

**IEEE** Computer Graphics  
AND APPLICATIONS

VOLUME 38, NUMBER 1

JANUARY/FEBRUARY 2018

# Image-Based Modeling & Rendering



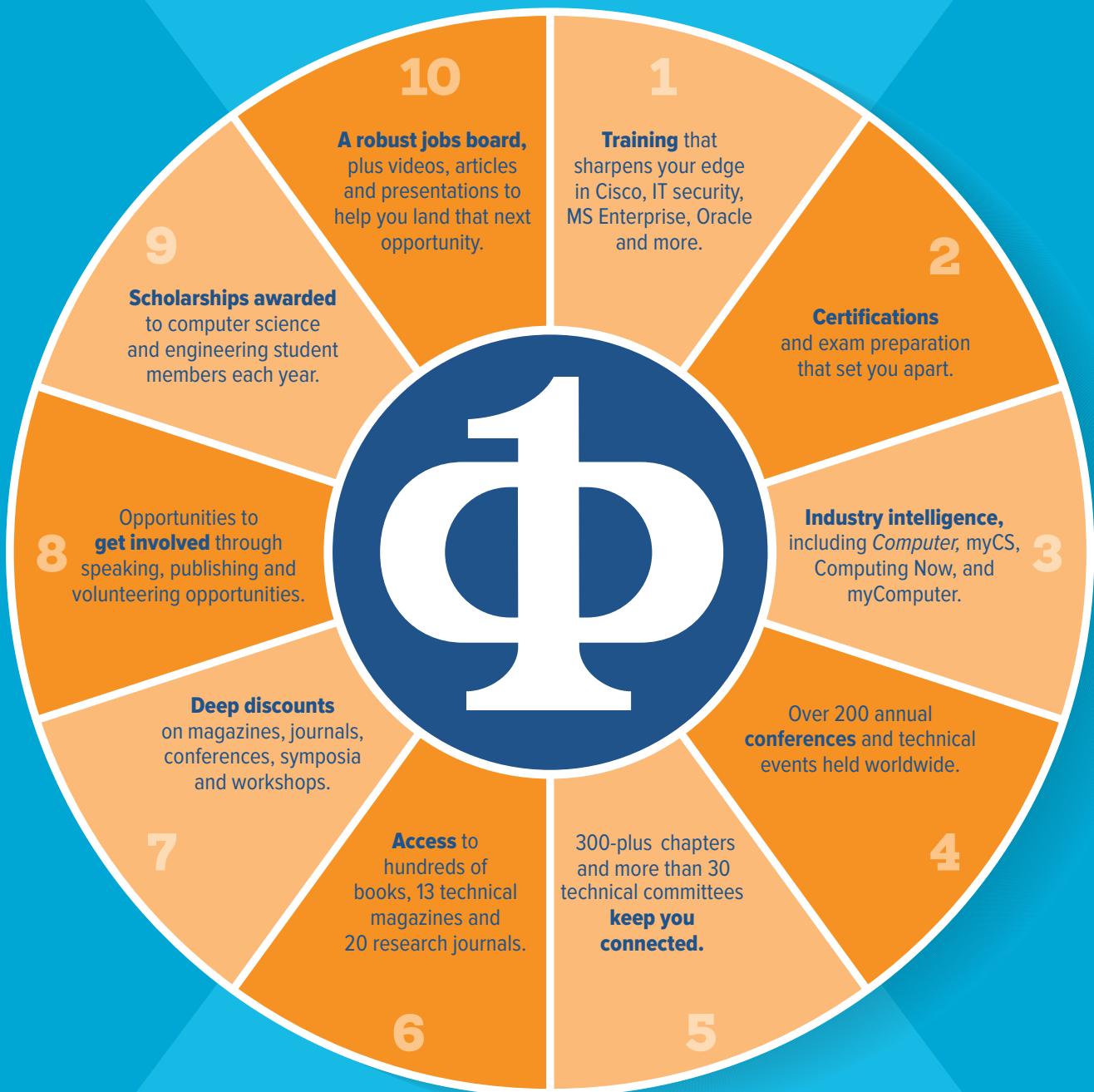
IEEE  
computer  
society

[www.computer.org/cga](http://www.computer.org/cga)

# IEEE COMPUTER SOCIETY: Be at the Center of It All

IEEE Computer Society membership puts you at the heart of the technology profession—and helps you grow with it.

**Here are 10 reasons why you need to belong.**



IEEE Computer Society—keeping you ahead of the game. Get involved today.

[www.computer.org/membership](http://www.computer.org/membership)

IEEE  
computer  
SOCIETY

**EDITOR IN CHIEF**

**Torsten Möller**, University of Vienna  
torsten.moeller@univie.ac.at

**ASSOCIATE EDITORS IN CHIEF**

**Kwan-Liu Ma**, University of California, Davis  
**Michael J. Potel**, Wildcrest Associates  
**Pak Chung Wong**, ACT, Inc.

