and non-player characters (NPCs) [108] in virtual environments, as well as, physical robot instructors and guiding helpers [38, 25, 93, 139, 131]. Despite that, the results disclosed in this work suggest that manipulating virtual representations of people is an effective technique to enhance the understanding of intentional communication in collaborative environments.

7.7 Chapter Summary

In this chapter, we introduced *Warping Deixis*, a novel approach to improving the perception of pointing gestures and facilitate communication in collaborative mixed reality environments. By warping the virtual representation of the pointing individual, we can match the pointing expression to the observer's perception. We evaluated our approach in a co-located side by side virtual reality scenario. Results suggest that our approach is effective in improving the interpretation of pointing gestures in shared virtual environments. In the following chapter, we take a step forward and address the reshaping of gestures aimed at enhancing workspace awareness in face-to-face remote collaborative settings.

Corresponding Publication

Part of the contents of this chapter previously appeared in the following publication:

[P5] Maurício Sousa, Rafael Kuffner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. **WARPING DEIXIS: Distorting Gestures to Enhance Collaboration.** In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 12 pages. DOI: https://doi.org/10.1145/3290605.3300838

8

Reshaping Gestures for Seamless Face-to-face Remote Collaboration

In this chapter, we depart from distant pointing and focus on proximal interactions in shared 3D workspaces. Therefore, we present a novel technique to improve face-to-face remote collaboration in mixed reality shared 3D workspaces using perception manipulation. We developed the present research on top of the knowledge that originated from the work introduced in the previous chapter. The ensuing investigation constitutes the final user evaluation that closes the research that encompasses this dissertation and tries to give a definitive answer to our research statement.

8.1 Motivation

For a telepresence meeting to be closer to a co-located experience, the *person space*, should rely on real size portrayal of remote people to allow understanding of nonverbal visual cues, such as facial expressions, gaze, body posture, gestures to indicate objects referred to in speech (*deictic gestures* [104]), or how people utilize the space and position themselves when communicating (*proxemics* [56]). Buxton also identified *task space* as the "space where the work appears" that can be either private or shared between participants, and *reference space* as the "space within which the remote party can use body language to reference the work."

Despite *person space* being initially considered separate from the *task space*, Ishii et al. [68] suggested that both concepts should be integrated when considering face-to-face meetings using a transparent display metaphor to maximize perception of other people's nonverbal cues. Indeed, with transparent displays, two videoconferencing participants can simultaneously see the other and digital content rendered between them, which can be jointly manipulated, minimizing attention shifts between the remote person and the shared workspace. Yet, people in a face-to-face conversation have no common orientation of right or left, reducing awareness and raising ambiguity. This lack of awareness negatively affects the quality of the cooperation [166] because it constrains the ability for people to use descriptions of relative positions. Ishii et al. [68] addresses this issue in Clearboard by mirror-reversing the remote person's video stream, producing gaze and pointing awareness since 2D graphics and text can thus be corrected to the participant's point-of-view.

This approach has been the subject of research for 2D content collaborative manipulation [164, 91, 166]. However, collaborating face-to-face with 3D digital content gives rise to multiple problems that impair the understanding of nonverbal communication and the overall awareness of the shared workspace. Contrary points-of-view can result in different perceptions or even serious communication missteps. Participants do not share the same *forward-backwards* orientation, and occlusions can affect the understanding of where or what the remote person is pointing at. Therefore, despite maintaining the sense of "being there" [64] enabled by the ability to communicate verbally, making eye contact, and observing gestures and facial expressions [69], remote collaborators in a face-to-face formation are unable to achieve a common understanding when interacting in a 3D workspace, as depicted in Figure 8.1A.

In this work, we propose *Altered Presence*, a novel approach to improving face-to-face collaboration in shared 3D workspaces by manipulating person space, task space and reference space *in subtle manners*. Our approach ensures that opposing participants share the same perspective of the workspace and distorts gestures performed by

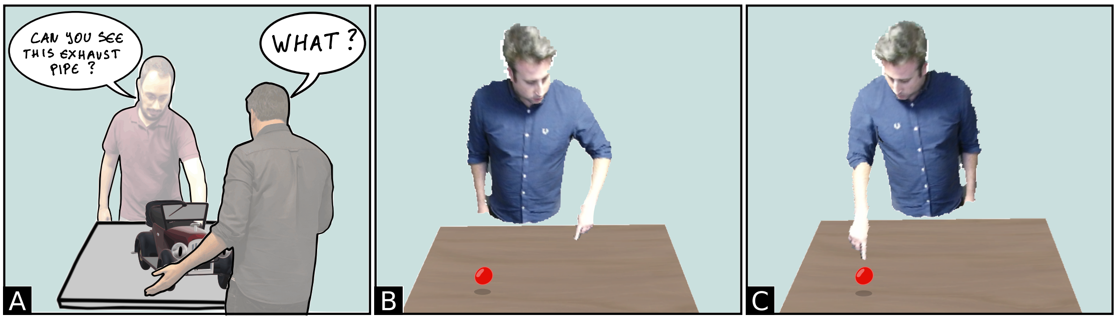


Figure 8.1: Despite leveraging nonverbal communication in virtual meetings, people communicating in a face-to-face with 3D content A) do not share the same perspective of the workspace leading to occlusions of information and misunderstandings. *Altered Presence* enables collaborators to share the same perspective while manipulating the virtual representation of remote people to improve the awareness of what is happening in the workspace. When interacting in a shared workspace B) our approach reshapes the gestures of remote people C) so that a common understanding is shared by both participants.

a remote person to present a corrected virtual representation of those gestures to the local participant using body warping.Figure 8.1B demonstrates a person performing a pointing gesture to a feature on a shared 3D workspace. An observing collaborator is able to see what is happening with the 3D artifacts in the workspace, but also perceive the gestures in a corrected reference space creating an illusion that the gesture was performed as intended because it matches the local participant's reference space, as shown in Figure 8.1C, thus improving the collaboration since both participants are always aware of the state of the workspace while being able to communicate non verbally. We developed a mixed reality telepresence environment implementing our approach using virtual representations of people rendered at life-size scale around a shared above-a-tabletop workspace. A user study validated the assumption that sharing the same perspective and reshaping remote peoples' gestures while in a face-to-face formation improves overall workspace awareness and facilitates collaboration as compared to having people collaborate in a side-to-side formation.

We contribute:

1. *Altered Presence* as a interactive space to integrate person-, task-, and reference spaces for face-to-face remote collaboration in mixed reality;

2. implementation details on warping techniques to reshape the gestures on virtual representations of people;

3. a user study, evaluating the impact of our approach on workspace awareness.

In what follows, we first present the design of our approach, and the implementation details, report on the user study and provide a discussion of the findings. Finally, we discuss design considerations for future face-to-face collaborative scenarios and directions for future work.

8.2 Altered Presence for Remote Collaboration

We introduce *Altered Presence* for face-to-face collaboration in shared object-centered 3D workspaces. Through *Altered Presence* remote people can meet and engage in computer-mediated collaborative design and review tasks in mixed reality environments. Our approach follows the Gutwin and Greenberg [54] assumption that communicating positions on the workspace, reading people's gestures, and switching attention from task space to person space and vice versa, require additional efforts to maintain collaboration. Accordingly, the main objective of our approach is to minimize participants' engagement on peripheral tasks to maintain collaboration and refocus their attention to the domain tasks.

Our vision for *Altered Presence* is to create a design space that improves collaboration by allowing remote people always to share the same perspective of the workspace. And, at the same time, be aware of the nonverbal intentional communication cues naturally provided by the full-body representation the remote person in front of them. Therefore, our approach allows people to have the same understanding of the workspace in face-to-face formations and presents manipulated versions of remote participants so that their gestures can make sense to a local observer. In this work, we focus on two people. While some aspects of our work may extend to multiple people, person-person communication is a typical critical case worth considering separately.

Next, we present a usage scenario exemplifying an instance of a collaborative interaction enabled by our approach. We further detail our improved integration of person space, task space, and reference space. And present our approach's method for manipulating virtual representations of people.

8.2.1 Remote Collaboration for Design Review of 3D models

Jessi and Chip, two automotive designers overhauling a classic car, are engaged in a remote meeting to discuss the new exhaust system. Chip is an industrial designer and requires Jessi's technical feedback since she is a mechanical engineer. So, they both agree to meet in their mixed reality workbenches powered by the *Altered Presence* approach. They meet facing each other across a virtual environment that includes life size depictions of their avatars in front of a workbench. Chip loads the model of the project he has been working. The 3D car model appears in the shared workspace between them above the workbench. Halfway though presenting his design proposal, Chip points to direct Jessi's attention to the rear of the 3D model. Jessi notices that the exhaust pipe location conflicts with the chosen rear bumper. Despite being face-to-face, both share the same perspective of the workspace, and so Jessi, using pointing gestures, indicates the optimal position for the newly redesigned exhaust pipe. Chip agrees to Jessi's assessment and places a virtual annotation marker where the exhaust pipe should be. In the end, they schedule a new virtual meeting a week forward to discuss the revised model.

8.2.2 Reshaping the representation of remote people

For two opposing collaborators facing each other to have the same perspective of the shared 3D workspace, their individual workspaces orientations are directed to them. Imposing the same perspective, at this moment, creates individual workspaces that are inverted and participants would have contradicting interpretations of left and right direction, and depth. To correct for this mismatched reference space, we propose altering the representation of the remote person. Therefore, *Altered Presence* employs two body warping techniques: mirroring the virtual representation of the remote person first to correct the left-right axis and readjust the arms interacting with the workspace to rectify the depth of the interaction.

Mirroring

Similarly to Clearboard [68] and 3D-Board [166], our approach introduces a horizontal mirror-reversal of the remote person's embodiment, using the "over a table" metaphor. With this, a local observer can perceive correctly remote interactions in the workspace's left-right axis. Furthermore, body language and gaze direction will also match horizontally with the local reference space. This method is only sufficient for vertical 2D workspaces since, in our scenario, there is also the need to correct for depth.

Move arm into position

To account for depth, we propose to artificially relocate the remote person's hand interacting with workspace, into the correct location along the forward-backward axis. Considering the hand as an end effector of the whole arm, moving the remote person's hand to the matching place in the local reference space dictates that the arm's upper arm and forearm must also follow the movement accordingly, as shown in Figure 8.1C.

8.2.3 Integrate Person-, Task-, and Reference Space

Following the style of face-to-face interactions presented in Clearboard [68], *Altered Presence* assumes an "above-the-table" metaphor with real-size virtual representations of remote people, allowing participants to see each other and share the same reachable space so they can easily discuss and interact with virtual artifacts. Also, interacting at a real-scale in a face-to-face formation promotes co-presence and enables people to be always aware of the other's actions. However, a naive approach to integrate the personal, task, and reference spaces to support the previously described scenario, has problems with occlusions that hinder workspace awareness.

Our approach ensures that the opposing participant has the same perspective of the workspace prevent people from having different views and understandings of the workspace, as depicted in Figure 8.2. Sharing an identical viewpoint ensures that participants can perceive the same regions of interest at the same time without occlusions. Also, for this to remain true, any manipulations to the workspace or to individual artifacts have to be similarly perceived on both sides. However, sharing the same perspective of the workspace breaks the reference space, since deictic gestures and gaze performed by the remote person would not match the correct intended location when perceived by the observing collaborator. *Altered Presence* distorts the remote partic-

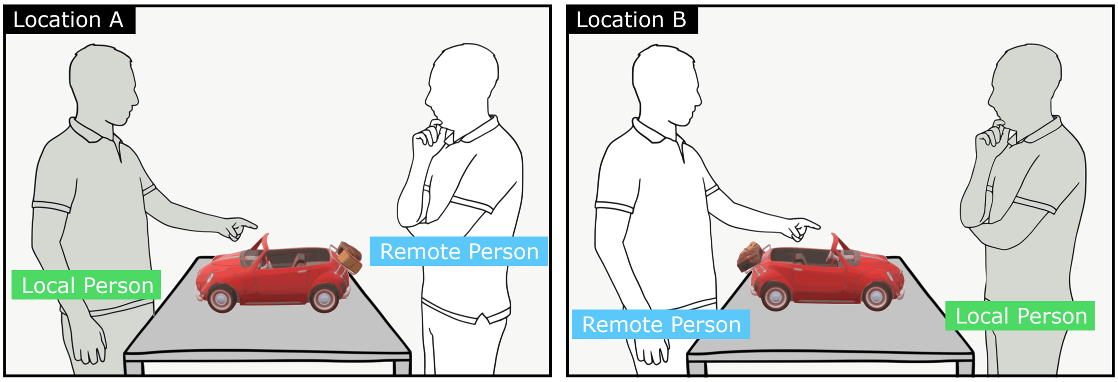


Figure 8.2: *Altered Presence* merges two remote locations into one single shared workspace, where collaborators can meet and engage in computer-mediated collaborative 3D design and review tasks

ipant's virtual representations to reshape their gestures to match the reference space of the local person. This is especially useful for tasks involving participants communicating specific features on shared digital models using proximal pointing, gesticulating relationships between workspace artifacts, or to ensure that everyone is aware of changes to virtual objects resulting from direct manipulation.

Therefore, we consider our approach to be a step forward in achieving workspace awareness in object-centered 3D collaboration because *Altered Presence* enables collaborators to seamlessly express and observe consequential communication, feedthrough, and intentional communication.

Consequential Communication.

In a full-scale face-to-face formation, the collaboration actors are always visible, reducing the necessity to shift focus between task space and person space. It requires fewer eye-movements and fewer body rotations away from the workspace. Indeed, people in front of each other can see the other person's body language and understand their actions.

Intentional Communication.

Altered Presence assures that gestures in the remote workspace are correctly converted to the local reference space. And since both participants have the same understanding of the workspace artifacts' whereabouts, deictic gestures and demonstrations per-

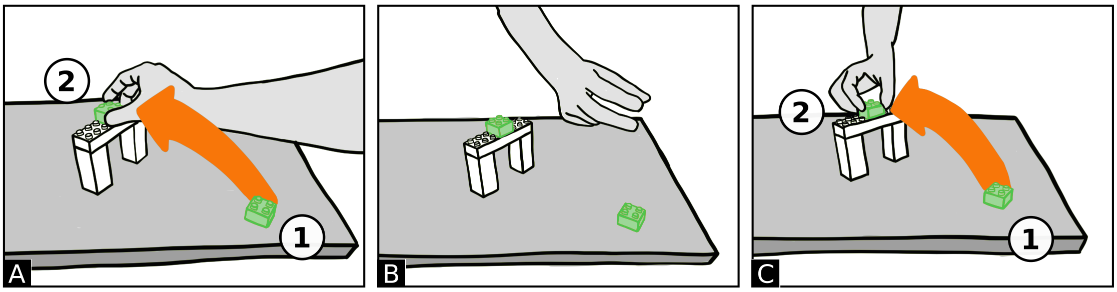


Figure 8.3: Object Manipulation: A) In a remote workspace a collaborator moves an object from location ① to ②. B) Sharing the same perspective of the workspace without any manipulations to people's actions do not match the local reference space. C) In *Altered Presence*, a local observer observes an identical action but with a mirrored path.

formed remotely remain meaningful when converted, allowing collaborators to communicate using natural gestures freely.

Feedthrough

Sharing the same perspective of the workspace allows collaborators to perceive the same artifacts' position and orientation. With AP, manipulations of artifacts remain identical in both local and remote workspaces. Figure 8.3A depicts a remote collaborator moving a virtual object from one location to another within the workspace. Reshaping the remote collaborator's gesture is fundamental to creating the illusion that the action matches the object's change in position. Thus, a local observer perceives an equivalent portrayal of that action. The remote collaborator's virtual representation acts accordingly to the local reference space (Figure 8.3B), thus communicating feedthrough.

8.3 Implementation

To study and evaluate our approach, we implemented a real-time end-to-end prototype. Our prototype platform merges the spaces of two physical remote workbenches into a single virtual workspace where local people can interact with a remote collaborator rendered in 3D in front of them, as depicted in Figure 8.2. Being a telepresence technique, our prototype requires implementing capturing, transmitting, and displaying remote people as if they were interacting in the same space (Figure 8.4).

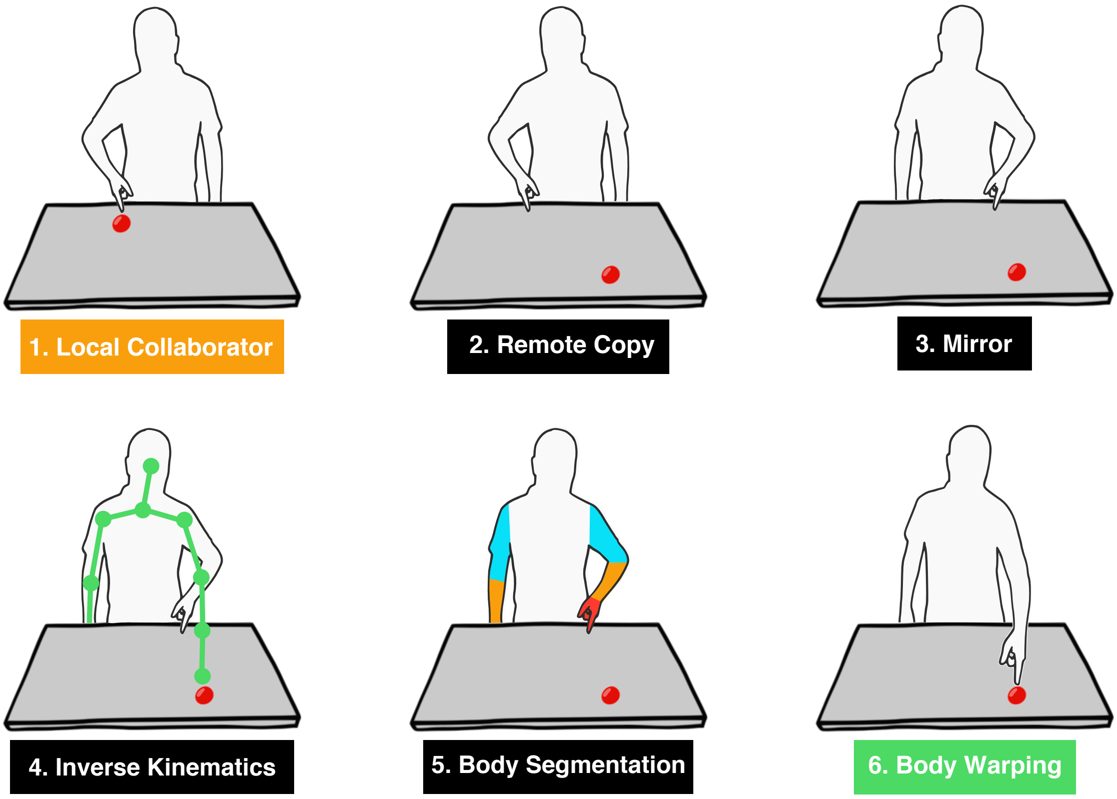


Figure 8.4: Overview of the *Altered Presence* implementation pipeline.

8.3.1 Capture

Similarly to other state-of-the-art approaches [17, 11, 95, 111, 113, 166], our implementation relies on commodity depth cameras to capture people's virtual embodiments. We used two instances of the Creepy Tracker Toolkit to capture in real-time a full color and depth point cloud data models of local and remote people. The data models are then transmitted through a local area network and reconstructed into a 3D textured mesh (similar to Maimone and Fuchs [95]) creating the appearance of captured remote person in the local collaborator's end. Following along the point cloud data, the skeleton model of the captured person is also transmitted. To reduce any noisy skeleton model data, we employed the 1€ filter [31].

8.3.2 *Body Manipulation*

Initially, we mirror the point cloud and the skeleton model in a simple horizontal reflection around the center of the workspace. If the remote speaker has his hands in the workspace/above the table, further body manipulation is used to reshape his gestures. We apply linear interpolations when the participants arms enter or leave the workspace to minimize abrupt movements that could break the illusion of people naturally performing that gesture. Body manipulation is performed in three steps.

First, from the set of body joints provided my the skeleton model, we derive a full list of segments, as depicted in Figure 8.5. For each point of the point-cloud, we observe its perpendicular to each body segment, associating it to the closest one. Second, we apply the inverse Kinematics correction. A double reflection of the participant's hand tip position inside the workspace gives us the position if the desired end effector. Having the endpoint for the hand tip effector, we use an inverse kinematic approach was based on Badler and Tolani [149] model for real-time inverse kinematics of the human arm, obtaining the end position for all the other arm joints. So, with the inverse kinematic-modified joints in place, a continuous linear interpolation is used to move the original skeleton model's arm joints towards the inverse kinemat-

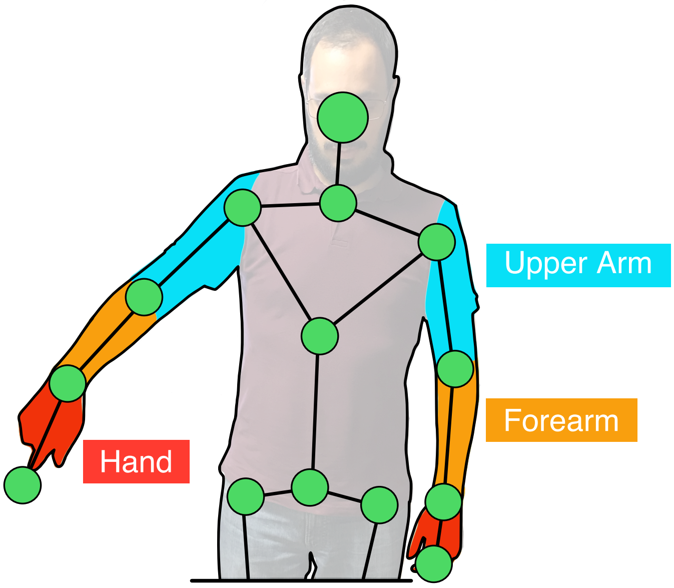


Figure 8.5: Body Segmentation and skeleton model upper arm, forearm and hand

ics calculated ones. This is done to guarantee that there is no sudden or abrupt arm movements, ensuring a smooth transition.

Finally, similarly to *Warping Deixis*, we use the obtained segmentation of the point cloud and its relation to the skeleton data, and transform their positions according to the applied transform to the arm. Differently than their approach, instead of a global transformation to the shoulder, we calculate 3 transformation matrices for each arm: the upper arm matrix, the forearm matrix, and the hand matrix, which come from the inverse kinematics calculation. The final result is a re-oriented 3D representation of the remote person, with the hand tip located at the desired spot, and the arm plausibly warped to that position.

8.4 User Evaluation

We conducted an user study using pairs of participants to assess whether *Altered Presence* enhances face-to-face collaboration by improving awareness in shared workspaces for 3D object-centered remote collaboration. The main goal of our study was to check whether people can maintain a shared understanding of the workspace while interacting in a face-to-face formation and sharing the same perspective of the workspace. Namely, we wanted to understand if our approach was successful to the point that when someone points at a specific point-of-interest, the observing collaborator can identify it in the workspace. Furthermore, we also evaluated co-presence participants to be aware of the workspace *and* their counterparts' personal space.

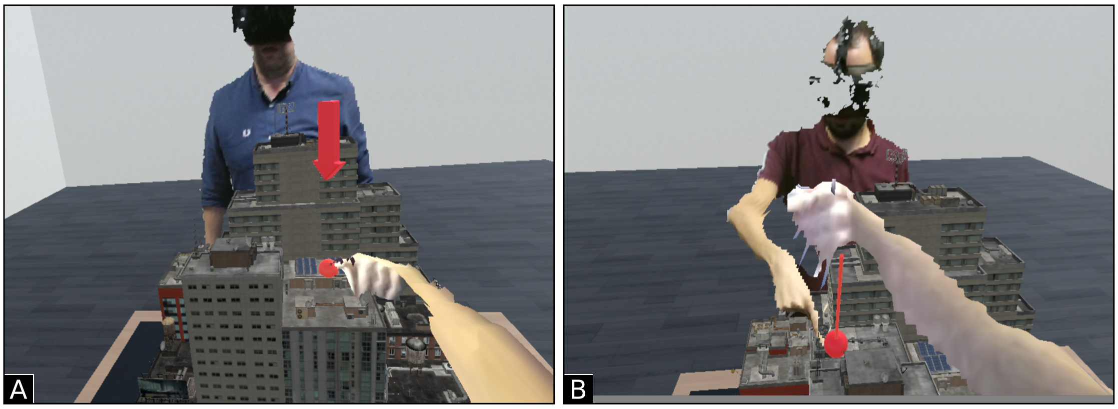


Figure 8.6: A) Instructor view and B) Operator view

8.4.1 Design

Our hypothesis throughout the experiment was that imperceptible distortions can be beneficial to remote virtual collaboration without negatively impacting task performance. Participants were asked to complete a pointing task, where one participant would point at a target and the other had to interpret the pointing gesture and identify the correct target position.The focus of the task addressed a challenge that arises frequently when collaborating over a 3D object, which is to understand the remote collaborator's perspective of the object [40]. One participant was designated *instructor* and the other had the role of *operator*. The instructor was shown a set of targets to point at (Figure 8.6A), one at a time, and the operator had to place a marker where he interpreted the instructor to be pointing at (Figure 8.6B). The targets were not visible to the operator, and the operator's marker was not visible to the instructor. The task consisted of eight repetitions, using different target positions and covering both sides of the workspace. We had six distinct sets of different targets, which were alternately assigned. To prevent occlusions by the own representation, and to make it less obvious for the instructor to know where the operator was placing the marker, there was an offset between the operator's hand and the marker. For each condition, participants completed the task twice, each time assuming a different role, and with a different set of targets. Participants were not allowed to use any verbal cues to aid target interpretation and marker placement. The evaluation followed a within-participants design, with three conditions:

- *Face-to-face with Altered Presence* (*AP*) – Participants were in a face-to-face formation in opposing sides of the workbench. The individual representations of the workspace are facing the participants for them to share the same perspective (Figure 8.7A).

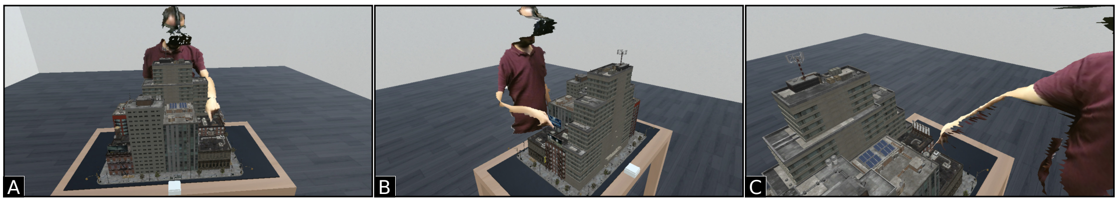


Figure 8.7: Evaluation conditions: A) *Face-to-face with Altered Presence (AP)*, B) *Veridical face-to-face (F2F)*, and C) *Side-to-side (S2S)*.

- *Veridical face-to-face* (*F2F*) – Participants were in a face-to-face formation in opposing sides of the workbench. The perspectives of the workspace followed a physical metaphor, meaning that participants had opposing points-of-view of the workspace (Figure 8.7B).

- *Side-to-side* (S2S) – Participants were in a side-to-side formation facing the workbench, therefore sharing the same perspective of the workspace (Figure 8.7C).

We decided to use baselines that mimic physical world interactions, since the performance they provide is typically the objective when developing remote collaboration approaches [113, 17]. The use of two baselines is related to the fact that one enables participants to engage in face-to-face communication (F2F), while the other allows them to observe the same side of the workspace (S2S), and no other approach exists that combines these two features. Participants were not made aware of any body warping manipulations.

8.4.2 Procedure

All evaluation sessions followed the same structure and lasted for about 45 minutes. Participants started by fulfilling a profile questionnaire and a consent form. They were then introduced to the evaluation, where we explained conditions, tasks, and roles. After, participants experimented all conditions. Conditions' order followed a Latin Square design, and participants were assigned instructor and operator roles randomly. For each condition, upon completing the task, participants switched roles and executed the task again with a different set of targets. After finishing the second task, they were asked to answer a questionnaire regarding that condition.

8.4.3 Measures

The user study included both objective and subjective measurements. Considering objective measures, we logged time participants took to complete the tasks, as well as distance errors between targets and markers placed by the operators. We also logged the distance between the participants' virtual representations and the angles between the forward vector of the operators' head to the workspace forward vector (hereinafter referred to as Look Angle). This is an indirect measure on how much of the participant's focus lies on the 3D model being discussed. As for subjective measures, we measured the social presence, perceived message understanding, and whether or not participants noticed anything unusual about the virtual representations, through questionnaires.

8.4.4 Setup and Apparatus

We built a prototype to evaluate the *Altered Presence* concept through a user study. This prototype was developed using Unity3D and was comprised of two different applications, one for each user. Each user was located in the same room, but in different spots separated by a curtain. These two applications were implemented in a way that allow the connection between those two physical spaces. To visualize the shared virtual environment each of them wore an Oculus Rift HMD, as shown in Figure 8.8. We chose this device over augmented reality glasses such as the Magic Leap or Hololens due to his increased field of view (FOV), which enabled them to view both the virtual content and the remote person representation simultaneously.

Either subject was able to see the representation of their remote partner, which were captured using MS/Kinect[1] devices and then mapped into a mesh representation. This representation was placed face-to-face to the local user or side-by-side, depending on the condition. Each of the physical spaces, comprising of a black table in a height that was comfortable for users to visualize the virtual content and the remote person

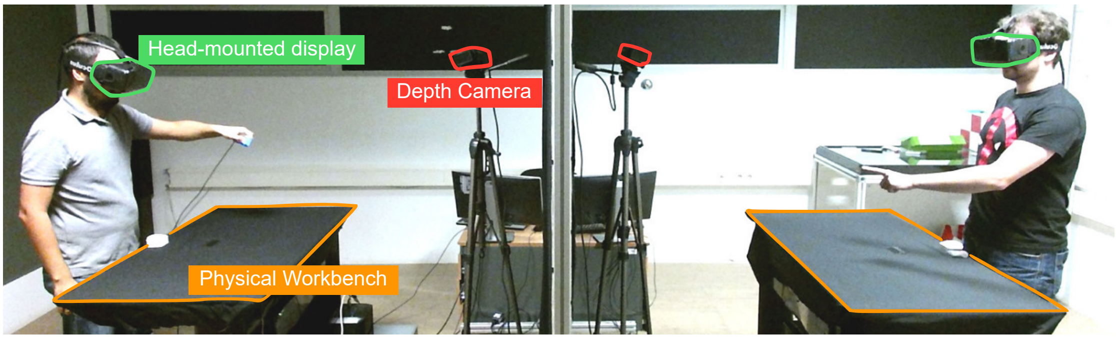


Figure 8.8: Evaluation setup.

[1] https://developer.microsoft.com/en-us/windows/kinect

accordingly. Both sides featured a physical pressing switch near the participants to start and advance the evaluation tasks. The virtual content was then placed on top of the workbench, between the local and remote user.

8.4.5 Participants

Our subject group included six pairs of participants (12 total, 6 female). Participants' ages ranged from 23 to 34 years ($M = 28.25$, $SD = 3.76$). All reported having previous experience both in virtual and augmented reality. Six participants reported rarely using virtual and augmented reality, three participants declared using these at least one a month, another three disclosed using such technologies at least once a week. Three participants reported seldom using video-conferencing solutions, four indicated using those at least once a month, and the remainder reported daily use.

8.5 Results

Throughout the evaluation trials we collected *Task Performance* and *Spatial Relationships* data using logs, and gathered *User Preferences* information from questionnaires filled up after the execution of the tasks correspondent to each condition. To analyse the gathered data, we first used Shapiro-Wilk test to check for data normality. To find significant differences in normal distributed data we ran a repeated measures ANOVA, checking for sphericity with the Mauchly's test and applying the Greenhouse-Geisser correction when it could not be assumed. For those data that did not follow a normal distribution we ran a Friedman non-parametric test. We applied the Bonferroni correction to Post-hoc tests. In what follows we present the results obtained from both task performance (time and errors) and user preference metrics.

8.5.1 Task Performance

We measured task performance using the error distance between the instructor's target and the last cursor position picked by the operator. Likewise, we measured the

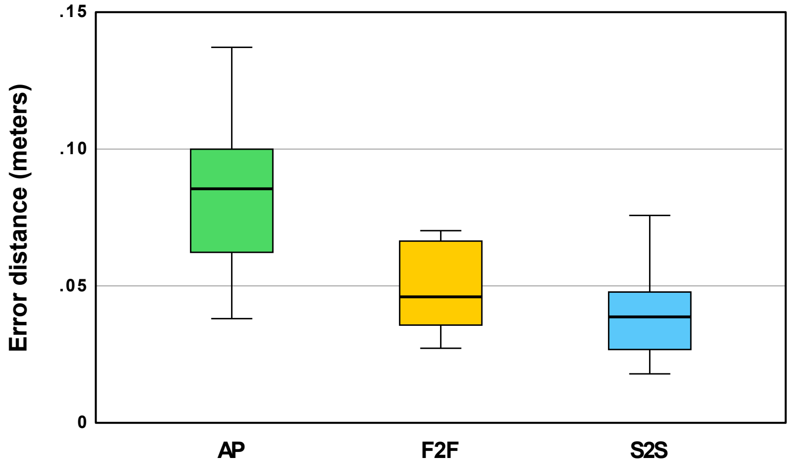


Figure 8.9: Tasks' error distance for each condition.

time spent from the start of each task to the timestamp of the operator's last cursor manipulation.

Figure 8.9 shows task error distances for each condition. The mean error distance differed significantly between conditions $F(1.090, 8.718) = 14.129, p = .004, \eta_p^2 = .645$. Post-hoc tests revealed no statistical difference between *AP* $(M = .079, SD = .008)$ and *F2F* $(M = .048, SD = .005)$. However, the statistical analysis revealed statistical differences between conditions *S2S* $(M = .034, SD = .004)$ and *AP* $(p = .003)$, and between *S2S* and *F2F* $(p = .014)$. Regarding task times (Figure 8.10), no statisti-

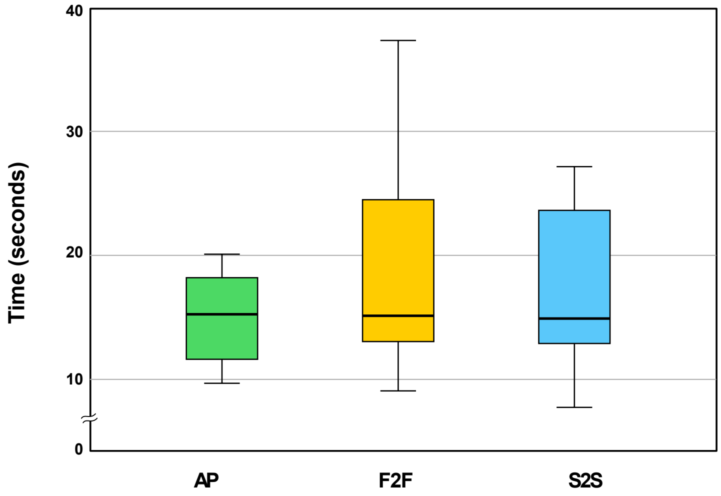


Figure 8.10: Task time for each condition.