**EDITORIAL BOARD**

**Ginger Alford**, Trinity Valley School, Fort Worth  
Museum of Science and History  
**Jan Bender**, RWTH Aachen University  
**Bruce D. Campbell**, Rhode Island School of Design  
**Yung-Yu Chuang**, National Taiwan University  
**Martin Hatchet**, Inria Bordeaux  
**Seok-Hee Hong**, University of Sydney  
**Jörn Kohlhammer**, Fraunhofer IGD  
**Joseph J. LaViola Jr.**, University of Central Florida  
**Seungyong Lee Pohang**, University of Science and Technology  
**Ligang Liu**, University of Science and Technology of China  
**Aditi Majumder**, University of California, Irvine  
**Michael Neff**, University of California, Davis  
**Manuel M. Oliveira**, Federal University of Rio Grande do Sul  
**Theresa-Marie Rhyne**, Consultant  
**Francesca Samsel**, University of Texas at Austin  
**Christian Sandor**, Nara Institute of Science and Technology  
**Gerik Scheuermann**, University of Leipzig  
**Beatriz Sousa Santos**, University of Aveiro  
**Frank Steinicke**, University of Hamburg  
**André Stork**, Technische Universität Darmstadt  
**Wolfgang Stuerzlinger**, Simon Fraser University  
**Melanie Tory**, Tableau Software  
**Wenping Wang**, University of Hong Kong  
**Peter Wonka**, KAUST

**ADVISORY COUNCIL**

**John Dill Simon**, Fraser University  
**Rae A. Earnshaw**, University of Bradford  
**L. Miguel Encarnação**, Innovation By Design Int'l Consulting  
**José Luis Encarnação**, Technische Universität Darmstadt  
**Jim Foley**, Georgia Institute of Technology  
**Markus Gross**, Disney Research, Zürich

**Dave Kasik**, Boeing

**Carol O'Sullivan**, Trinity College, Dublin  
**Lawrence J. Rosenblum**, US National Science Foundation  
**Holly Rushmeier**, Yale University  
**Maureen Stone Tableau**, Software  
**Gabriel Taubin**, Brown University

**EDITORIAL STAFF**

**Staff Editor/Magazine Contact:** Brian Brannon, bbrannon@computer.org  
**Contributing Editor:** Gary Singh  
**Senior Advertising Coordinator:** Debbie Sims  
**Manager, Editorial Services:** Brian Brannon  
**Publisher:** Robin Baldwin  
**Director, Products & Services:** Evan Butterfield  
**Director of Membership:** Eric Berkowitz

**CS MAGAZINE OPERATIONS COMMITTEE**

**George K. Thiruvathukal (Chair)**, Gul Agha, M. Brian Blake, Irena Bojanova, Jim X. Chen, Shu-Ching Chen, Lieven Eeckhout, Nathan Ensmenger, Sumi Helal, Marc Langheinrich, Torsten Möller, David Nicol, Diomidis Spinellis, VS Subrahmanian, Mazin Yousif

**CS PUBLICATIONS BOARD**

**Greg Byrd (VP for Publications)**, Erik Altman, Ayse Basar Bener, Alfredo Benso, Robert Dupuis, David S. Ebert, Davide Falessi, Vladimir Getov, Avi Mendelson, Dimitrios Serpanos, Forrest Shull, George K. Thiruvathukal

**EDITORIAL OFFICE**

**Publications Coordinator:** cga-ma@computer.org  
**Authors:** www.computer.org/cga/author.htm  
**Letters to the Editors:** bbrannon@computer.org  
**Subscribe:** www.computer.org/subscribe  
**Subscription change of address:** address.change@ieee.org  
**Missing or damaged copies:** help@computer.org  
**Reprints of articles:** cga-ma@computer.org  
IEEE Computer Graphics and Applications  
c/o IEEE Computer Society  
10662 Los Vaqueros Circle, Los Alamitos, CA 90720 USA  
Phone +1 714 821 8380; Fax +1 714 821 4010  
www.computer.org/cga



**Reuse Rights and Reprint Permissions:** Educational or personal use of this material is permitted without fee, provided such use: 1) is not made for profit; 2) includes this notice and a full citation to the original work on the first page of the copy; and 3) does not imply IEEE endorsement of any third-party products or services. Authors and their companies are permitted to post the accepted version of IEEE-copyrighted material on their own web servers without permission, provided that the IEEE copyright notice and a full citation to the original work appear on the first screen of the posted copy. An accepted manuscript is a version which has been revised by the author to incorporate review suggestions, but not the published version with copyediting, proofreading, and formatting added by IEEE. For more information, please go to: [www.ieee.org/publications\\_standards/publications/rights/paperversionpolicy.html](http://www.ieee.org/publications_standards/publications/rights/paperversionpolicy.html). Permission to reprint/republish this material for commercial, advertising, or promotional purposes or for creating new collective works for resale or redistribution must be obtained from IEEE by writing to the IEEE Intellectual Property Rights Office, 445 Hoes Lane, Piscataway, NJ 08854-4141 or pubpermissions@ieee.org. Copyright © 2018 IEEE. All rights reserved. **Abstracting and Library Use:** Abstracting is permitted with credit to the source. Libraries are permitted to photocopy for private use of patrons, provided the per-copy fee indicated in the code at the bottom of the first page is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

**Circulation:** *IEEE Computer Graphics and Applications* (ISSN 0272-1716) is published bimonthly by the IEEE Computer Society. IEEE Headquarters, Three Park Avenue, 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Circle, PO Box 3014, Los Alamitos, CA 90720-1314; voice +1 714 821 8380; fax +1 714 821 4010; IEEE Computer Society Headquarters, 1828 L St. NW, Suite 1202, Washington, DC 20036. Visit [www.computer.org/subscribe](http://www.computer.org/subscribe) for subscription information.