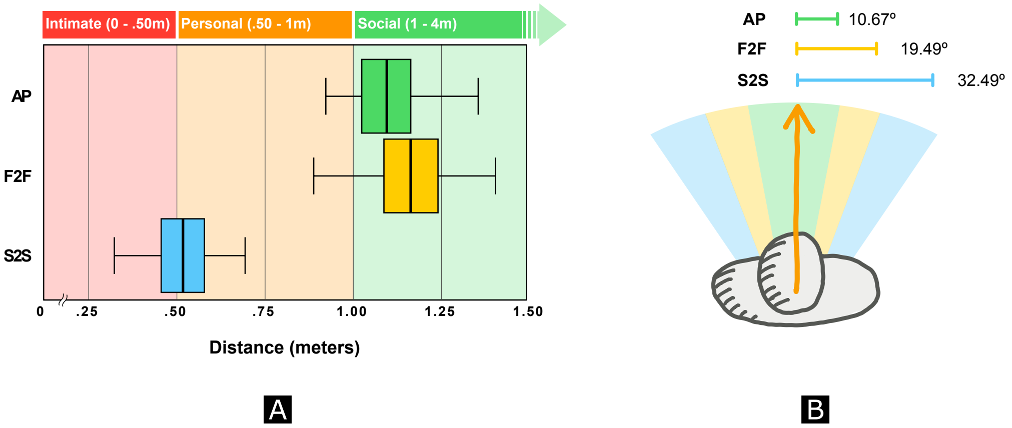


Figure 8.11: Spatial Relationships: A) Distance between participants for each condition; B) Mean values for head angle for each condition.

cally significant distances existed between conditions *AP* ($M = 14.9s, SD = 1.0$), *F2F* ($M = 18.5s, SD = 2.4$), and *S2S* ($M = 17.1s, SD = 1.8$).

8.5.2 Spatial Relationships

Distances between participants in the virtual space, are depicted as horizontal bars in Figure 8.11A. The reported results for distance indicate significant statistical differences between the different conditions ($F(2, .032) = 151.548, p < .001, \eta_p^2 = .947$). Namely, S2S ($M = .509, SD = .021$) was significantly different from both AP ($M = 1.067, SD = .032, p < .001$) and F2F ($M = 1.15, SD = .047, p < .001$).

As for the look angle, depicted in Figure 8.11B, we can also observe statistically significantly differences between the three conditions ($F(1.273, 14.004) = 19.106, p < .001$, $\eta_p^2 = .634$). Both AP ($M = 10.667$,SD = .484, $p < .001$) and F2F ($M = 19.494$,SD = 2.322, $p = .050$) showed smaller values than S2S ($M = 32.487$,SD = 3.703). Additionally, AP was also smaller than F2F ($p = .006$).

8.5.3 User Preferences

We assessed people's subjective responses using questionnaires, which were based in the Network Mind Measure, for Social Presence [57] and in the Single Ease Question [126] regarding usability. They were comprised of twelve statements using a 6-point Likert scale, where a value of 1 means they totally disagreed with a statement

Question	*AP*	*F2F*	*S2S*
1. I felt present in the virtual environment.	5(.25)	5(1)	5.5(1)
2. I felt that my partner was present in the virtual environment.	5(5.5)	6(1)	6(1)
3. I felt that my presence was evident to my partner.	5(.5)	5(1)	6(1)
4. I remained focused on my partner throughout our interaction.	4.5(1)	5(.25)	4.5(2)
5. I remained focused on the workspace throughout our interaction.	5(2)	5(.5)	4.5(2)
6. I felt that my partner could remain focused on me throughout our interaction.	4.5(1)	5(1)	4.5(2)
7. I felt that my partner could remain focused on the workspace throughout our interaction.	5(1.25)	5(0)	5(1)
8. I could understand my partner's actions.	5(1.5)	5(.5)	5(1)
9. I felt that my partner could understand my actions.	5(1.25)	5(.5)	5(1.25)
10. I felt that I was pointing to where I wanted to point.	4(2)	5(1.25)	5(1)
11. It was easy to understand where my partner was pointing.	4.5(2)	4(1.25)	5(1.25)
12. It was easy to understand what was happening in the workspace.	4.5(1.25)	5(1.25)	5(1)

Table 8.1: Results for the user preference questionnaires (Median, Inter-quartile Range).

and a value of 6 signifies that they totally agreed with it. The questionnaires and associated results are summarized in Table 8.1. Interestingly, participants reported no marked preference for either of the three different setups. They seemingly felt comfortable with either setup and would adapt after a preamble negotiation with the other party.

Regarding warping, no participant was seemingly aware of distorted representations of the other party, although most reported artifacts in the kinect-collected point clouds. One person noticed that the image of the other participant was mirrored, due to altered text on their t-shirt. Another person remarked that being inside other people "is confusing" during the actual experiment, although none reported this after the experiment was over. However, we noted that most participants took pains to establish a minimally comfortable distance from the other, in the S2S condition.

8.6 Discussion

Our results show that AP seemingly works best since occlusions are far less likely to exist. This was confirmed by participants' remarks during the experiment and agrees well with classical results from CSCW. However, our approach while being perceived as natural, requires better tracking to avoid depth perception errors. Our approach significantly contributed for participants to spend more time looking forward, since they could simultaneously see both the workspace and the remote collaborator without having to turn their head. Moreover, our approach did not suffer from intimate space violation as observed in S2S. This seemingly bothered some participants, as some asked their partners to move away during the actual experiment. However, this disturbance was not reflected in user preference questionnaires because after they adjusted themselves, virtual body overlap was no longer an issue to the successful completion of the task.

8.6.1 Task Times

AP was on a par with the baseline conditions, although performance figures feature a lower dispersion, which seemingly indicates more predictable outcomes. As we can see from the results, S2S led to fewer errors, which is unsurprising, given that participants were ”forced” to assume similar POVs. Performance results seemingly go against our expectation that AP would best both S2S and F2F. This may be explained because of perfectible tracking (Creepy Tracker showed performance limitations and the number of cameras was insufficient to avoid artifacts. On the other hand, using too many depth cameras severely increases lag, due to network bandwidth limitations). While errors with AP were larger than the baseline condition, location errors remained below 15cm across experiments, which shows that AP was successful in conveying an approximate location, and much better than if we applied no limb deformation, since people would be pointing at the opposite side of the workspace. This is seemingly partially related to the fact that our warping of limbs remains faithful to the real person anatomy, i.e. we chose not to stretch the participant’s limbs, so that distortions were not perceived , a limitation of our approach. So, in cases people are pointing to a location near the closest side of the workspace while standing a bit far from it, our reshaping mechanism could not make the finger to touch the corrected position, since the arm would not have enough length, making the person only point at it from a distance. This naturally contributed for increased errors in perception [61].

Among the current limitations of our experiment, we do not warp the remote person’s gaze. Still, we argue that gaze promotes distal pointing and is thus inherently ambiguous and imprecise. Furthermore, HMDs cover the face of participants. While several approaches exist to remove HMDs and promote eye contact, this was not the focus of the present research.

8.6.2 Distances

Edward Hall's Proxemics theory [56] can be used to classify the interpersonal spatial relationships between both participants. Based on it, Edward Hall identifies four categories of Proxemic distances, namely, intimate, personal, social and public, as depicted on the horizontal bars of Figure 8.11A. We omitted the public proxemic distance since it falls outside of the range of our virtual space around the virtual workbench. Given the high values for presence self reported by all participants, it is not surprising that the distances assumed in the context of the virtual meeting, correlate well with social ranges both in AP and F2F scenarios. However, for unrelated participants the distances informally forced by the S2S condition are seemingly uncomfortable, between the intimate and personal ranges. Unsurprisingly people adopted different distances re: F2F and AP feature seemingly greater (more comfortable) distance between participants than S2S. We could observe from the behaviour evidenced by participants that they would try to make as much distance as possible between themselves prior to undertaking the task proper. This is consistent with the reported sense of presence, and makes their behaviour correspond to that predicted by proxemics.

8.6.3 Look Angle

Look Angle describes how the participants tilted their head to the workspace outer normal. It is worst in S2S, since people need to turn their heads to look at each other. Among the other two conditions, F2F led to larger head angles as compared to AP, since people had to peek between buildings in our experimental setup. The lower values typical of AP and F2F indicate that visual contact leads to more favorable interactions between participants, allowing them to scan the 3D model without losing visual contact, whereas in S2S, participants were "forced" to shift their gaze sideways, to look at the other party, making this situation less comfortable and distracting from their focus on the shared 3D model.

8.7 Conclusions

In virtual meetings, face-to-face formation improves collaboration in 2D task spaces since it promotes the sense of co-presence and facilitates the awareness of other participants and facilitates nonverbal communication. Furthermore, integrating person space, task space, and reference space minimize the need for meeting participants to constantly switch attention between other participants' space and the workspace, thus improving collaboration. However, 3D digital content in a shared virtual environment may cause distractions that affect and impair workspace awareness. This is because participants do not share the same forward-backwards orientation. Also, occlusions can make it unclear where or what the remote person is pointing at. Additionally, contrary points-of-view can result in different perceptions and induce serious communication mishaps. We introduced *Altered Presence*, a design space for mixed reality environments that enables people to engage in object-centered collaboration in a shared workspace using an "above the table" metaphor. AP allows participants to have the same POV/perspective of the workspace; And the remote participant's virtual representation is subtly transformed so that gestures performed in the remote workspace are *virtually* correct in the local workspace. Results from our user study suggest that participants felt more comfortable with AP over both the baseline and S2S conditions, even though artifacts in the tracking system led to higher but manageable errors in deictic co-location.

8.8 Chapter Summary

In this chapter, we introduced *Altered Presence*, a novel approach to improve remote collaboration in shared 3D workspaces by allowing participants to communicate through nonverbal cues while sharing the same perspective, integrating task-, person-, and reference-space seamlessly. Our approach enables face-to-face remote collaboration by distorting the gestures of the remote participant's avatar so that they correctly apply to the local person's reference space. Results from a mixed-reality evaluation suggest that our approach is effective in enabling face-to-face collaboration, and the manipulations were not noticeable, leading to improved awareness, presence, and interactions between remote collaborators. In the next and final chapter, we conclude with the contributions of the research completed in this dissertation and discuss future research opportunities.

9

Conclusions and Future Work

FULL-BODY FACE-TO-FACE TELEPRESENCE promotes the sense of co-presence and improves communication by permitting natural nonverbal cues such as gaze, body posture, and gestures. These nonverbal communication devices impact, to a great extent, collaboration over a virtual shared workspace, since people can use them to demonstrate concepts and refer to positions or task-related artifacts. In remote face-to-face collaboration requiring cooperative selection and manipulation of 3D virtual objects, opposing points-of-view and objects occluding other workspace elements hinders workspace awareness. As a result, previous face-to-face approaches focused on 2D workspaces offer limited workspace awareness when extended to interactions with 3D virtual objects. In this dissertation, we explored perception manipulation to enable seamless face-to-face remote collaboration in 3D object-centered shared workspaces.

This chapter concludes the research presented in this dissertation. Accordingly, in the next sections, we summarize the work conducted in this thesis, discuss its main results, and present directions for future research.

9.1 Dissertation Overview

To accomplish the proposed goal of improving workspace awareness in face-to-face remote collaboration, we employed a triangulation approach that takes in a theoretical, technological, and an experimental design perspective.

First, in Chapter 2, we began by surveying the related work on workspace awareness and nonverbal communication in computer-supported collaborative work. And, then continued reviewing previous literature regarding virtual representations of remote people. A discussion of the state-of-the-art allowed us to identify trends and open challenges that informed the theoretical foundations of our approach presented in Chapter 3. Thus, concluding the first part of our research.

In the second part of this dissertation, we addressed the technological bases needed to design and evaluate our proposed approach. Therefore, in Chapter 4, we presented the Creepy Tracker Toolkit as a rapid prototyping engine focused on enabling co-located and remote user experiences. Then, in Chapter 5, we demonstrated the capabilities of our toolkit in archetype interactive scenarios.

In the final part of this dissertation, we addressed the research developed to validate our research statement from an experimental design perspective.

We started, in Chapter 6, by studying the impact of manipulating the observer's point-of-view, workspace, and embodiment on the ability for collaborators to be aware of the activities occurring in a shared workspace. For this, we introduced the *Negative Space*, a telepresence approach to evaluate instructor-assembler user trials, where participants jointly collaborated in different workspace conditions. Results showed having the same point-of-view is advantageous for people to maintain the same perception of the workspace. However, these trials also revealed that people have difficulty in accurately pinpoint the target of deictic gestures in mixed reality settings.

Taking these previous results into account, we then focused on improving the perception of deixis in mixed reality environments (Chapter 7). And we introduced

Warping Deixis, an approach to rectify the pose of a pointing person to match the way other observers usually perceive deictic gestures. We demonstrated the effectiveness of this technique in an experimental evaluation by comparing our warping method with normal pointing conditions. Results implied that this perception manipulation technique reduces the ambiguity of pointing gestures successfully.

Finally, in Chapter 8, we presented *Altered Presence*, an approach focused on allowing remote participants to collaborate face-to-face while sharing the same perspective, integrating task-, person-, and reference-space. We evaluated *Altered Presence* in a mixed-reality environment, and results suggested that our approach is effective in enabling face-to-face collaboration. Also, body warping manipulations were not noticeable. And overall, our approach can lead to improved awareness, presence, and interactions between collaborators. This dissertation demonstrated how perception manipulation could be used as a technique to reduce the ambiguity of nonverbal communication and enable seamless face-to-face remote interactions in shared 3D workspaces.

9.2 Addressing the Research Questions and Objectives

As explained in the beginning of this dissertation in Chapter 1, the overall research statement that was addressed is: "*Perception manipulation can be used to increase workspace awareness and improve face-to-face remote collaboration in shared 3D workspaces.*" Summarized, we divided this research statement in three research questions:

Question 1: **Can two collaborators share the same perspective of a shared 3D workspace?**

Question 2: **Can perception manipulations improve the understanding of nonverbal communication?**

Question 3: **Can two opposing collaborators share the same understanding of a 3D workspace?**

Consequently, to address these questions, we devised a series of five research objectives to be accomplished in the scope of the research presented in this dissertation.

Next, we describe the progress we have made towards validating our research statement by reviewing the status of the proposed objectives and relating them with the research questions.

Objective 1: We will define perception manipulations for face-to-face collaboration.

To achieve this goal, in Chapter 3, we began by providing a comprehensive review of relevant literature on perception manipulation techniques for improving user experience in mixed-reality. We identified vital approaches to manipulating the characteristics of virtual worlds and people's virtual embodiments. Based on this previous knowledge, we developed the theoretical foundations of our approach. We determined that the integration of task-, person-, and reference space could be extended to 3D object-centered collaboration assuming that people could correctly understand the reference space presented. As a consequence, we identified perspective-sharing, workspace warping, and body warping as the atomic tools to shape how people perceive their environment without breaking workspace awareness. The efforts to complete this goal shaped our technological requirements and the experimental design. We consider this objective **accomplished**.

Objective 2: We will design and implement rapid prototyping tools to make building remote interactions accessible.

We **accomplished** this goal in Part II of this dissertation. As described in Chapter 4, we designed and implemented the *Creepy Tracker Toolkit* allowing the rapid-prototyping of co-located and remote user experiences. Our toolkit encapsulates people's full-body positional tracking data and point-cloud avatar-based data into streams of high-level information. Thus, simplifying the development efforts, as demonstrated by the sample usage scenarios presented in Chapter 5. Furthermore, the Creepy Tracker employs multiple commodity depth cameras to do away with equipping users with intrusive reflective markers and infrared trackers. A system performance evaluation showed that our toolkit provides reliable tracking data despite being slightly less precise than marker-based optical systems. And, the network distributed architecture allows for deploying multiple tracker instances, making it easier for students and researchers to start prototyping new telepresence approaches. The easy access to body tracking information and virtual avatars allowed us to develop the approaches meant to validate our research statement.

Objective 3: **We will evaluate workspace awareness using variations of a shared workspace, individual point-of-view, and remote person's virtual representation.**

To address this goal, Chapter 6 introduced Negative Space, a telepresence approach that creates a virtual 3D workspace between two physical spaces, where interactions with 3D objects can occur. We developed this collaborative design space to evaluate different combinations of embodiment, workspace, and point-of-view manipulations. The aim was to identify the ideal conditions to preserve the reference space between participants while promoting workspace awareness. Therefore, the presented user evaluations were comprised of instructor-assembler trials, where participants jointly solved a different puzzle in all workspace conditions. Results suggest that the combination of sharing the same point-of-view with mirroring the embodiment of remote people may participants to use gestures naturally as a complement to clear verbal instructions. However, these results showed that participants can successfully collaborate in the *Negative Space*, and suggest that having the same point-of-view is beneficial. And, thus, this objective was **accomplished**. Despite these results, this completion of this objective was not sufficient to answer **Question 1** and **Question 2**.

Objective 4: **We will contribute body manipulation techniques to improve deictic gestures.**

In Chapter 7, we demonstrated *Warping Deixis*, a body warping technique to improve how pointing gestures are interpreted in mixed-reality environments. Warping Deixis allows for body warping adjustments to the avatar of a person performing a pointing gesture to make distal referents more explicit and more accessible to be identified by an observing collaborator. First, in this enclosed body of research, we developed and evaluated the technology to redirect arm poses that can be used not only in avatar-based embodiments but also in high definition point-cloud mesh-based virtual representations. The psychological fundamentals of how people perceive pointing gestures motivated our approach, which relies on a predictive Bayesian model to know where the pointer's arm should be to reduce errors in identifying the target. Indeed, results from a user evaluation showed that our body warping approach does reduce errors drastically when identifying targets. This result alone makes us consider this objective **accomplished**. Furthermore, evaluation participants were unable to distinguish any implausible movements or actions from the person who was pointing. *Warping Deixis* partially addressed **Question 2** in the sense that it only proves that perception manipulations are effective in improving distal pointing gestures.

Objective 5: **We will contribute perception manipulation techniques to improve close face-to-face collaboration.**

We addressed this final goal in Chapter 8, where we proposed *Altered Presence*, our approach to improving face-to-face remote collaboration in mixed reality shared 3D task spaces. *Altered Presence* focused on addressing the core research problem of this dissertation by allowing participants to communicate through nonverbal cues while sharing the same perspective, integrating task-, person-, and reference-space in a seamless manner. *Altered Presence* contributes to improving collaboration since both participants are always aware of the state of the workspace while being able to communicate non verbally We refined the approach presented in Chapter 7 to enable more complex body warping manipulations, namely a total reshaping of people's upper limbs. Furthermore, *Altered Presence* is a collaborative interactive space that distorts gestures to present a corrected virtual representation of those gestures on the local participant reference space. We performed a user evaluation aimed at verifying whether people could maintain a shared understanding of the workspace while interacting in a face-to-face formation and sharing the same perspective of the workspace. This evaluation also featured an investigation on co-presence and the ability of participants to be aware of their counterparts' personal space. Results suggest that our approach is effective in dealing with workspace occlusions and significantly contributed for participants to spend more time looking forward to the task space and to their remote collaborator. And therefore improving close face-to-face collaborative interactions, which makes us consider this objective to be **accomplished**.

In accomplishing these objectives, we achieved the path proposed in Chapter 1 to prove our research statement.

The research developed in **Objective 3** and **5** suggested not only that a shared perspective is natural and do not interfere with collaboration, but also is useful in guaranteeing that participants share the same understanding of the actions occurring in the task space. Yet, sharing the same point-of-view alone is not a condition for the naturality of the technique. It is the combination of the shared perspective with the other participant's actions corrected to the local reference space that makes our approach not break the suspension of disbelief. Since people accept that there is nothing wrong with what they perceive because everyone perceives the same thing. Therefore, this dissertation is able to confirm **Question 1**.

Both the *Warping Deixis* and the *Altered Presence* approaches exploit perception manipulation techniques the understanding of nonverbal intentional communication devices, thus addressing **Question 2**. As demonstrated in Chapter 7, *Warping Deixis* significantly reduces distal pointing misunderstandings. And, correspondingly, *Altered Presence* performs a reference space correction of proximal pointing gestures. *Altered Presence* significantly contributed to people to face each other, which, in combination with the reference space correction, enables participants always to be aware and understand each other's gestures. Furthermore, the perception manipulation techniques employed in the *Altered Presence* stops the occlusion problem introduced in Chapter 1 and are capable of supporting gestures communicating spatial positions, thus answering **Question 3**.

Finally, the final results suggest that using perception manipulations in a collaborative 3D workspace environment is best since collaborators spent more time facing each other when interacting in a shared workspace, while being able to communicate freely using nonverbal cues. Therefore, we can prove our research statement.

9.3 Future Work

This work focused on the study remote face-to-face collaboration in shared 3D workspaces. While we succeeded in improving workspace awareness using perception manipulation techniques, we describe possible directions for future research extending the work presented in this dissertation.

Exploration of Altered Presence design space: In this dissertation, we demonstrated the benefits of collaborating face-to-face in a shared 3D ”above-the-table” workspace. We focused on workspace awareness and awareness of remote people's actions and nonverbal cues while contributing a design space for remote collaboration. In that matter, the contributions of this dissertation open different opportunities for future research. Since the scope of our thesis focused on improving the exchange of information between local and remote collaborators, studying selection and manipulation techniques for design and review of 3D models combined with novel turn-taking approaches could help improve remote collaboration in new exciting ways. Furthermore, our research did not address group interactions because this fell out of scope. Yet, it would be interesting to venture into group interactions around shared 3D workspaces and study if perception manipulation techniques are sufficient to deal with a mixture of local and remote people.

Improve body manipulations: Future work using better depth cameras will investigate whether deictic disparities can improve in the findings of this dissertation. A promising direction is to study other forms of body warping, focusing on ensuring that manipulated actions do not force people to convey different meanings than they originally intended and further explore warping in real use case scenarios. Another direction is looking at elastic exaggerated yet minimally perceptible distortions to ascertain whether they have the potential to either improve or hinder collaboration. The motivation is to find a balance between what is real and familiar, with what is fabricated and overkill without breaking the suspension of disbelief. While historically virtual reality research has focused on faithfully reproducing reality, it is our strong belief that manipulating perception both in subtle and not so subtle ways, can be used to greater advantage in the future.

Learning activities supported by body warping: The body warping techniques proposed in this dissertation can be extended for people to perceive their bodies performing gestures our of their control in mixed-reality environments. Seeing our own body performing complex tasks for the first time, could have the potential to reduce the need for training and could promote walk-up use of sophisticated professional machinery. The body warping techniques proposed in this dissertation can be extended for people to perceive their bodies performing gestures our of their control in mixed-reality environments. Seeing our own body performing complex tasks for the first time, has the potential to reduce the feed for training and could promote walk-up use of sophisticated professional machinery. These body manipulations could be automatic or controlled by a remote expert. Hence, a future research direction is studying the impact on learning from observing instructional actions in a first-person point-of-view rather than observing others. As mixed-reality technologies become ubiquitous in the workplace, we envision future scenarios where instead of extensive training, people may learn job-specific tasks by observing themselves doing it. Furthermore, it would be interesting to study how people perceive agency ("who did what") in such scenarios.

9.4 Final Remarks

In conclusion, we have accomplished our research goals and validated our thesis. We have shown that perception manipulation techniques can successfully guarantee mutual perception when remote people collaborate face-to-face in shared 3D workspaces.

We have an optimistic expectation that our insights can help researchers and designers to create better remote presence interactions. With this work, we not only contributed technology to ease the development of remote experiences but also revealed the notion that interactive telepresence systems do not need to solely focus on communicating what is real when transferring nonverbal communication. Instead, telepresence approaches can deliver altered and improved versions of nonverbal cues, but so that the message to be transmitted continues to remain genuine.

References

[1] Abtahi, P., Landry, B., Yang, J. J., Pavone, M., Follmer, S., & Landay, J. A. (2019). Beyond the force: Using quadcopters to appropriate objects and the environment for haptics in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 359).: ACM.

[2] Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human factors*, 37(1), 85–104.

[3] Aggarwal, J. K. & Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3), 16.

[4] Annett, M., Grossman, T., Wigdor, D., & Fitzmaurice, G. (2011). Medusa: a proximity-aware multi-touch tabletop. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 337–346).: ACM.

[5] Antifakos, S. & Schiele, B. (2002). Beyond position awareness. *Personal and Ubiquitous Computing*, 6(5-6), 313–317.

[6] Argelaguet, F. & Andujar, C. (2009). Visual feedback techniques for virtual pointing on stereoscopic displays. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (pp. 163–170).: ACM.

[7] Ark, W. S. & Selker, T. (1999). A look at human interaction with pervasive computers. *IBM systems journal*, 38(4), 504–507.

[8] Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff, T., Becker, T., Holz, C., & Baudisch, P. (2010). Multitoe: High-precision interaction with back-projected floors based on high-resolution multi-touch input. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10 (pp. 209–218). New York, NY, USA: ACM.

[9] Azer, S. A. & Azer, S. (2016). 3d anatomy models and impact on learning: A review of the quality of the literature. *Health Professions Education*, 2(2), 80 – 98.

[10] Azimi, M. (2012). *Skeletal Joint Smoothing White Paper*. Technical report. http://msdn.microsoft.com/en-us/library/jj131429.aspx.

[11] Azmandian, M., Grechkin, T., Bolas, M., & Suma, E. (2016a). The redirected walking toolkit: a unified development platform for exploring large virtual environments. In *2016 IEEE 2nd Workshop on Everyday Virtual Reality (WEVR)* (pp. 9–14).: IEEE.

[12] Azmandian, M., Hancock, M., Benko, H., Ofek, E., & Wilson, A. D. (2016b). 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).: ACM.

[13] Ballendat, T., Marquardt, N., & Greenberg, S. (2010). Proxemic interaction: Designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10 (pp. 121–130). New York, NY, USA: ACM.

[14] Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, 15(6), 415–419.

[15] Bangerter, A. & Oppenheimer, D. M. (2006). Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture*, 6(1), 85–102.

[16] Beck, E. E. (1993). A survey of experiences of collaborative writing. In *Computer supported collaborative writing* (pp. 87–112). Springer.

[17] Beck, S., Kunert, A., Kulik, A., & Froehlich, B. (2013). Immersive group-to-group telepresence. *Visualization and Computer Graphics, IEEE Transactions on*.

[18] Bekker, M. M., Olson, J. S., & Olson, G. M. (1995). Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *Proceedings of the 1st Conference on Designing Interactive Systems: Processes, Practices, Methods, & Techniques*, DIS '95 (pp. 157–166). New York, NY, USA: ACM.

[19] Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C., & Snowdon, D. (1995). User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95 (pp. 242–249). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co.

[20] Benko, H., Jota, R., & Wilson, A. (2012). Miragetable: Freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12 (pp. 199–208). New York, NY, USA: ACM.

[21] Billinghurst, M. & Kato, H. (1999). Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality* (pp. 261–284).

[22] Bolt, R. A. (1980). Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '80 (pp. 262–270). New York, NY, USA: ACM.

[23] Bouma, H. & Andriessen, J. (1968). Perceived orientation of isolated line segments. *Vision Research*, 8(5), 493–507.

[24] Bowman, D. A., Johnson, D. B., & Hodges, L. F. (2001). Testbed evaluation of virtual environment interaction techniques. *Presence: Teleoperators & Virtual Environments*, 10(1), 75–95.

[25] Bremner, P. & Leonards, U. (2016). Iconic gestures for robot avatars, recognition and integration with speech. *Frontiers in psychology*, 7, 183.

[26] Butterworth, G. (2003). Pointing is the royal road to language for babies. In *Pointing* (pp. 17–42). Psychology Press.

[27] Butterworth, G. & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, 18(1), 25–50.

[28] Buxton, B. (2009). *Mediaspace – Meaningspace – Meetingspace*, (pp. 217–231). Springer London: London.

[29] Buxton, W. (1992). Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface* (pp. 123–129).

[30] Buxton, W. (1997). Living in augmented reality: Ubiquitous media and reactive environments. *Video Mediated Communication.*, (pp. 363–384).

[31] Casiez, G., Roussel, N., & Vogel, D. (2012). 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2527–2530).: ACM.

[32] Chen, W.-C., Towles, H., Nyland, L., Welch, G., & Fuchs, H. (2000). Toward a compelling sensation of telepresence: Demonstrating a portal to a distant (static) office. In *Proceedings of the Conference on Visualization '00*, VIS '00 (pp. 327–333). Los Alamitos, CA, USA: IEEE Computer Society Press.

[33] Congdon, B. J., Wang, T., & Steed, A. (2018). Merging environments for shared spaces in mixed reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (pp. 1–8).

[34] Dey, A. K., Abowd, G. D., & Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-computer interaction*, 16(2), 97–166.

[35] Dourish, P. & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work* (pp. 107–114).: ACM.

[36] Duval, T., Nguyen, T. T. H., Fleury, C., Chauffaut, A., Dumont, G., & Gouranton, V. (2014). Improving awareness for 3d virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues. *Journal on Multimodal User Interfaces*, 8(2), 187–197.

[37] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1), 32–64.

[38] Faber, F., Bennewitz, M., Eppner, C., Gorog, A., Gonsior, C., Joho, D., Schreiber, M., & Behnke, S. (2009). The humanoid museum tour guide robotinho. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on* (pp. 891–896).: IEEE.

[39] Fails, J. A. & Olsen, J. D. (2002). Light widgets: interacting in every-day spaces. In *Proceedings of the 7th international conference on Intelligent user interfaces* (pp. 63–69).: ACM.

[40] Feick, M., Mok, T., Tang, A., Oehlberg, L., & Sharlin, E. (2018). Perspective on and re-orientation of physical proxies in object-focused remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 281).: ACM.

[41] Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., & Kramer, A. D. (2004a). Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19(3), 273–309.

[42] Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., & Kramer, A. D. I. (2004b). Gestures over video streams to support remote collaboration on physical tasks. *Hum.-Comput. Interact.*, 19(3), 273–309.

[43] Galegher, J. & Kraut, R. E. (1994). Computer-mediated communication for intellectual teamwork: An experiment in group writing. *Information systems research*, 5(2), 110–138.

[44] Genest, A. & Gutwin, C. (2011). *Characterizing Deixis over Surfaces to Improve Remote Embodiments*, (pp. 253–272). Springer London: London.

[45] Genest, A. M., Gutwin, C., Tang, A., Kalyn, M., & Ivkovic, Z. (2013). Kinectarms: a toolkit for capturing and displaying arm embodiments in distributed tabletop groupware. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 157–166).: ACM.

[46] Gibson, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of experimental psychology*, 16(1), 1.

[47] Gotsch, D., Zhang, X., Merritt, T., & Vertegaal, R. (2018). Telehuman2: A cylindrical light field teleconferencing system for life-size 3d human telepresence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18 (pp. 522:1–522:10). New York, NY, USA: ACM.

[48] Greenberg, S., Gutwin, C., & Cockburn, A. (1996a). Awareness through fisheye views in relaxed-wysiwis groupware. In *Graphics interface*, volume 96 (pp. 28–38).

[49] Greenberg, S., Gutwin, C., & Roseman, M. (1996b). Semantic telepointers for groupware. In *Computer-Human Interaction, 1996. Proceedings., Sixth Australian Conference on* (pp. 54–61).: IEEE.

[50] Greenhalgh, C. & Benford, S. (1995). Massive: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3), 239–261.

[51] Grønbæk, K., Iversen, O. S., Kortbek, K. J., Nielsen, K. R., & Aagaard, L. (2007). Igamefloor: A platform for co-located collaborative games. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, ACE '07 (pp. 64–71). New York, NY, USA: ACM.

[52] Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Van Gool, L., Lang, S., Strehlke, K., Moere, A. V., & Staadt, O. (2003). Blue-c: A spatially immersive display and 3d video portal for telepresence. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03 (pp. 819–827). New York, NY, USA: ACM.

[53] Gugenheimer, J., Stemasov, E., Frommel, J., & Rukzio, E. (2017). 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*, CHI '17 (pp. 4021–4033). New York, NY, USA: ACM.

[54] Gutwin, C. & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3), 411–446.

[55] Gutwin, C., Greenberg, S., & Roseman, M. (1996). Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation. In *People and Computers XI* (pp. 281–298). Springer.

[56] Hall, E. T. (1966). *The hidden dimension*. Doubleday & Co.

[57] Harms, C. & Biocca, F. (2004). Internal consistency and reliability of the networked minds measure of social presence.

[58] Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., & Rogers, Y. (2009). Around the table: Are multiple-touch surfaces better than single-touch for children's collaborative interactions? In *Proceedings of the 9th international conference on Computer supported collaborative learning-Volume 1* (pp. 335–344).: International Society of the Learning Sciences.

[59] Harrison, S. (2009). *A Brief History of Media Space Research and Mediated Life*, (pp. 9–16). Springer London: London.

[60] Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2), 262–271.

[61] Herbort, O. & Kunde, W. (2016). Spatial (mis-) interpretation of pointing gestures to distal referents. *Journal of Experimental Psychology: Human Perception and Performance*, 42(1), 78.

[62] Herbort, O. & Kunde, W. (2018). How to point and to interpret pointing gestures? instructions can reduce pointer–observer misunderstandings. *Psychological research*, 82(2), 395–406.

[63] Higuchi, K., Chen, Y., Chou, P. A., Zhang, Z., & Liu, Z. (2015). Immerseboard: Immersive telepresence experience using a digital whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2383–2392).: ACM.

[64] Hollan, J. & Stornetta, S. (1992). Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 119–125).: ACM.

[65] Ichikawa, Y., Okada, K.-i., Jeong, G., Tanaka, S., & Matsushita, Y. (1995). Majic videoconferencing system: experiments, evaluation and improvement. In *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work ECSCW'95* (pp. 279–292).: Springer.

[66] Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., & Schiele, B. (2016). Deepercut: A deeper, stronger, and faster multi-person pose estimation model. *arXiv preprint arXiv:1605.03170*.

[67] Ishihara, S. (1960). *Tests for colour-blindness*. Kanehara Shuppan Company.

[68] Ishii, H. & Kobayashi, M. (1992). Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92 (pp. 525–532). New York, NY, USA: ACM.

[69] Ishii, H., Kobayashi, M., & Arita, K. (1994). Interactive design of seamless collaboration media. *Communications of the ACM*, 37(8), 83–98.

[70] Ishii, H., Kobayashi, M., & Grudin, J. (1993). Integration of interpersonal space and shared workspace: Clearboard design and experiments. *ACM Transactions on Information Systems (TOIS)*, 11(4), 349–375.

[71] Ives, H. E., Gray, F., & Baldwin, M. W. (1930). Image transmission system for two-way television. *The Bell System Technical Journal*, 9(3), 448–469.

[72] Jo, D., Kim, K.-H., & Kim, G. J. (2016). Effects of avatar and background representation forms to co-presence in mixed reality (mr) tele-conference systems. In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments* (pp.12̃).: ACM.

[73] Jokela, T., Ojala, J., & Olsson, T. (2015). A diary study on combining multiple information devices in everyday activities and tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15 (pp. 3903–3912). New York, NY, USA: ACM.

[74] Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., & Debevec, P. (2009). Achieving eye contact in a one-to-many 3d video teleconferencing system. *ACM Transactions on Graphics (TOG)*, 28(3), 64.

[75] Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., Ofek, E., MacIntyre, B., Raghuvanshi, N., & Shapira, L. (2014). Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14 (pp. 637–644). New York, NY, USA: ACM.

[76] Jota, R., Nacenta, M. A., Jorge, J. A., Carpendale, S., & Greenberg, S. (2010a). A comparison of ray pointing techniques for very large displays. In *Proceedings of graphics interface 2010* (pp. 269–276).: Canadian Information Processing Society.