**Postmaster:** Send undelivered copies and address changes to *IEEE Computer Graphics and Applications*, Membership Processing Dept., IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854-4141. Periodicals Postage Paid at New York, NY, and at additional mailing offices. Canadian GST #125634188. Canada Post Publications Mail Agreement Number 40013885. Return undeliverable Canadian addresses to PO Box 122, Niagara Falls, ON L2E 6S8, Canada. Printed in the USA. **Editorial:** Unless otherwise stated, bylined articles, as well as product and service descriptions, reflect the author's or firm's opinion. Inclusion in *IEEE Computer Graphics and Applications* does not necessarily constitute endorsement by the IEEE or the Computer Society. All submissions are subject to editing for style, clarity, and space. IEEE prohibits discrimination, harassment, and bullying: For more information, visit [www.ieee.org/web/aboutus/whatis/policies/p9-26.html](http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html).

---

## TABLE OF CONTENTS

### Image-Based Modeling & Rendering

- 30 Crosswalk Localization from Low Resolution Satellite Images to Assist Visually Impaired People  
Marcelo Cabral Ghilardi, Julio Jacques Junior, and Isabel Manssour
- 47 Objective Quality Assessment and Perceptual Compression of Screen Content Images  
Shiqi Wang, Ke Gu, Kai Zeng, Zhou Wang, and Weisi Lin
- 59 Efficient C2-Weighting for Image Warping  
Chuhua Xian, Shuo Jin, and Charlie C.L. Wang
- 77 Parametric Reshaping of Portrait Images for Weight-change  
Haiming Zhao, Xiaogang Jin, Xiaojian Huang, Menglei Chai, and Kun Zhou

- 91 ARIES: Enabling Visual Exploration and Organization of Art Image Collections

Lhaylla Crissaff, Louisa Wood Ruby, Samantha Deutch, R. Luke DuBois, Jean-Daniel Fekete, Juliana Freire, and Claudio Silva

### Columns and Departments

- 4 FROM THE EDITOR IN CHIEF  
Building Bridges between Research and Applications  
Torsten Möller
- 8 ABOUT THE COVER  
Improving Visual Communication  
Gary Signh
- 11 SPATIAL INTERFACES  
A Taxonomy of Spatial Interaction Patterns and Techniques  
Jason Jerald

**20 APPLICATIONS**

**ColorMoves: Real-time Interactive Colormap Construction for Scientific Visualization**

Francesca Samsel, Sebastian Klaassen, and David H. Rogers

**109 GRAPHICALLY SPEAKING**

**10 Years of MegaMol: The Pain and Gain of Creating Your Own Visualization Framework**

Michael Krone, Sebastian Grottel, Guido Reina, Christoph Müller, and Thomas Ertl

**115 TOOLS AND PRODUCTS**

**Putting VR/AR to Work**

Amit Agrawal

**119 VISUALIZATION VIEWPOINTS**

**Optimizing Scientist Time through In Situ Visualization and Analysis**

John Patchett and James Ahrens

## Also in This Issue

**1 Masthead**

---

For more information on computing topics, visit the Computer Society Digital Library at [www.computer.org/cSDL](http://www.computer.org/cSDL).

# Building Bridges between Research and Applications

**Torsten Möller**  
University of Vienna

Newly appointed Editor in Chief Torsten Möller thanks outgoing EIC Miguel Encarnaçāo, discusses the history of *IEEE Computer Graphics and Applications* and the current issue, and welcomes new members of the magazine's Editorial Board.

*IEEE Computer Graphics and Applications* magazine was launched in 1980, when I was in third grade on the other side of the Iron Curtain. In its first issue, then Editor in Chief Michael J. Wozny talked about the growing importance of computer graphics and applications in industry and science. This magazine "will appeal to the end user as well as to the specialist. Each issue will cover a broad spectrum of research areas and applications." The world has changed dramatically since then, but it is clearer now more than ever that the impact of computer graphics and its applications is broad and wide and is felt in every aspect of our lives today. I am greatly honored to take over as the editor in chief of such a long-lasting, impactful magazine.

While I know that it will be hard to live up to the expectations of a long list of amazing previous editors in chief, I am grateful for a very dedicated and hard-working editorial board as well as the IEEE Computer Society staff that make this magazine possible. I especially would like to thank our previous editor in chief, Miguel Encarnaçāo, who has done a tremendous job in running the magazine smoothly and on time. He prepared to take over as editor in chief in 2013. Co-incidentally, this was the year when I moved from North America to Austria to start a research group on visualization and data analysis at the University of Vienna. Thanks to Miguel and his team, we see an increased interest in the magazine, which shows through a rising impact factor. He pioneered a link to the IEEE Visualization conference as well as the IEEE Virtual Reality conference. Since 2015, authors can present their work previously published in *CG&A* at the IEEE Visualization conference. Based on the success with IEEE VIS, we are extending this model to the IEEE Virtual Reality conference in 2018. Miguel also created a Twitter account (@ieeeca) and a Facebook community ([www.facebook.com/ieeeca](http://www.facebook.com/ieeeca)) for *CG&A* to reach a new readership. Miguel has kindly agreed to continue his service to the magazine by joining its advisory council.