[77] Jota, R., Nacenta, M. A., Jorge, J. A., Carpendale, S., & Greenberg, S. (2010b). A comparison of ray pointing techniques for very large displays. In *Proceedings of Graphics Interface 2010*, GI '10 (pp. 269–276). Toronto, Ont., Canada, Canada: Canadian Information Processing Society.

[78] Junuzovic, S., Inkpen, K., Tang, J., Sedlins, M., & Fisher, K. (2012). To see or not to see: A study comparing four-way avatar, video, and audio conferencing for work. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, GROUP '12 (pp. 31–34). New York, NY, USA: ACM.

[79] Kendon, A. (2010). Spacing and orientation in co-present interaction. In *Development of Multimodal Interfaces: Active Listening and Synchrony* (pp. 1–15). Springer.

[80] Kirk, D., Crabtree, A., & Rodden, T. (2005). Ways of the hands. In *ECSCW 2005* (pp. 1–21).: Springer.

[81] Kita, S. (2003). *Pointing: Where language, culture, and cognition meet.* Psychology Press.

[82] Knill, D. C. & Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12), 712–719.

[83] Kooima, R. (2009). Generalized perspective projection. *J. Sch. Electron. Eng. Comput. Sci.*

[84] Körding, K. P. & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971), 244.

[85] Krauss, R. M. & Fussell, S. R. (1990). Mutual knowledge and communicative effectiveness. *Intellectual teamwork: Social and technological foundations of cooperative work*, (pp. 111–146).

[86] Kunz, A., Nescher, T., & Kuchler, M. (2010). Collaboard: a novel interactive electronic whiteboard for remote collaboration with people on content. In *Cyberworlds (CW), 2010 International Conference on* (pp. 430–437).: IEEE.

[87] Langbehn, E., Bruder, G., & Steinicke, F. (2016). Subliminal reorientation and repositioning in virtual reality during eye blinks. In *Proceedings of the 2016 symposium on spatial user interaction* (pp. 213–213).: ACM.

[88] Lecuyer, A., Coquillart, S., Kheddar, A., Richard, P., & Coiffet, P. (2000). Pseudo-haptic feedback: Can isometric input devices simulate force feedback? In *Proceedings of the IEEE Virtual Reality 2000 Conference*, VR '00 (pp. 83–). Washington, DC, USA: IEEE Computer Society.

[89] Lee, S., Seo, J., Kim, G. J., & Park, C.-M. (2003). Evaluation of pointing techniques for ray casting selection in virtual environments. In *Third international conference on virtual reality and its application in industry*, volume 4756 (pp. 38–45).: International Society for Optics and Photonics.

[90] Leithinger, D., Follmer, S., Olwal, A., & Ishii, H. (2014). Physical telepresence: Shape capture and display for embodied, computer-mediated remote collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14 (pp. 461–470). New York, NY, USA: ACM.

[91] Li, J., Greenberg, S., Sharlin, E., & Jorge, J. (2014a). Interactive two-sided transparent displays: Designing for collaboration. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14 (pp. 395–404). New York, NY, USA: ACM.

[92] Li, J., Greenberg, S., Sharlin, E., & Jorge, J. (2014b). Interactive two-sided transparent displays: Designing for collaboration. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14 (pp. 395–404). New York, NY, USA: ACM.

[93] Liles, K. R., Perry, C. D., Craig, S. D., & Beer, J. M. (2017). Student perceptions: The test of spatial contiguity and gestures for robot instructors. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17 (pp. 185–186). New York, NY, USA: ACM.

[94] Lécuyer, A. (2009). Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback. *Presence: Teleoperators and Virtual Environments*, 18(1), 39–53.

[95] Maimone, A. & Fuchs, H. (2011). Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on* (pp. 137–146).: IEEE.

[96] Marquardt, N., Ballendat, T., Boring, S., Greenberg, S., & Hinckley, K. (2012a). Gradual engagement: facilitating information exchange between digital devices as a function of proximity. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (pp. 31–40).: ACM.

[97] Marquardt, N., Diaz-Marino, R., Boring, S., & Greenberg, S. (2011a). The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 315–326).

[98] Marquardt, N., Diaz-Marino, R., Boring, S., & Greenberg, S. (2011b). The proximity toolkit: Prototyping proxemic interactions in ubiquitous computing ecologies. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11 (pp. 315–326). New York, NY, USA: ACM.

[99] Marquardt, N. & Greenberg, S. (2012). Informing the design of proxemic interactions. *IEEE Pervasive Computing*, 11(2), 14–23.

[100] Marquardt, N., Hinckley, K., & Greenberg, S. (2012b). Cross-device interaction via micro-mobility and f-formations. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12 (pp. 13–22). New York, NY, USA: ACM.

[101] Marquardt, N., Jota, R., Greenberg, S., & Jorge, J. (2011c). The continuous interaction space: interaction techniques unifying touch and gesture on and above a digital surface. *Human-Computer Interaction–INTERACT 2011*, (pp. 461–476).

[102] Marshall, P., Hornecker, E., Morris, R., Dalton, N. S., & Rogers, Y. (2008). When the fingers do the talking: A study of group participation with varying constraints to a tabletop interface. In *Horizontal Interactive Human Computer Systems, 2008. TABLETOP 2008. 3rd IEEE International Workshop on* (pp. 33–40).: IEEE.

[103] Matthews, T., Dey, A. K., Mankoff, J., Carter, S., & Rattenbury, T. (2004). A toolkit for managing user attention in peripheral displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology* (pp. 247–256).: ACM.

[104] McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago press.

[105] Mendes, D., Relvas, F., Ferreira, A., & Jorge, J. (2016). The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology*, VRST '16 (pp. 261–268). New York, NY, USA: ACM.

[106] Morikawa, O. & Maesako, T. (1998). Hypermirror: Toward pleasant-to-use video mediated communication system. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, CSCW '98 (pp. 149–158). New York, NY, USA: ACM.

[107] Nguyen, D. & Canny, J. (2005). Multiview: Spatially faithful group video conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05 (pp. 799–808). New York, NY, USA: ACM.

[108] Noma, T., Zhao, L., & Badler, N. I. (2000). Design of a virtual human presenter. *IEEE Comput. Graph. Appl.*, 20(4), 79–85.

[109] Norman, D. A. (1993). Things that make us smart.

[110] Oda, O., Elvezio, C., Sukan, M., Feiner, S., & Tversky, B. (2015). Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 405–415).: ACM.

[111] Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P. L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P. A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S. B., Kohli, P., Lutchyn, Y., Keskin, C., & Izadi, S. (2016). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16 (pp. 741–754). New York, NY, USA: ACM.

[112] Pechmann, T. & Deutsch, W. (1982). The development of verbal and nonverbal devices for reference. *Journal of experimental child psychology*, 34(2), 330–341.

[113] Pejsa, T., Kantor, J., Benko, H., Ofek, E., & Wilson, A. (2016). Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16 (pp. 1716–1725). New York, NY, USA: ACM.

[114] Pierce, J. S., Forsberg, A. S., Conway, M. J., Hong, S., Zeleznik, R. C., & Mine, M. R. (1997). Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics* (pp. 39–ff).: ACM.

[115] Piumsomboon, T., Day, A., Ens, B., Lee, Y., Lee, G., & Billinghurst, M. (2017a). Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications* (pp.16).: ACM.

[116] Piumsomboon, T., Lee, G. A., Hart, J. D., Ens, B., Lindeman, R. W., Thomas, B. H., & Billinghurst, M. (2018). Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18 (pp. 46:1–46:13). New York, NY, USA: ACM.

[117] Piumsomboon, T., Lee, Y., Lee, G., & Billinghurst, M. (2017b). Covar: a collaborative virtual and augmented reality system for remote collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies* (pp.3).: ACM.

[118] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., & Fuchs, H. (1998). The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98 (pp. 179–188). New York, NY, USA: ACM.

[119] Razzaque, S., Kohn, Z., & Whitton, M. C. (2005). *Redirected walking*. Citeseer.

[120] Rekimoto, J. & Nagao, K. (1995). 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).: ACM.

[121] Robinson, M. (1991). Computer-supported cooperative work: Cases and concepts. In *proceedings of Groupware*, volume 91 (pp. 59–75).

[122] Robles-De-La-Torre, G. (2006). The importance of the sense of touch in virtual and real environments. *Ieee Multimedia*, 13(3), 24–30.

[123] Rock, I. & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, 143(3606), 594–596.

[124] Salomon, A. D. (1947). Visual field factors in the perception of direction. *The American journal of psychology*, 60(1), 68–88.

[125] Samad, M., Gatti, E., Hermes, A., Benko, H., & Parise, C. (2019). Pseudo-haptic weight: Changing the perceived weight of virtual objects by manipulating control-display ratio. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 320).: ACM.

[126] Sauro, J. & Dumas, J. S. (2009). Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1599–1608).: ACM.

[127] Schilit, B., Adams, N., & Want, R. (1994). Context-aware computing applications. In *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on* (pp. 85–90).: IEEE.

[128] Schmidt, C. L. (1999). Adult understanding of spontaneous attention-directing events: What does gesture contribute? *Ecological Psychology*, 11(2), 139–174.

[129] Sellen, A. J. (1992). Speech patterns in video-mediated conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92 (pp. 49–59). New York, NY, USA: ACM.

[130] Seyed, T., Azazi, A., Chan, E., Wang, Y., & Maurer, F. (2015). Sod-toolkit: A toolkit for interactively prototyping and developing multi-sensor, multi-device environments. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, ITS '15 (pp. 171–180). New York, NY, USA: ACM.

[131] Shen, Z. & Wu, Y. (2016). Investigation of practical use of humanoid robots in elderly care centres. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, HAI '16 (pp. 63–66). New York, NY, USA: ACM.

[132] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., & Moore, R. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1), 116–124.

[133] Slater, M. & Usoh, M. (1994). Body centred interaction in immersive virtual environments. *Artificial life and virtual reality*, 1, 125–148.

[134] Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., & Maciocci, G. (2013). Bethere: 3d mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 179–188).: ACM.

[135] Sousa, M., Mendes, D., Anjos, R. K. D., Medeiros, D., Ferreira, A., Raposo, A., Pereira, J. a. M., & Jorge, J. (2017). Creepy tracker toolkit for context-aware interfaces. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17 (pp. 191–200). New York, NY, USA: ACM.

[136] Sousa, M., Mendes, D., Ferreira, A., Pereira, J. M., & Jorge, J. (2015). Eery space: Facilitating virtual meetings through remote proxemics. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2015* (pp. 622–629). Cham: Springer International Publishing.

[137] Stefik, M., Bobrow, D. G., Foster, G., Lanning, S., & Tatar, D. (1987a). Wysiwis revised: early experiences with multiuser interfaces. *ACM Transactions on Information Systems (TOIS)*, 5(2), 147–167.

[138] Stefik, M., Foster, G., Bobrow, D. G., Kahn, K., Lanning, S., & Suchman, L. (1987b). Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. *Commun. ACM*, 30(1), 32–47.

[139] Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., Hagita, N., & Anzai, Y. (2006). Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection science*, 18(4), 379–402.

[140] Tang, A., Boyle, M., & Greenberg, S. (2004). Display and presence disparity in mixed presence groupware. In *Proceedings of the fifth conference on Australasian user interface-Volume 28* (pp. 73–82).: Australian Computer Society, Inc.

[141] Tang, A., Neustaedter, C., & Greenberg, S. (2007). Videoarms: embodiments for mixed presence groupware. In *People and Computers XX—Engage* (pp. 85–102). Springer.

[142] Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J. C., & Buxton, B. (2010). Three's company: understanding communication channels in three-way distributed collaboration. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 271–280).: ACM.

[143] Tang, J. C. & Minneman, S. (1991). Videowhiteboard: Video shadows to support remote collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91 (pp. 315–322). New York, NY, USA: ACM.

[144] Tang, J. C. & Minneman, S. L. (1990). Videodraw: A video interface for collaborative drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90 (pp. 313–320). New York, NY, USA: ACM.

[145] Tang, R., Alizadeh, H., Tang, A., Bateman, S., & Jorge, J. A. (2014). Physio@home: Design explorations to support movement guidance. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14 (pp. 1651–1656). New York, NY, USA: ACM.

[146] Tanner, P. & Shah, V. (2010). Improving remote collaboration through side-by-side telepresence. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3493–3498).: ACM.

[147] Taylor, J. L. & McCloskey, D. (1988). Pointing. *Behavioural Brain Research*.

[148] Terveen, L. G. (1995). Overview of human-computer collaboration. *Knowledge-Based Systems*, 8(2-3), 67–81.

[149] Tolani, D. & Badler, N. I. (1996). Real-time inverse kinematics of the human arm. *Presence: Teleoperators & Virtual Environments*, 5(4), 393–401.

[150] Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child development*, 78(3), 705–722.

[151] Vermeulen, J., Luyten, K., Coninx, K., Marquardt, N., & Bird, J. (2015). Proxemic flow: Dynamic peripheral floor visualizations for revealing and mediating large surface interactions. In *Human-Computer Interaction* (pp. 264–281).: Springer.

[152] Vogel, D. & Balakrishnan, R. (2004). Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04 (pp. 137–146). New York, NY, USA: ACM.

[153] Voida, S., Podlaseck, M., Kjeldsen, R., & Pinhanez, C. (2005). A study on the manipulation of 2d objects in a projector/camera-based augmented reality environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 611–620).: ACM.

[154] Ware, C. & Lowther, K. (1997). Selection using a one-eyed cursor in a fish tank vr environment. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 4(4), 309–322.

[155] Weiser, M. (1991). The computer for the 21st century. *Scientific american*, 265(3), 94–104.

[156] Weiser, M. (2002). The computer for the 21st century. *IEEE pervasive computing*, 1(1), 19–25.

[157] Wen, W.-C., Towles, H., Nyland, L., Welch, G., & Fuchs, H. (2000). Toward a compelling sensation of telepresence: demonstrating a portal to a distant (static) office. In *Proceedings Visualization 2000. VIS 2000 (Cat. No.00CH37145)*, VIS '00 (pp. 327–333). Los Alamitos, CA, USA: IEEE Computer Society Press.

[158] Wilson, A. D. (2007). Depth-sensing video cameras for 3d tangible tabletop interaction. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on* (pp. 201–204).: IEEE.

[159] Wilson, A. D. & Benko, H. (2010). Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (pp. 273–282).: ACM.

[160] Wnuczko, M. & Kennedy, J. M. (2011). Pivots for pointing: Visually-monitored pointing has higher arm elevations than pointing blindfolded. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5), 1485.

[161] Wolff, R., Roberts, D. J., Steed, A., & Otto, O. (2007). A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology*, 29(1), 11–26.

[162] Wong, N. & Gutwin, C. (2010). Where are you pointing?: the accuracy of deictic pointing in cves. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1029–1038).: ACM.

[163] Wong, N. & Gutwin, C. (2014). Support for deictic pointing in cves: still fragmented after all these years'. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 1377–1387).: ACM.

[164] Wood, E., Taylor, J., Fogarty, J., Fitzgibbon, A., & Shotton, J. (2016). Shadowhands: High-fidelity remote hand gesture visualization using a hand tracker. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*, ISS '16 (pp. 77–84). New York, NY, USA: ACM.

[165] Wu, C.-J., Houben, S., & Marquardt, N. (2017). Eaglesense: Tracking people and devices in interactive spaces using real-time top-view depth-sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17 (pp. 3929–3942). New York, NY, USA: ACM.

[166] Zillner, J., Rhemann, C., Izadi, S., & Haller, M. (2014). 3d-board: A whole-body remote collaborative whiteboard. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14 (pp. 471–479). New York, NY, USA: ACM.

A
Appendices

A.1 Elements of Workspace Awareness

Table A.1: Elements of workspace awareness relating to the present (Gutwin and Greenberg [54]).

Category	Element	Specific questions
Who	Presence	Is anyone in the workspace?
	Identity	Who is participating? Who is that?
	Authorship	Who is doing that?
What	Action	What are they doing?
	Intention	What goal is that action part of?
	Artifact	What object are they working on?
Where	Location	Where are they working?
	Gaze	Were are they looking?
	View	Where can they see?
	Reach	Where can they reach?

Table A.2: Elements of workspace awareness relating to the past (Gutwin and Greenberg [54]).

Category	Element	Specific questions
How	Action history	How did that operation happen?
	Identity history	How did this artifact come to be in this state?
When	Event history	When did that event happen?
Who (past)	Presence history	Who was here, and when?
Where (past)	Location history	Where has a person been?
What (past)	Action history	What has a person been doing?

A.2 Exploratory Work

In this dissertation, we also conducted work focused on topics verging the thesis critical path. The respective research contributed with valuable knowledge and expertise, both with tracking people's body movements and working in a virtual workspace. Below we present two exploratory works related to this dissertation , as well as the scientific publications they originated.

A.2.1 *Augmented Reality for Rehabilitation using Realtime Feedback*

In this exploratory work, we presented an intelligent user interface that allows people to perform rehabilitation exercises by themselves under the offline supervision of a therapist. For this, we proposed SleeveAR, a novel approach that enhances patient awareness to guide them during rehabilitation exercises. SleeveAR aims at providing the means for patients to precisely replicate these exercises, especially prescribed for them by a knowledgeable health professional, as demonstrated in Video Figure 4. Since the rehabilitation process (Figure A.1) relies on repetition of exercises during the physiotherapy sessions, our approach contributes to the correct performance of the therapeutic exercises while offering reports on the patient's progress. Furthermore, without rendering the role of the therapist obsolete, our approach builds on the notion that with proper guidance, patients can autonomously execute rehabilitation exercises.