## UNIQUE DEPARTMENTS

Computer graphics and its applications are broad and *CG&A* is one of the only places where interesting research from such areas as computer graphics, augmented and virtual reality, visualization, and human-computer interaction comes together. What is special and unique for our

community, however, are the departments. Some have been running for many years. Mike Potel's first Application entry was in November of 1994 and has been going strong ever since (see also <http://www.wildcrest.com/Potel/Portfolio/AppsDeptArchive.html>)! When Theresa-Marie put together her first Visualization Viewpoints (by Turner Whitted) in 1999, I had just defended my Ph.D. thesis and started a tenure-track position at Simon Fraser University (see [http://theresamarierhyne.com/Theresa-Marie\\_Rhynes\\_Viewpoint/VisViews\\_Dept\\_Listing.html](http://theresamarierhyne.com/Theresa-Marie_Rhynes_Viewpoint/VisViews_Dept_Listing.html)). Others are brand new. Our newest department "People in Practice" will be on evaluations in research and debut later this year. It will be headed by Melanie Tory, a senior scientist at Tableau research. Melanie recently moved from academia to industry and deeply cares about the user in our applications. She has been a long-time board member of *CG&A* and has been the magazine liaison with IEEE VIS. By coincidence, she had also been my first PhD student and we worked together at Simon Fraser University.

Another new department editor is Amit Agrawal, who is heading Tools and Products with his first curated entry in the current issue. Amit has over 20 years of experience in multimedia technologies. Among others, he was the executive director of software engineering at Sony Pictures Imageworks, where he supervised software development for both computer animated and visual effects feature films including *The Polar Express*, *Spider Man 1 and 2*, *Stuart Little*, *Hollow Man*, and the Oscar-winning short film *Chub Chubs*.

In addition, I'm pleased to announce that Christian Sandor joined Frank Steinicke as the co-lead on the department of Spatial Interfaces. He is one of the forces behind the augmented reality revolution and has gained experience in both industry and academia.

Further to these changes, the following departments have been going strong and round out the success of *CG&A*: Art on Graphics by Bruce Campbell and Francesca Samsel; Dissertation Impact by Jim Foley; Education by Ginger Alford and Beatriz Sousa Santos; as well as Graphically Speaking by André Stork. Please contact us with your ideas and contributions for interesting content.

Lastly, I would like to thank, also in the name of the departing editor in chief, the following departing department heads: Wolfgang Stuerzlinger for co-leading the Spatial Interface department; and Associate Editor Victoria L. Interrante and Lisa S. Avila and Mike Bailey for their co-editorship of the Tools and Products department.

As the editor in chief, I intend to continue building bridges between research and applications, bringing novel ideas from a broad range of computer graphics and its applications to our readership. Hence, "each issue will cover a broad spectrum of research areas and applications." I will do my best to engage the research and practitioner communities to be part and contribute to the journal. After all, it is a unique place for our communities.

## IN THIS ISSUE

The January/February 2018 issue features five great papers from the general queue. All of these papers highlight new research in image-based modeling and image analysis. The article by Marcelo Cabral Ghilardi, Julio C.S. Jacques Jr., and Isabel Harb Manssour on "Crosswalk Localization from Low Resolution Satellite Images to Assist Visually Impaired People" puts algorithms and novel data sources into the services of the needy. It uses an SVM classifier to great effect. The work by Shiqi Wang, Ke Gu, Kai Zeng, Zhou Wang, and Weisi Lin on "Objective Quality Assessment and Perceptual Compression of Screen Content Images" focuses on the compression of images for applications like thin-client applications. They adapt general ideas on perceptual quality metrics for compression towards the specifics of images with lots of text and user interface widgets.

Then, there are two papers on warping of images. The work by Chuhua Xian, Shuo Jin, and Charlie C. L. Wang on "Efficient C<sup>2</sup>-Weighting for Image Warping" introduces a new framework for a higher order blending algorithm for image warping. On the other hand, the work by Haiming Zhao, Xiaogang Jin, Xiaojian Huang, Menglei Chai, and Kun Zhou on "Parametric Reshaping of Portrait Images for Weight-Change" focuses on the manipulation of faces in headshots, helping to thin or thicken the overall shape of the face.

Lastly, “ARIES: Enabling Visual Exploration and Organization of Art Image Collections” by Lhaylla Crissaff, Louisa Ruby, Samantha Deutch, Luke DuBois, Jean-Daniel Fekete, Juliana Freire and Cláudio T. Silva provides a very specific application of image-based modeling to exhibit artifacts from art history. They provide what was once accomplished with so-called light boxes in the physical world in the digital, online world.

We hope that you enjoy this issue. Whether you do or don’t, please engage with the authors and their work, engage with our editorial board or with me directly. Help us make this *your* magazine! If your life isn’t entwined with the magazine as much as mine yet, we hope to help you get there.

## NEW EDITORIAL BOARD MEMBERS



**Christian Sandor** (chris.sandor@gmail.com) is an associate professor at one of Japan’s most prestigious research universities, Nara Institute of Science and Technology (NAIST), where he is co-directing the Interactive Media Design Lab together with Professor Hirokazu Kato. Since 2000, his foremost research interest is Augmented Reality, as he believes that it will have a profound impact on the future of mankind.

In 2005, he obtained a doctorate in computer science from the Munich University of Technology, Germany under the supervision of Gudrun Klinker and Steven Feiner. He decided to explore the research world in the spirit of Alexander von Humboldt and has lived outside of Germany ever since to work with leading research groups at institutions including: Columbia University (New York, USA), Canon’s Leading-Edge Technology Research Headquarters (Tokyo, Japan), Graz University of Technology (Austria), University of Stuttgart (Germany), and Tohoku University (Japan).