With SleeveAR, patients are able to formally assess feedback combinations suitable for movement guidance while solving some of the perception problems. SleeveAR also applies projection-based visual feedback both on the user's body (arm and forearm) and on his surrounding floor. The ground projection shows the movements in

Video Figure 4. SleeveAR overview.
`http://web.ist.utl.pt/~antonio.sousa/videos/sousa2016-acmiui-sleevear.mp4`
`(File size: 15.8 MB)`

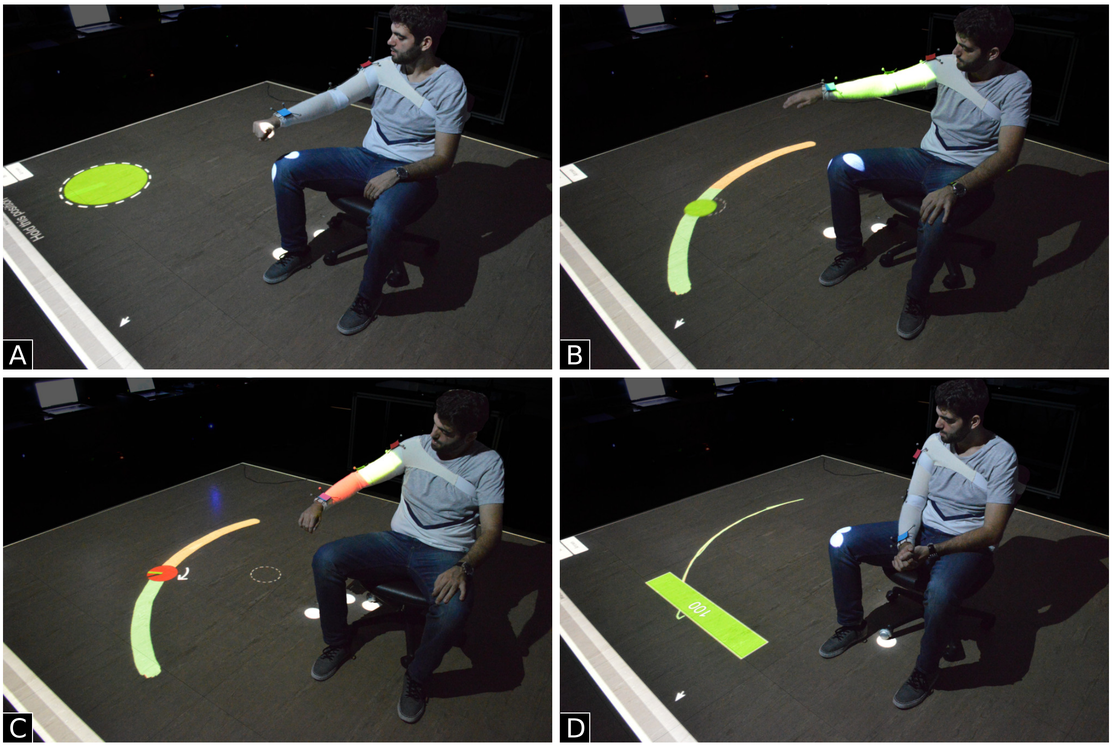


Figure A.1: SleeveAR addresses new active projection-based strategies for providing user feedback during rehabilitation exercises. A) Initial position. B) Mid-performance. C) Sleeve Feedback. D) Progress report.

all axes and allows the sleeve projection to continue in the patient's peripheral field of view.

Empirical evaluation showed the effectiveness of our approach as compared to traditional video-based feedback. Our experimental results showed that SleeveAR can successfully guide subjects through an exercise prescribed (and demonstrated) by a physical therapist, with performance improvements between consecutive executions, a desirable goal to successful rehabilitation.

Corresponding Publication

Part of the contents of this section previously appeared in the following publication:

[P1] Maurício Sousa, João Vieira, Daniel Medeiros, Artur Arsenio, and Joaquim Jorge. **SleeveAR: Augmented Reality for Rehabilitation using Realtime Feedback.** In Proceedings of the 21st International Conference on Intelligent

User Interfaces (IUI '16). ACM, New York, NY, USA, 175-185. DOI: https://doi.org/10.1145/2856767.2856773

A.2.2 Virtual Reality for Radiologists in the Reading Room

We first tackled interaction with 3D virtual data in a real work healthcare scenario. The main focus of this iteration is on how to interact with a workspace designed to house 3D virtual content. Therefore, we purposely decided to circumvent remote collaboration at this stage and contribute first contribute a novel, cost-effective, and portable method to study medical images. Reading room conditions such as illumination, ambient light, human factors, and display luminance, play an essential role in how radiologists analyze and interpret images. Indeed, severe diagnostic errors can appear when observing images through everyday monitors. Typically, these occur whenever professionals are ill-positioned with respect to the display or visualize images under improper light and luminance conditions. In this work, we show that virtual reality can assist radiodiagnostics by considerably diminishing or cancel out the effects of unsuitable ambient conditions. Our approach combines immersive head-mounted displays with interactive surfaces to support professional radiologists in analyzing medical images and formulating diagnostics, as depicted in Figure A.2. We evaluated our prototype with two senior medical doctors and four seasoned radiology fellows. Our evaluation with experts suggests that Virtual Reality is a viable approach to overcome existing ergonomic, ambient, and illumination conditions. While interacting with the desk surface helps to promote its adoption by medical professionals. Additionally, participants were able to identify organs, and all participants correctly identified fractures and the prosthesis, even though none were prompted to do so.

Video Figure 5. VRRRRoom Overview.
http://web.ist.utl.pt/~antonio.sousa/videos/sousa2017-acmchi-video-figure.mp4
(File size: 20.6 MB)

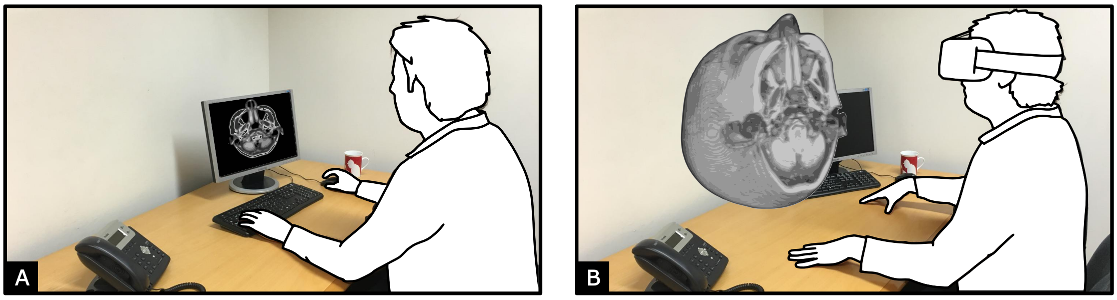


Figure A.2: A) A typical radiology reading room; B) Our VRRRRoom approach combining virtual reality and desktop touch interactions.

VRRRRoom disregards self-representation because we relied on touch input, providing with somesthesis [122] feedback about the movement and position of his hands. Yet, when in remote collaboration, being able to observe other people's full-body representation is essential to convey deictics and other nonverbal cues.

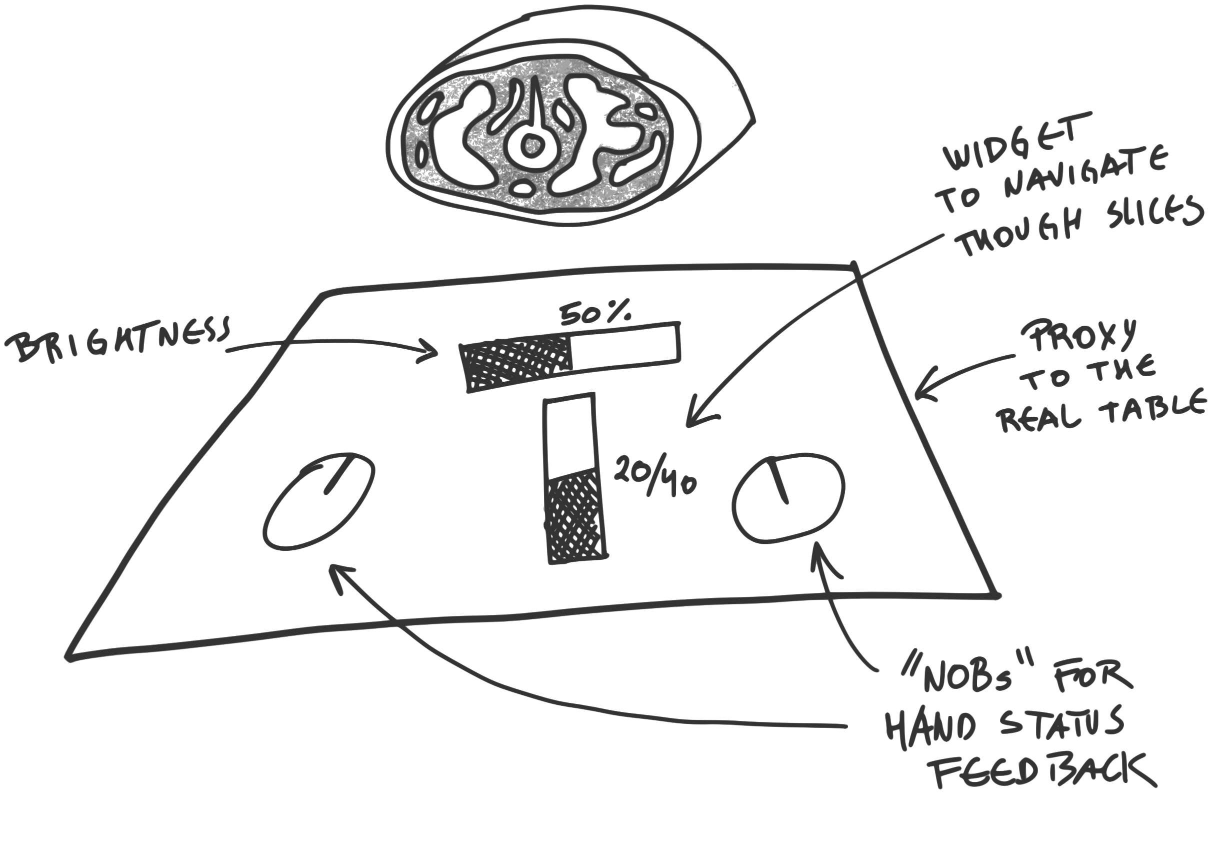
Corresponding Publication

Part of the contents of this section previously appeared in the following publication:

[P2] Maurício Sousa, Daniel Mendes, Soraia Paulo, Nuno Matela, Joaquim Jorge, and Daniel Simões Lopes. **VRRRRoom: Virtual Reality for Radiologists in the Reading Room.** In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 4057-4062. DOI: https://doi.org/10.1145/3025453.3025566