Before joining NAIST, he directed the Magic Vision Lab (<http://www.magicvisionlab.com>). Together with his students, he won awards at the premier Augmented Reality conference, IEEE International Symposium on Mixed and Augmented Reality: best demo (2011, 2016) and best poster honourable mention (2012, 2013). He has presented several keynotes and has acquired 2 million USD funding; in 2012, Magic Vision Lab was the first, and still only, Australian lab to be awarded in Samsung’s Global Research Outreach Program. In 2014, he received a Google Faculty Award for creating an Augmented Reality X-Ray system for Google Glass.

Sandor is an editorial board member of *Elsevier Computers & Graphics* and *IEEE Computer Graphics and Applications*. He is a steering committee member of ACM Symposium on Spatial User Interaction and IEEE International Symposium on Mixed and Augmented Reality.



**Amit Agrawal** (amit.agrawal@me.com) is a senior technology executive specializing in new product development leveraging new technology. He has over 20 years of experience in multimedia technologies and spent a few years exploring the application of these technologies in the educational domain. These days, he is principal at Kleene Closure Consulting where he helps companies apply emerging technologies including machine learning and VR to their workflow.

Previously, as CTO of Auryn Inc., he was responsible for developing their ground breaking proprietary technologies which brought nonphotoreal rendering technologies to market in the form of animated films and children’s apps. Before joining Auryn, Amit was the Executive Director of Software Engineering at Sony Pictures Imageworks, where, he supervised software development for both computer animated and visual effects feature films including *The Polar Express*, *Spider-Man 1 and 2*, *Stuart Little*, *Hollow Man*, and the Oscar-winning short *Chub Chubs*. Agrawal holds a BTech in computer science from the prestigious Indian Institute of Technology as well as a PhD from the Computer Science Department at the University of Southern California.



**Melanie Tory** (mtory@tableau.com) is a senior research scientist focusing on interactive visual data analysis. Her research explores techniques and tools to help people analyze data more effectively. This includes intuitive interactions with visualizations and the design and evaluation of tools that support the holistic data analysis process, including sensemaking, analytical guidance, and collaboration. Before joining Tableau, Melanie was an associate professor in visualization at the University of Victoria. She earned her PhD in Computer Science from Simon Fraser University and her BSc from the University of British Columbia. She is associate editor of *IEEE Computer Graphics and Applications* and has served as Papers Co-chair for the IEEE Information Visualization and ACM Interactive Surfaces and Spaces conferences. She has also served as a member of the IEEE InfoVis Steering Committee since 2017.

# Improving Visual Communication

**Gary Singh**

A tool that creates realtime interactive color maps for scientific visualization helped enhance the dynamics of a major research project for the Climate, Ocean, and Sea Ice Modeling team at Los Alamos National Laboratory.

Artists and scientists work from different perspectives that can either become a hindrance or an advantage. Creative work tends to emerge from a first person, subjective experience, while scientists tend to require a third person, objective viewpoint. But there is no need to assume the two perspectives are incompatible. A meeting of the minds can indeed occur, often leading to more improved methods of visually communicating a scientific problem.

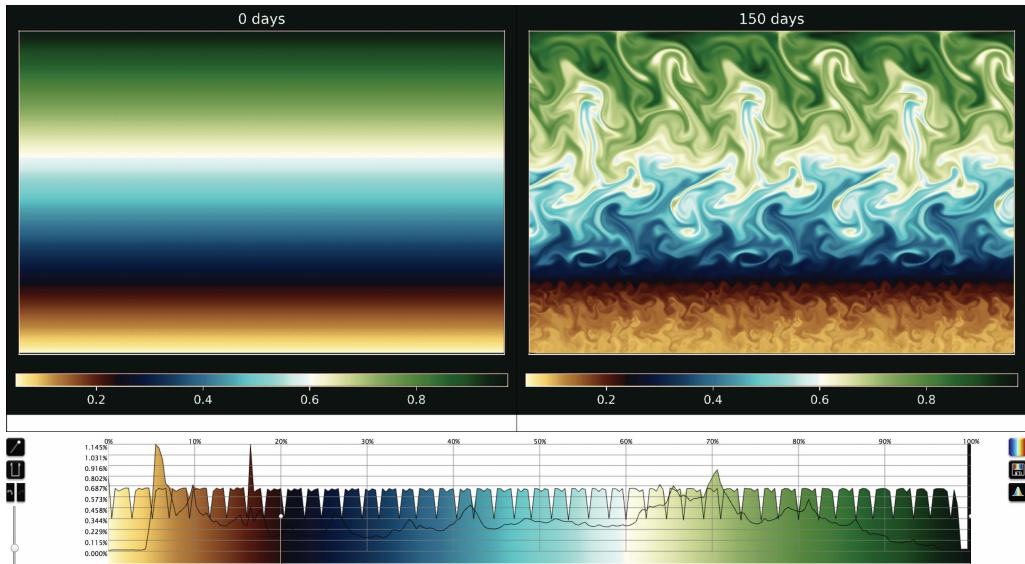
Enter artist Francesca Samsel, a research associate at the University of Texas at Austin and member of the *IEEE Computer Graphics and Applications* Editorial Board, who actually didn't even think about any such differences of perspective when collaborating with the Los Alamos National Laboratory (LANL) in New Mexico. By introducing a tool called ColorMoves that creates realtime interactive color maps for scientific visualization, Samsel was able to enhance the dynamics of a major research project in the Climate, Ocean, and Sea Ice Modeling (COSIM) team at LANL. (An article on ColorMoves is also presented in this issue.)