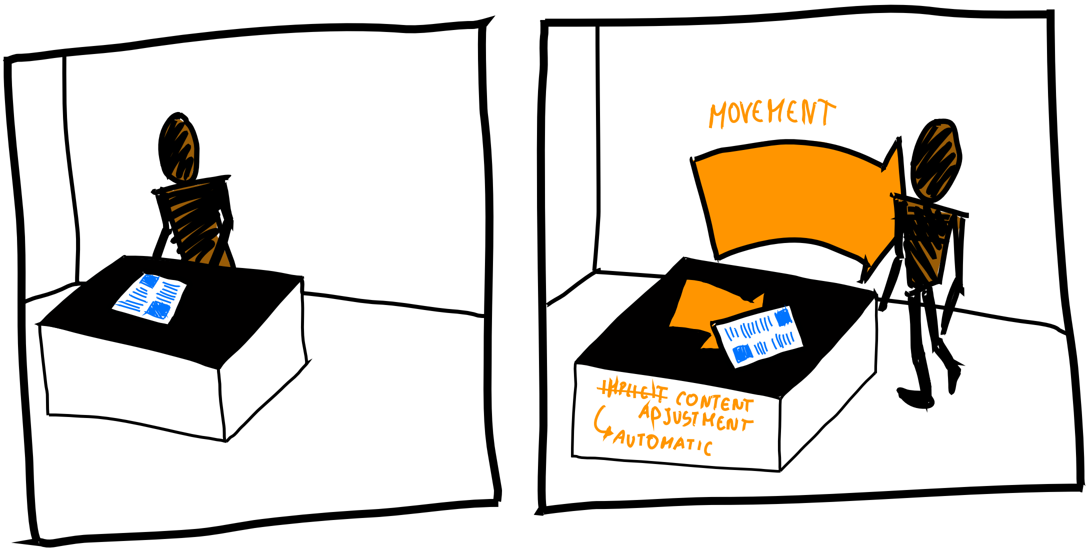
SKETCHBOOK

VRRRROOM

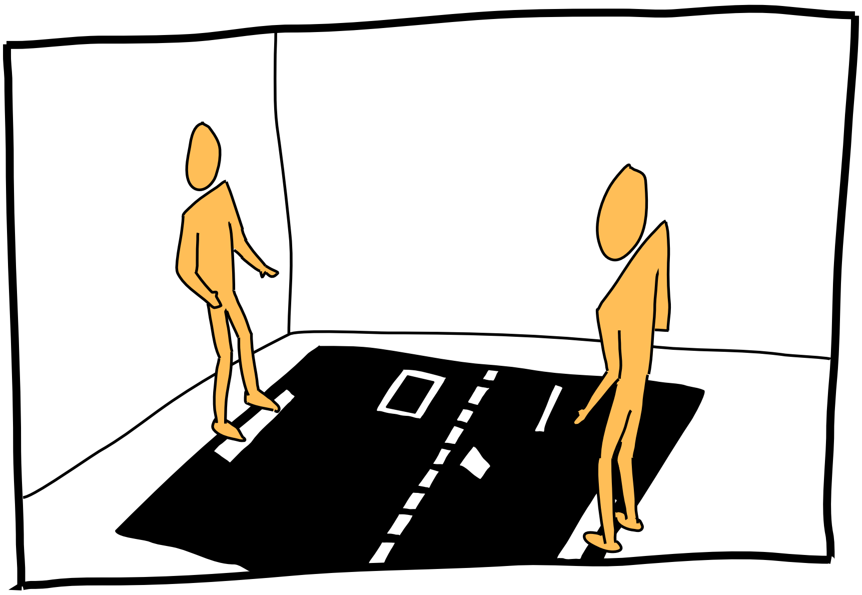

WIDGET
TO NAVIGATE
THOUGH SLICES
BRIGHTNESS
50%
PROXY
TO THE
REAL TABLE
20/40
"NOBs" FOR
HAND STATUS
FEEDBACK

CREEPY TRACKER

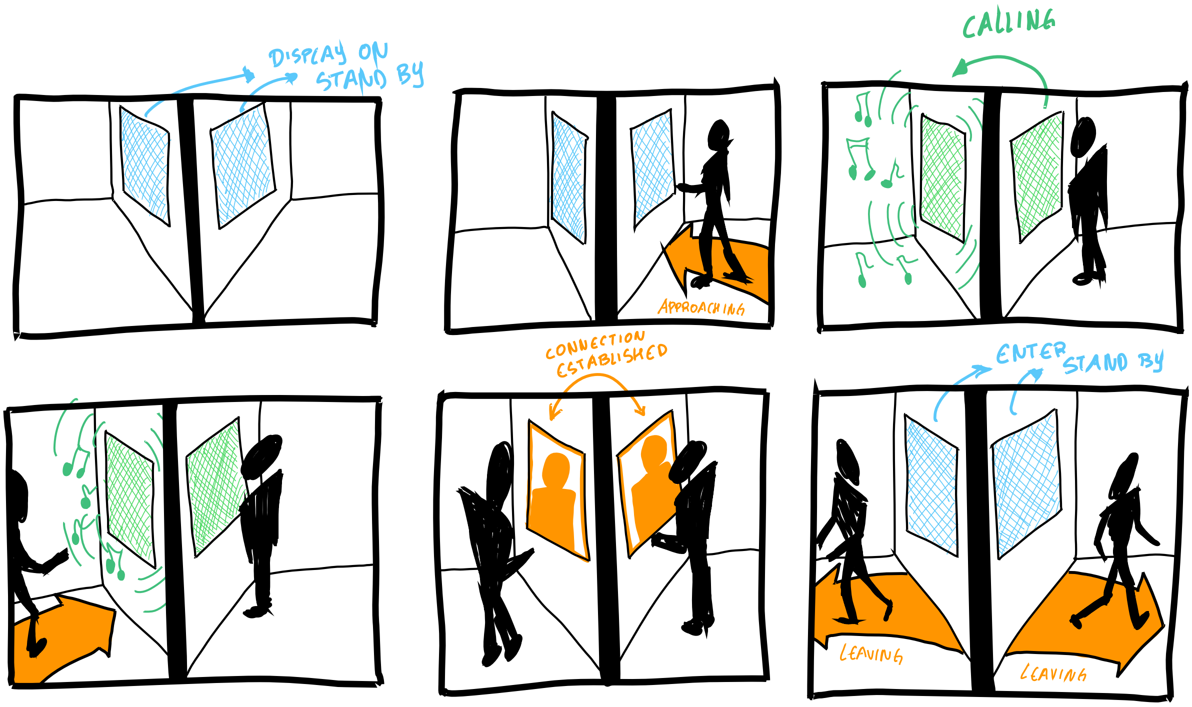
CONTEXT-AWARE INTERACTIVE SURFACES



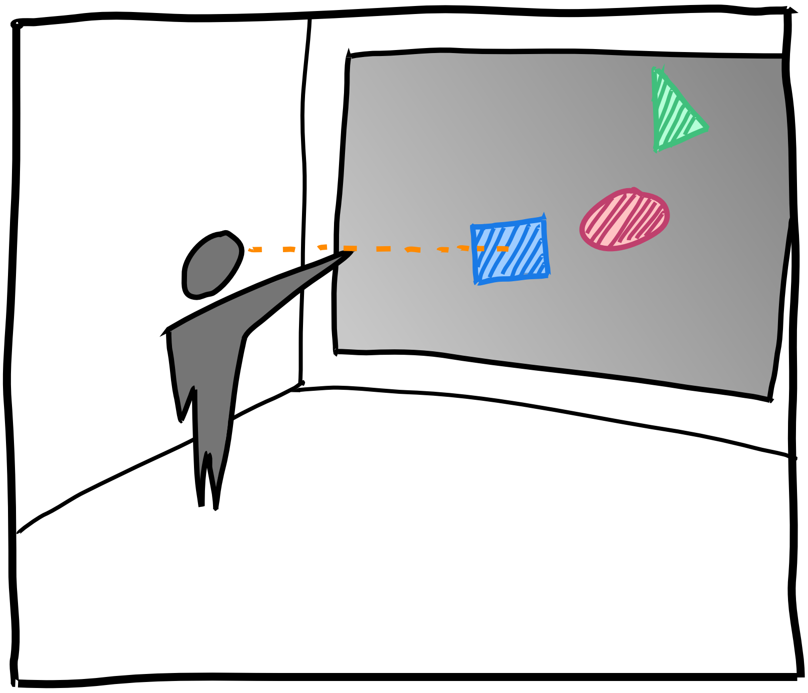
FLOOR SURFACE ENTERTAINMENT SCENARIO



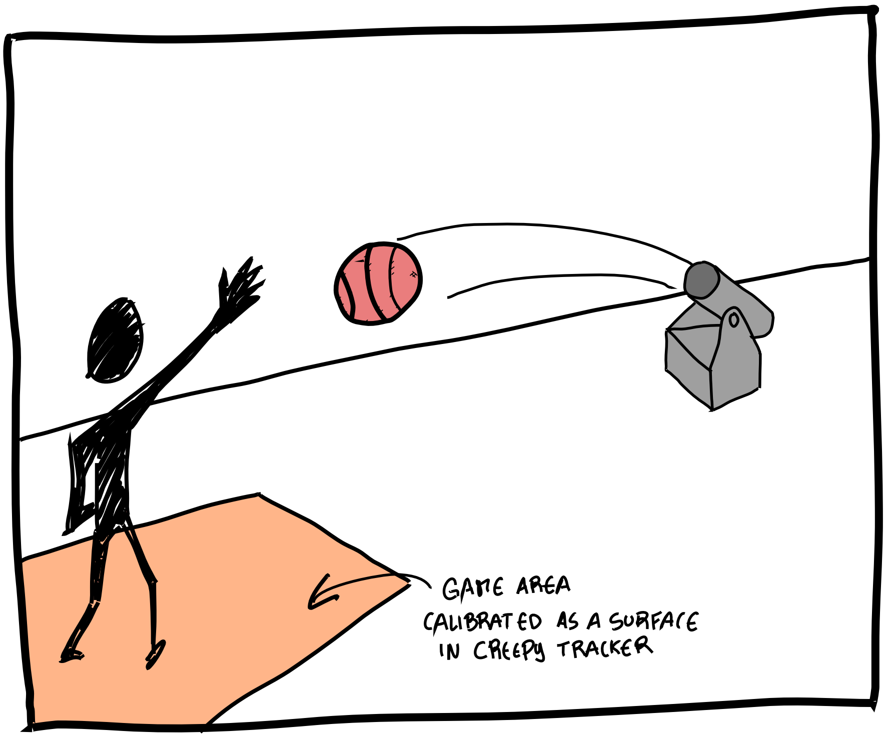
CONTEXT-AWARE TELEPRESENCE



WALL-SIZED DISPLAY :
"PUT-THAT-THERE" SCENARIO

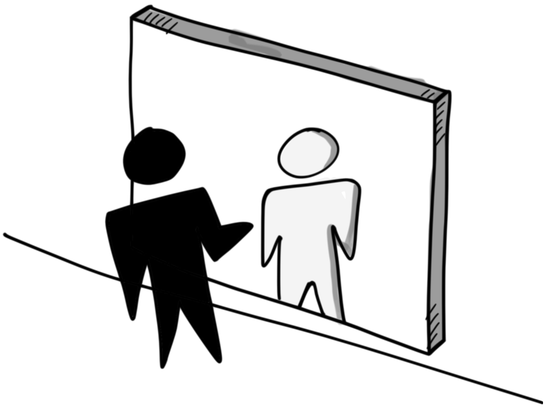


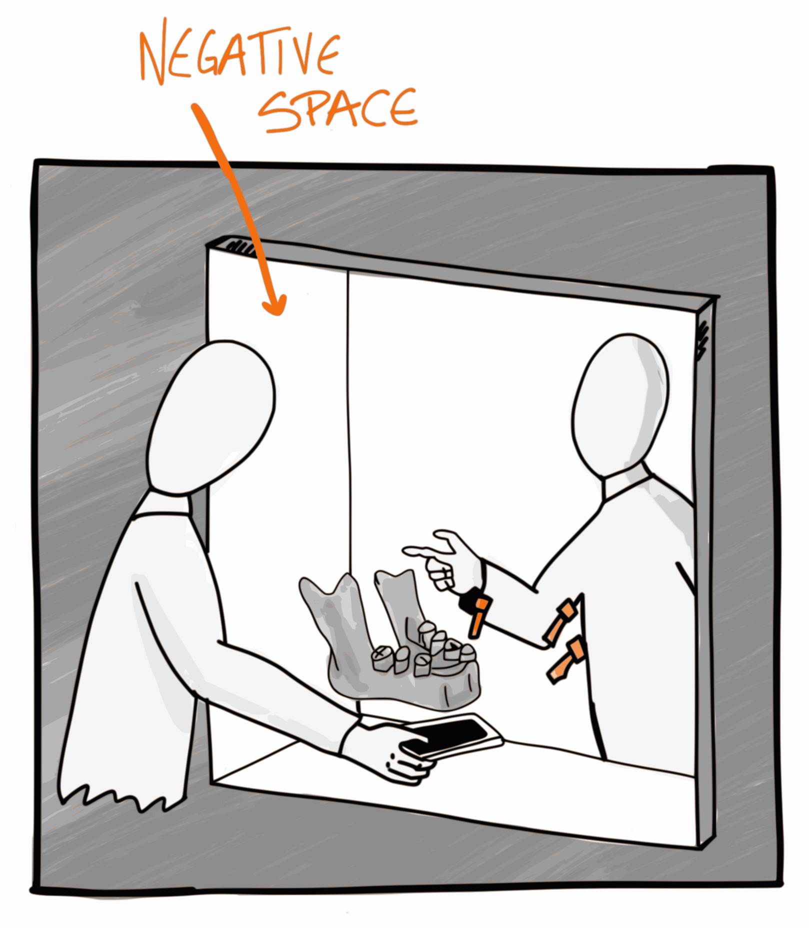
VIRTUAL REALITY
POINT CLOUD AVATAR

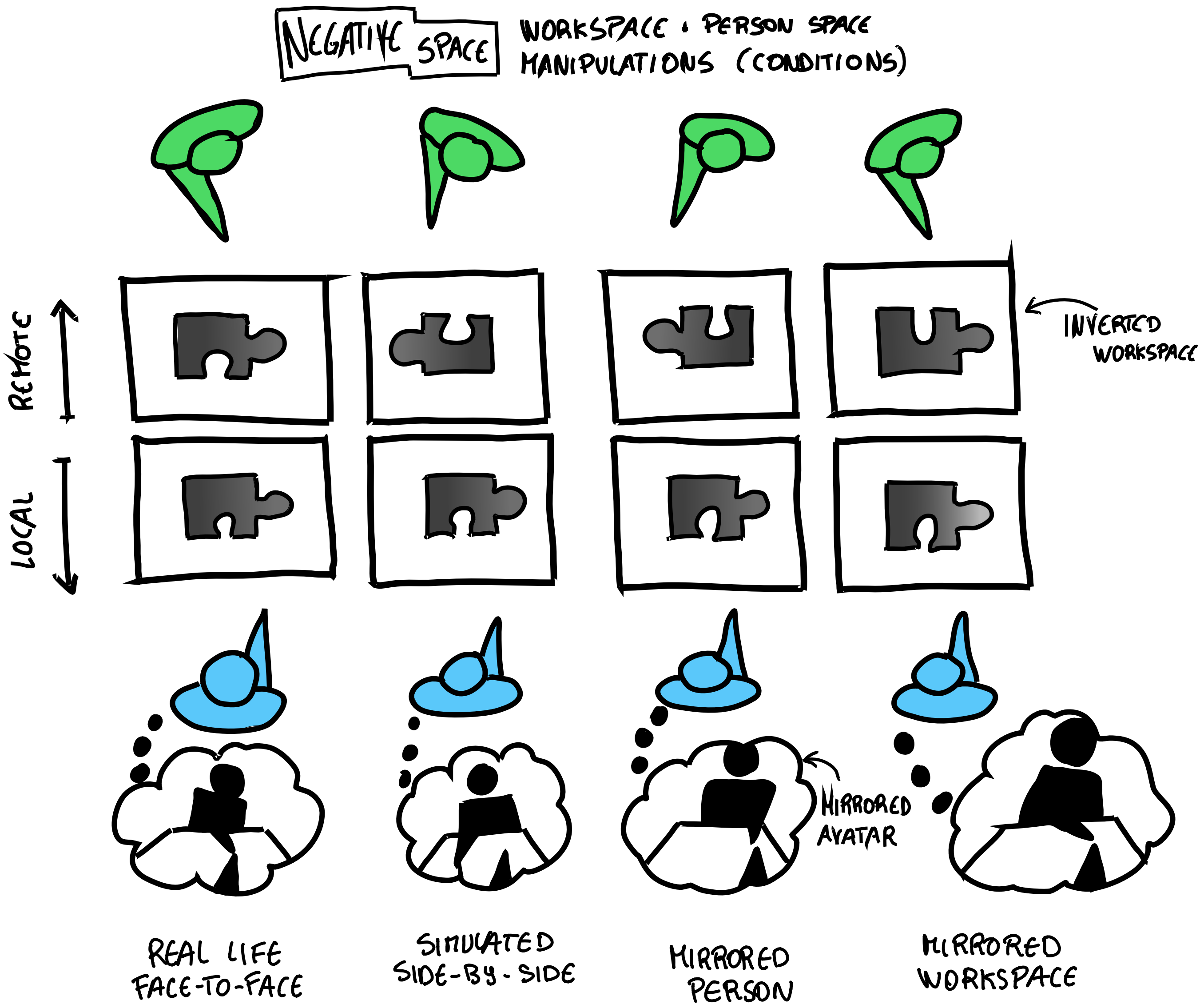


GAME AREA
CALIBRATED AS A SURFACE
IN CREEPY TRACKER

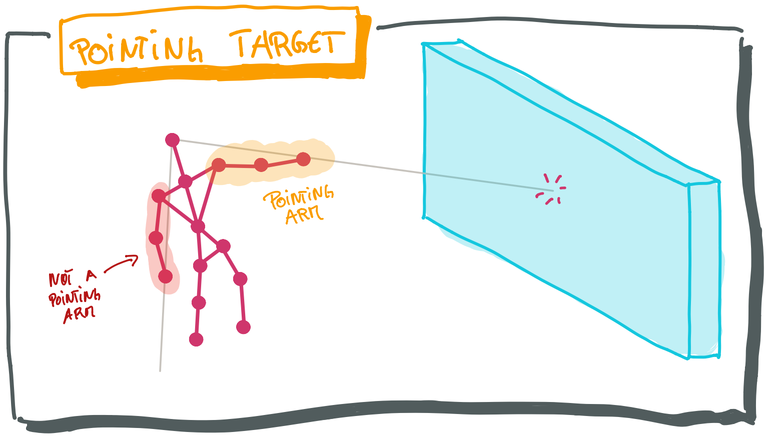
NEGATIVE SPACE

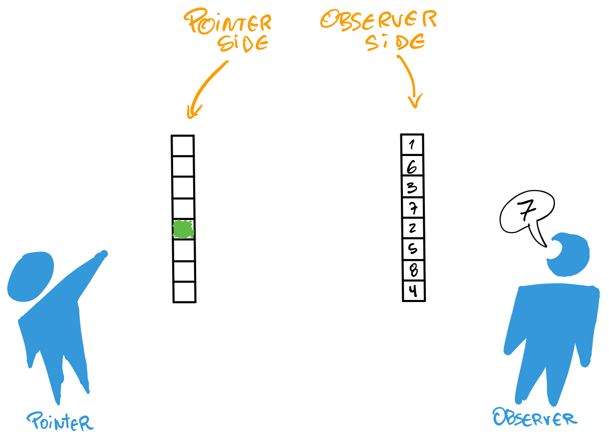


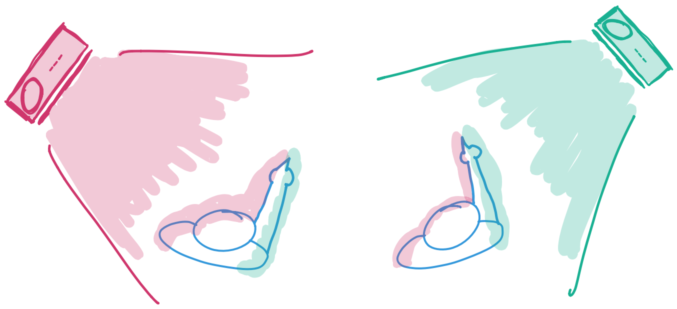



NEGATIVE SPACE
WORKSPACE & PERSON SPACE
MANIPULATIONS (CONDITIONS)
REMOTE
LOCAL
INVERTED WORKSPACE
MIRRORED AVATAR
REAL LIFE FACE-TO-FACE
SIMULATED SIDE-BY-SIDE
MIRRORED PERSON
MIRRORED WORKSPACE

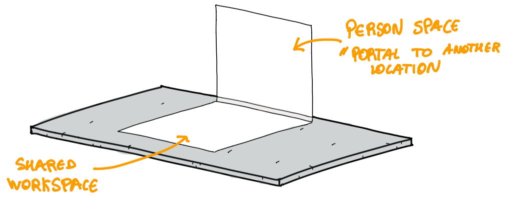
WARPING DEIXIS

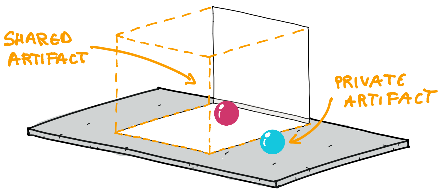


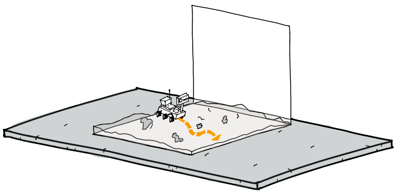


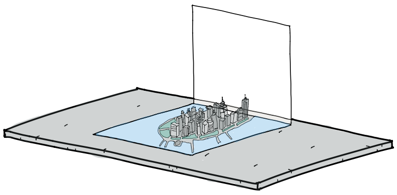


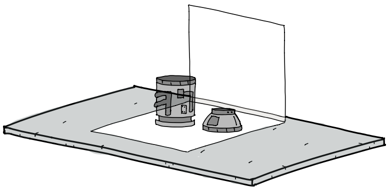
ALTERED PRESENCE ⚡ RESHAPING GESTURES

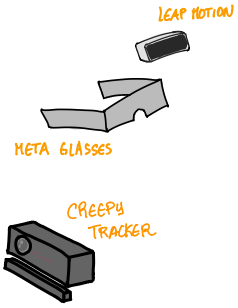


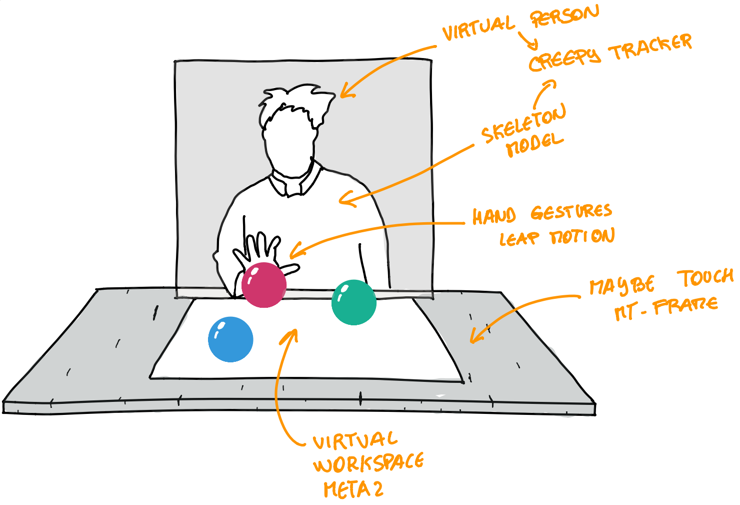


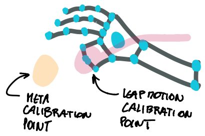


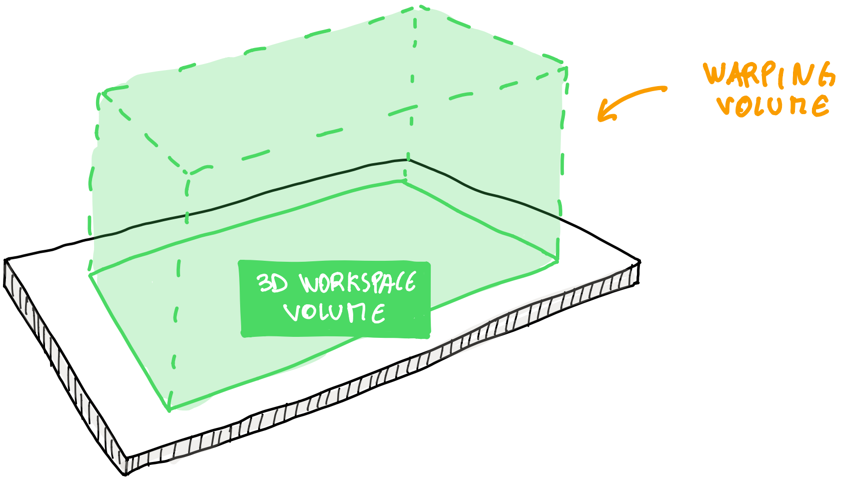


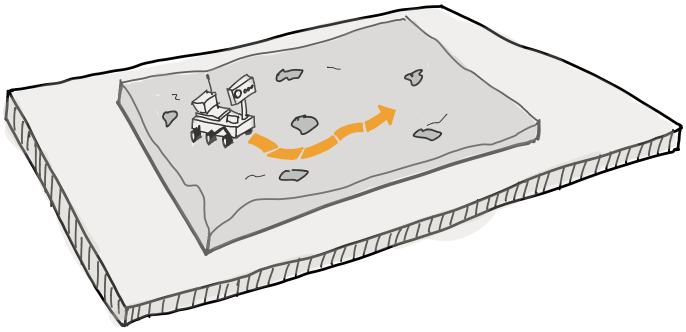


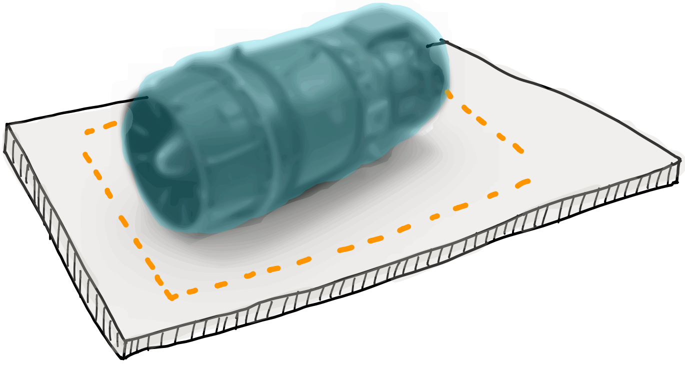


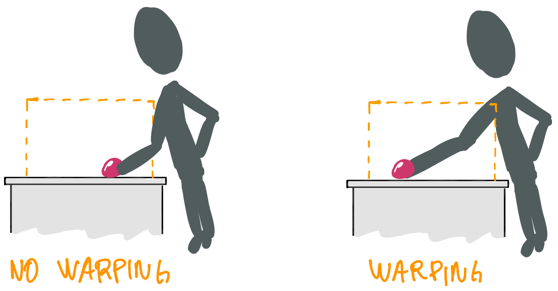


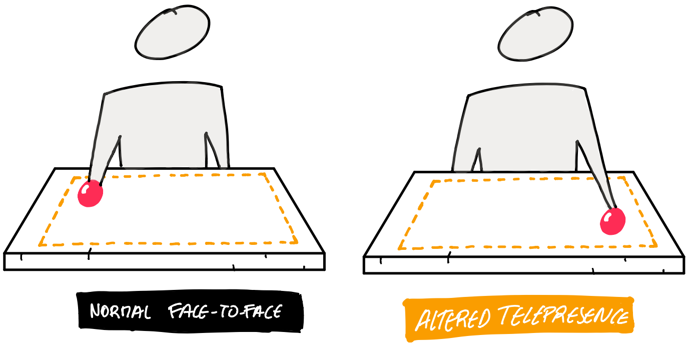


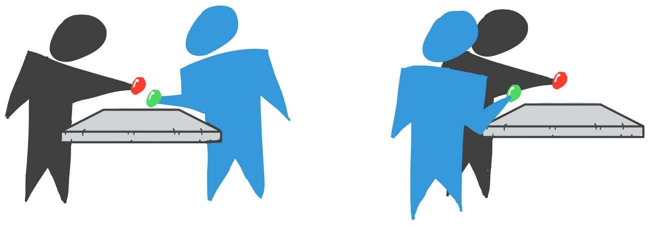

WARPING VOLUME
3D WORKSPACE VOLUME

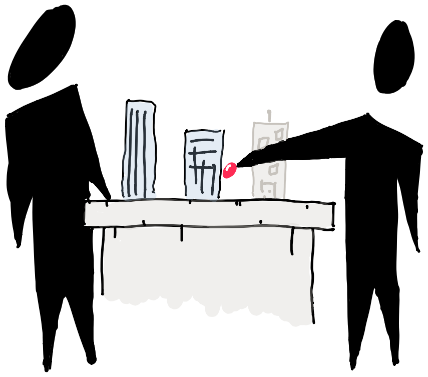


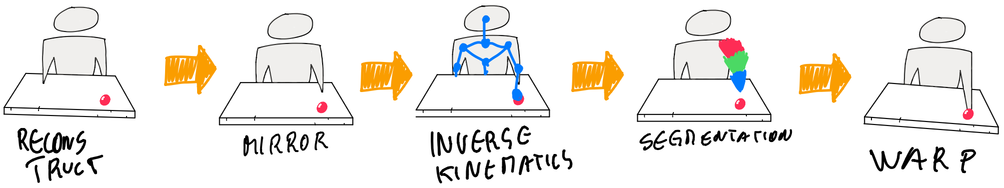


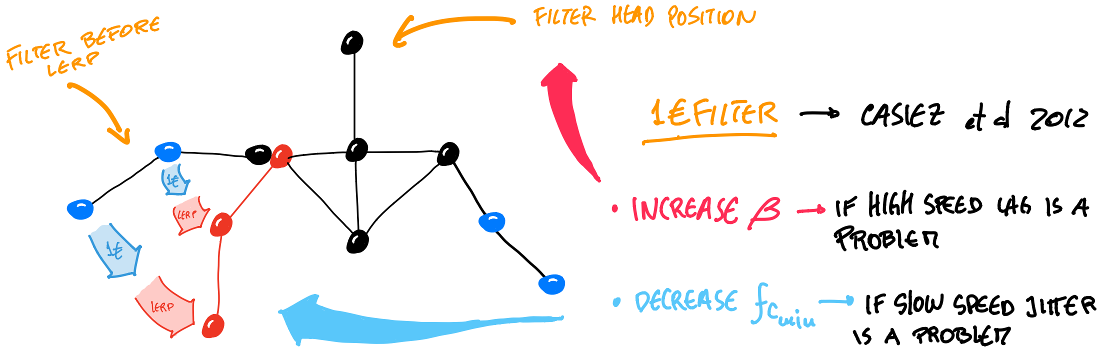

NO WARPING
WARPING

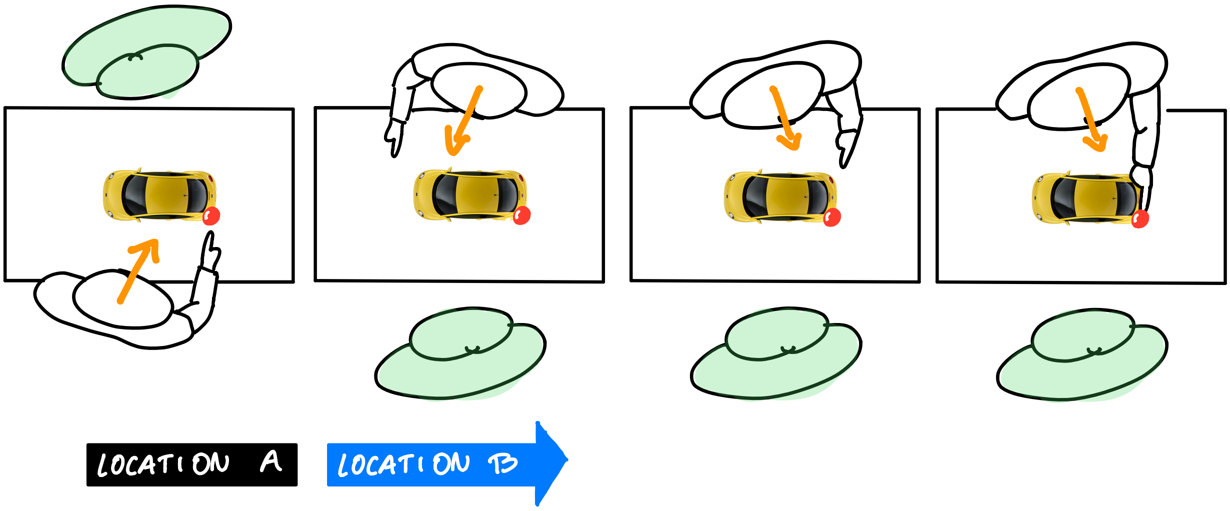

NORMAL FACE-TO-FACE
ALTERED TELEPRESENCE

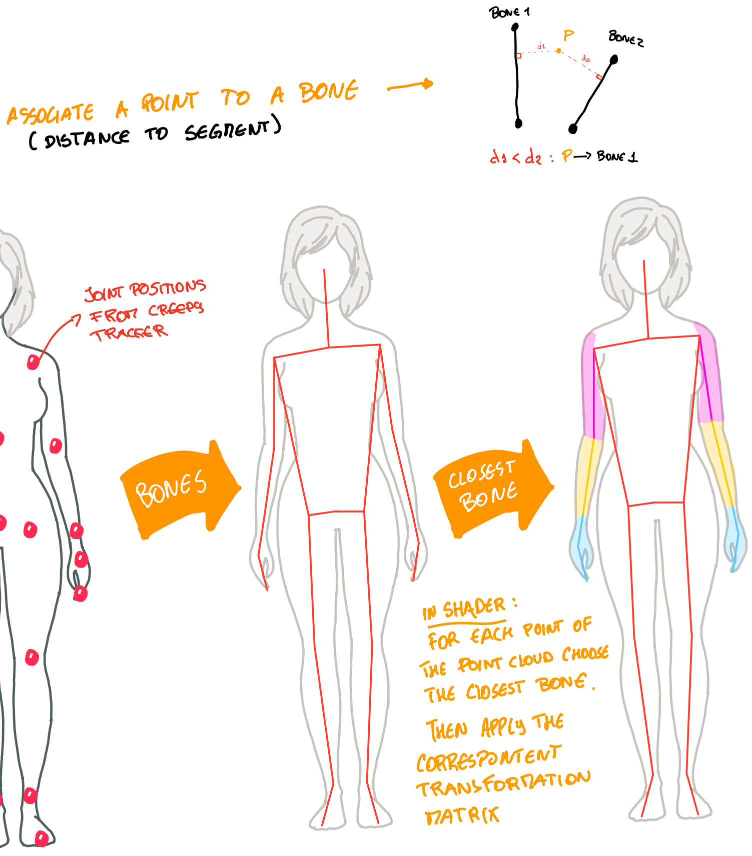


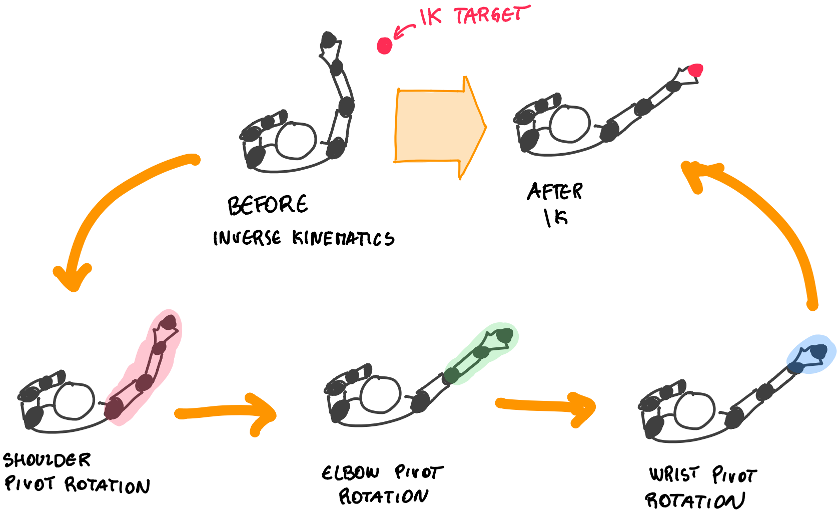


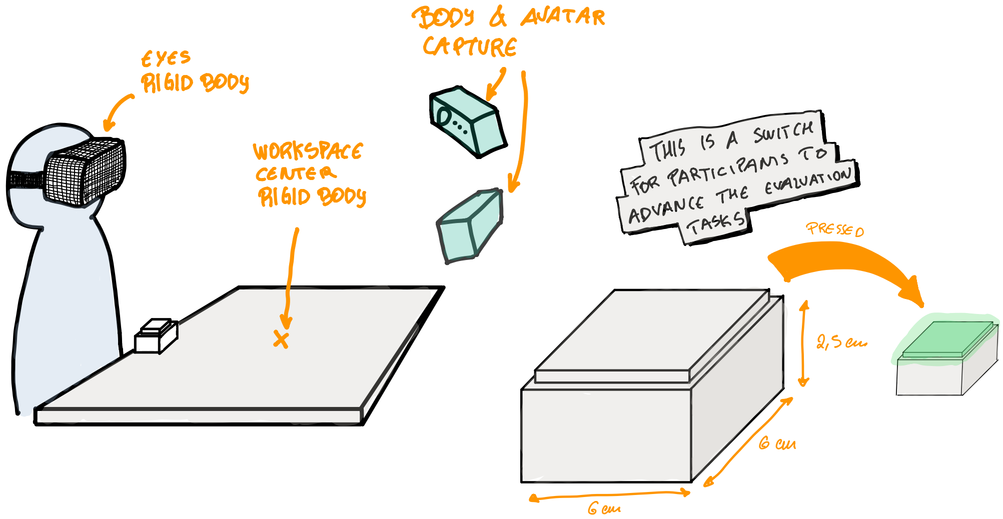

RECONS TRUCT
MIRROR
INVERSE KINEMATICS
SEGMENTATION
WARP

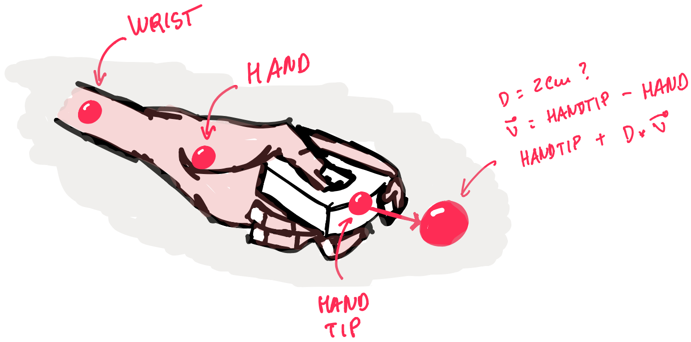

FILTER BEFORE LERP
FILTER HEAD POSITION
1€
LERP
1€
LERP
1€FILTER → CASIEZ et al 2012
• INCREASE β → IF HIGH SPEED LAG IS A PROBLEM
• DECREASE fc_min → IF SLOW SPEED JITTER IS A PROBLEM

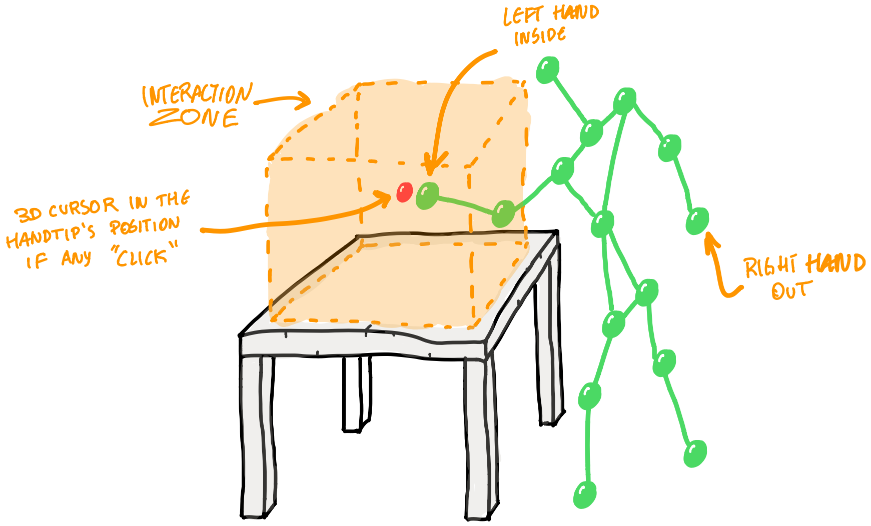

LOCATION A
LOCATION B


ASSOCIATE A POINT TO A BONE
(DISTANCE TO SEGMENT)
BONE 1
P
d1
d2
BONE 2
d1 < d2 : P → BONE 1
JOINT POSITIONS FROM CREEPY TRACKER
BONES
CLOSEST BONE
IN SHADER :
FOR EACH POINT OF THE POINT CLOUD CHOOSE THE CLOSEST BONE.
THEN APPLY THE CORRESPONTENT TRANSFORMATION MATRIX


IK TARGET
BEFORE
INVERSE KINEMATICS
AFTER
IK
SHOULDER
PIVOT ROTATION
ELBOW PIVOT
ROTATION
WRIST PIVOT
ROTATION


EYES RIGID BODY
BODY & AVATAR CAPTURE
WORKSPACE CENTER RIGID BODY
THIS IS A SWITCH FOR PARTICIPANTS TO ADVANCE THE EVALUATION TASKS
PRESSED
2,5 cm
6 cm
6 cm


WRIST
HAND
D = 2cm ?
$\vec{v}$ = HANDTIP - HAND
HANDTIP + D$\cdot\vec{v}$
HAND TIP


LEFT HAND INSIDE
INTERACTION ZONE
3D CURSOR IN THE HANDTIP'S POSITION IF ANY "CLICK"
RIGHT HAND OUT

mdias

THIS DISSERTATION WAS TYPESET using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 12 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above figure is a point-cloud representation of me, captured using the Creepy Tracker Tookit, in the João Lourenço Fernandes Laboratory at Instituto Superior Técnico, where most of the prototyping effort was done and user evaluations were conducted. This dissertation was written in Overleaf using an adaptation of a template that can be used to format a dissertation and is under the permissive AGPL license and can be found online at github.com/suchow/Dissertate. All figures were digitally designed by me using a combination of an iPad Pro, an Apple Pencil, the illustrator app Procreate, and the vector graphics editors Affinity Designer for iPad and Inkscape.