The resulting cover image, supplied by Samsel, is a snapshot of a particular dataset from Phil Wolfram's work in the COSIM team. The colors indicate mixing by ocean currents due to ocean vortices and waves after 100 days in an idealized Antarctic Circumpolar Current. Figure 1 illustrates the starting condition and the mixed state at 150 days. Initially, the colors were constant across the horizontal axis and smoothly varying in a gradient up and down, but they are now mixed by the eddies and Rossby waves of the simulation. Throughout their research using the ColorMoves interface, the scientists were able to pull out detail in their visualizations that was previously hard to understand without a background in color theory.

"The image for the cover is a good snapshot of why fluids matter," Wolfram says. "Mixing is one of the most fascinating physical phenomena and it was really easy to see how studying ocean turbulence that mixes heat and CO<sub>2</sub> from the atmosphere into the ocean is a way to bridge my interests and needs in the earth system modeling sector."

A civil engineer, Wolfram completed his undergraduate degree at the Colorado School of Mines and then went to graduate school at Stanford before discovering computational fluid mechanics and taking a PhD with coastal oceanographer Oliver Fringer and a postdoc with global oceanographer Todd Ringler. Among other subjects in the COSIM team, he studies ocean mixing and its effect on temperature and the global climate. He is particularly concerned with doing science in

the large, to address issues affecting the planet on which we live and breathe, as well as the natural resources we need to survive.



**Figure 1.** This visualization is based on the same variable as that shown on the cover but illustrates the starting condition and the mixed state at 150 days.

Samsel comes from a decades-long background in arts academia with visual design skills, plus video and sculpture experience. Before moving to Texas, she spent 20 years in New York, producing work for galleries and large scale public art projects. In Austin, she was eventually asked to do some work at the Texas Advanced Computing Center Visualization Lab, but from an artist's point of view. Through that work, she did a large series of pieces for an 80-monitor display wall, a hybrid of visualization, drawing and inspiration from other sources. All of which eventually led to applying for an artist in residence position at Los Alamos. Science has always influenced her work, so she was a natural at embedding herself within a scientific team.

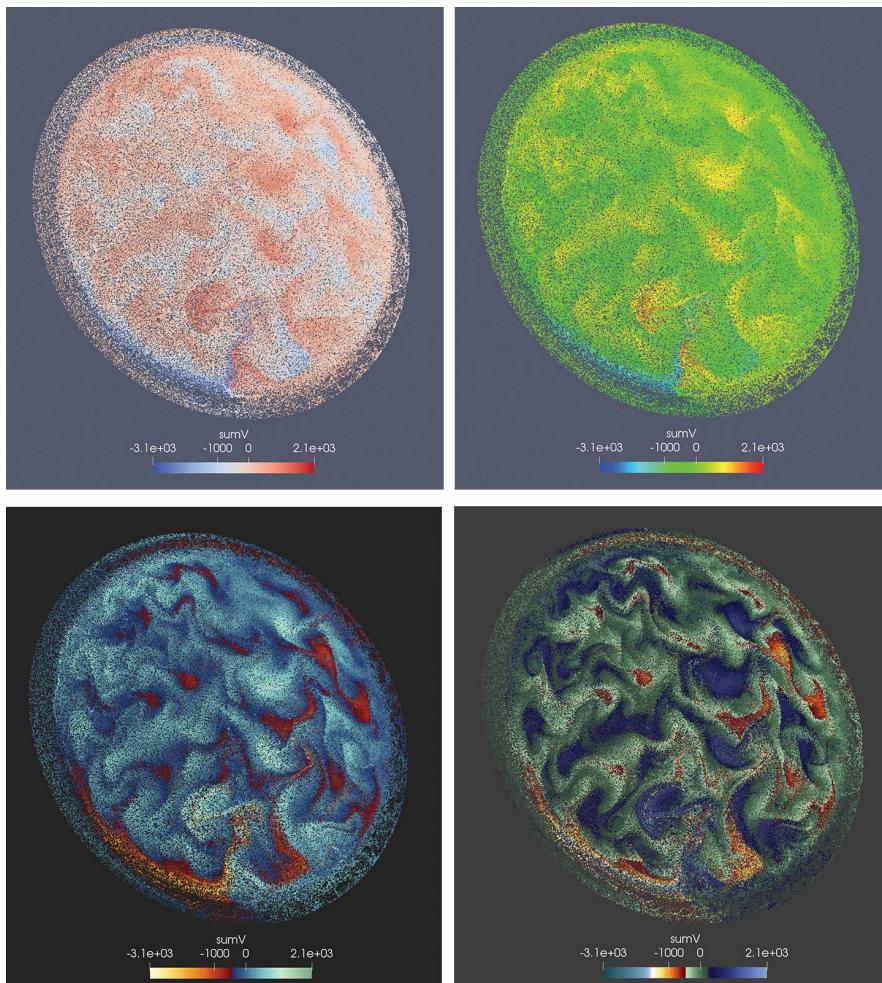
"The key to any successful kind of art-sci-tech collaboration is learning each other's language and thus being able to figure out where the skills and expertise from each discipline can impact one another," Samsel says. "The conversations propel both fields forward."

In collaborative scenarios, scientists might intrinsically believe the artist is only present to help provide more interesting visuals and not play a serious role in the research process. But Samsel says she can recount several examples where scientists were surprised at how an artist actually altered the course of the project in a positive fashion, even if just by asking a lot of questions. According to Samsel, she has been surprised by scientist's comments such as, "You impacted the course of our research." and more numerously, "Are you here to kick us in the head again?" meaning are you here to shake things up again, all of which are compliments, as far as artists are concerned.

"What it's really about, is not that the research 'changes,' but that the combination of the discussions transforms both the art and the science," Samsel explains. "That's what makes the collaborative process so interesting."

At LANL, Samsel came into the picture and eventually began working with Wolfram on numerous visualization challenges. Recent collaborations involved the ColorMoves project. Wolfram says he took away a deeper belief that the key to art is clarity in a visual presentation.

"We all can be better communicators and often this is discussed in the context of written or spoken word," he says. "However, the images we make are potentially more important because they affect the scientific conclusions we draw. Even something simple like picking the right colorbar is key."



**Figure 2.** A particle simulation shows different colormaps generated in ColorMoves from Model for Prediction Across Scales-Ocean (MPAS-Ocean) data dealing with mixing. It speaks to highlighting different features versus standard colormaps that give an overall view.

Which is precisely where the ColorMoves plays a key role. Wolfram says ColorMoves provides a fast and efficient way to design colorbars with an emphasis on how to best select a colorbar to best understand the data (see Figure 2). Nothing similar previously existed.

“It is possible to design colorbars by hand, however, one simply can’t build them quickly or as effectively as with this tool,” Wolfram says, adding that such a tool isn’t needed for every single application, but in many instances this information is vital to understanding a scientific problem. “I suspect many innovative problems will benefit from this tool, especially when scientists want to quickly and effectively communicate with others. A good image makes the science clear and a tool like ColorMoves is indispensable toward this goal.”

At the moment, ColorMoves is still an exterior tool. Researchers still have to import data into it, and then export the information back into their visualization tool. For the future, Samsel hopes to find ways for scientists to plug the interface directly into their own systems.

## ABOUT THE AUTHOR

**Gary Singh** lives and writes in San Jose, California.

# A Taxonomy of Spatial Interaction Patterns and Techniques

**Jason Jerald**  
NextGen Interactions

**Editors:**  
Frank Steinicke  
[frank.steinicke@uni-hamburg.de](mailto:frank.steinicke@uni-hamburg.de)

Christian Sandor  
[chris.sandor@gmail.com](mailto:chris.sandor@gmail.com)

This article presents a taxonomy of spatial interaction patterns and techniques and discusses specific examples.

What is the best spatial interaction technique? Like any great question, the answer is “it depends.” There are numerous factors to consider when it comes to designing interfaces such as high-level goals, hardware constraints, and the target audience.

Instead of having a one-solution-fits-all answer, spatial interaction designers and developers who have a diverse understanding of different options will be able to innovate more quickly across a wider range of use cases. Understanding such tradeoffs is especially important now that a diverse set of hardware input options are becoming less expensive and more available. One way to build such expertise is by studying what others have done in 3D user interface research.<sup>1</sup> In this article, I review a taxonomy of interactions<sup>2</sup> and discuss some examples that fit within that taxonomy.

## OVERVIEW OF PATTERNS AND TECHNIQUES

A *taxonomy* is a classification of things or concepts. Although taxonomies are not required to be hierarchical, structuring interaction concepts hierarchically enables us to think about interactions at different levels of abstraction. In the taxonomy described here, we have meta patterns at the highest level, more specific patterns at the middle level, and interaction techniques at the most detailed level. This structure provides the following benefits<sup>2,3</sup>:

- Broader pattern names and concepts enable easier communication of high-level interaction concepts.
- Mid-level groupings facilitate systematic analysis and comparison.
- Organizing interaction techniques under the umbrella of an interaction pattern makes it easier to consider appropriate design possibilities by focusing on conceptual utility and higher-level design decisions before worrying about more specific details.
- There are too many existing interaction techniques with many names and characterizations to remember, and certainly many more will be developed in the future.

- When a specific technique fails, then other techniques within the same pattern can be more easily thought about and explored, resulting in better understanding of why that specific interaction technique did not work as intended.

I define an *interaction pattern* to be a high-level interaction concept that can be used repeatedly across different applications to achieve common user goals. Interaction patterns are quite different than software design patterns that readers may be familiar with. Whereas software design patterns describe software architecture, interaction patterns describe conceptual relationships and interactions between the user and the virtual world from the user's perspective.

I define an *interaction technique* to be a more specific description of an actual implementation that is more technology dependent than an interaction pattern. Different interaction techniques that are similar are grouped under the same interaction pattern. For example, the walking pattern consists of moving the legs to navigate and covers several walking interaction techniques<sup>4</sup> including real walking that works with room-scale VR setups such as with the HTC Vive, redirected walking<sup>5</sup> as used by The Void ([www.thevoid.com](http://www.thevoid.com)), treadmill walking as done with the Virtuix Omni ([www.virtuix.com](http://www.virtuix.com)), and walking-in-place<sup>6</sup> as implemented by Stompz ([www.roaddrivr.com/ground-control-stompz-two-natural-vr-controllers-feet](http://www.roaddrivr.com/ground-control-stompz-two-natural-vr-controllers-feet)). Each of these specific walking techniques has advantages and disadvantages.

Such an interaction taxonomy provides models to experiment with, suggestions and warnings of use, and starting points for innovative new designs. Interaction designers benefit by having a library of options to choose from depending on their needs and a base to innovate upon. Each pattern or technique has strengths and weaknesses depending on many factors such as application goals and the intended audience. By better understanding distinctions and tradeoffs of the various options, designers can empower themselves to create higher-quality interactive experiences.

I considered more numerous interaction techniques and organized those techniques into five pattern groupings and 16 patterns as shown in Figure 1.

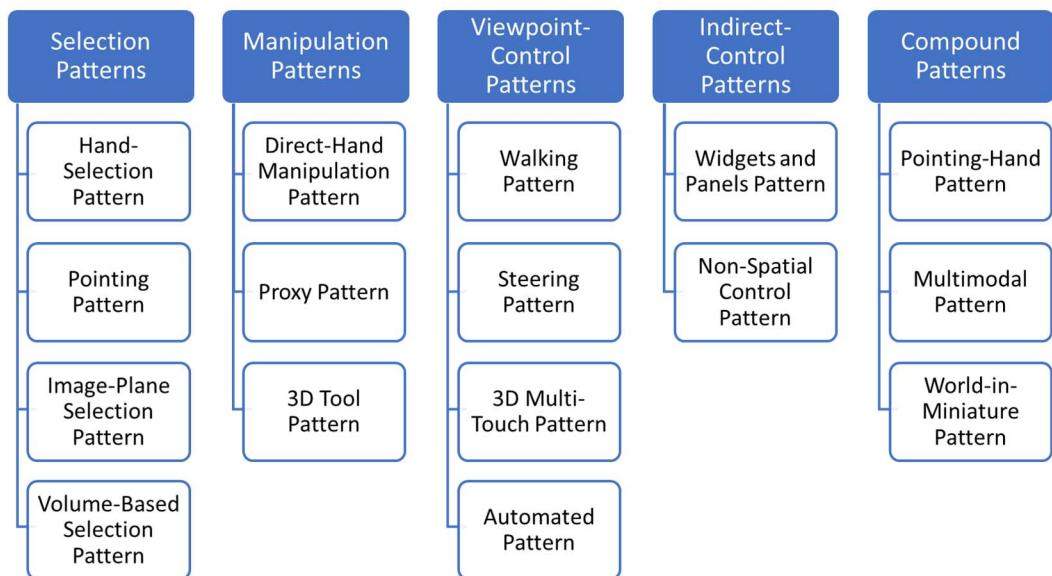


Table 1. 16 Interaction Patterns organized under five pattern groupings. More specific techniques are implementations of the 16 patterns.

Whether you use these suggested interaction patterns or form your own patterns, thinking about such interactions at a high level can be quite useful when designing for spatial interactions by enabling your team to think creatively beyond the most commonly used interaction metaphors. There is not time or space to go into detail of all patterns and techniques in this short article.

However, some examples of Selection Patterns, Viewpoint Control Patterns, and Compound Patterns are discussed below.

## SELECTION PATTERNS EXAMPLES

*Selection* is the specification of objects, data, or space and can be used to specify the target to which a command will be applied, denote the beginning of a manipulation task, or specify a target to travel toward.<sup>7</sup> Selection is not necessarily obvious for spatial interfaces, especially when objects are located at a distance from the user. Selection Patterns include the Hand-Selection Pattern, Pointing Pattern, Image-Plane Selection Pattern, and Volume-Based Selection Pattern. Here, the Hand-Selection Pattern is discussed.

### The Hand-Selection Pattern

The *Hand-Selection Pattern* is a way to select objects via direct object-touching that mimics real-world interaction—the user directly reaches out the hand to intersect some virtual object and then triggers a grab (e.g., pushing a button on a controller, making a fist, or uttering a voice command).

As shown in Figure 2, the visual implementation of hand selection can range from realistic hands (e.g., a high-quality embodied avatar with arms) to no hands where only the hand-held controllers are rendered). Each of these have their own advantages and disadvantages. For example, hands with arms can look great when the full body is tracked, or the user is seated and not expected to move or rotate her torso. However, imperfect inverse kinematics can cause breaks-in-presence due to the arms behaving in strange ways when only the head and hands are tracked and assumptions are violated.

The mapping of physical hand locations does not necessarily need to have perfect spatial compliance with the virtual hands. The *Go-Go Interaction Technique*<sup>8</sup> is a subset of a non-realistic or semi-realistic hand that enables the user to reach far beyond personal space. The virtual hand is directly mapped to the physical hand when within 2/3 of the full arm's reach. However, when extended further, the arm stretches in a nonlinear manner enabling the user to reach further into the environment. This technique enables closer objects to be selected (and manipulated) with greater accuracy while allowing further objects to be easily reached. Physical aspects of arm length and height have been found to be important for the go-go technique, so measuring arm length should be considered when using this technique. Arm length can be estimated by simply asking the user to hold out the hands in front of or to the side of the body at the start of the application. Bowman and Hodges<sup>9</sup> describes extensions to the go-go technique, such as providing rate control (i.e., velocity) options that enable infinite reach, and compares these to pointing techniques. *Non-isomorphic hand rotations*<sup>10</sup> are similar but scale rotations instead of position.

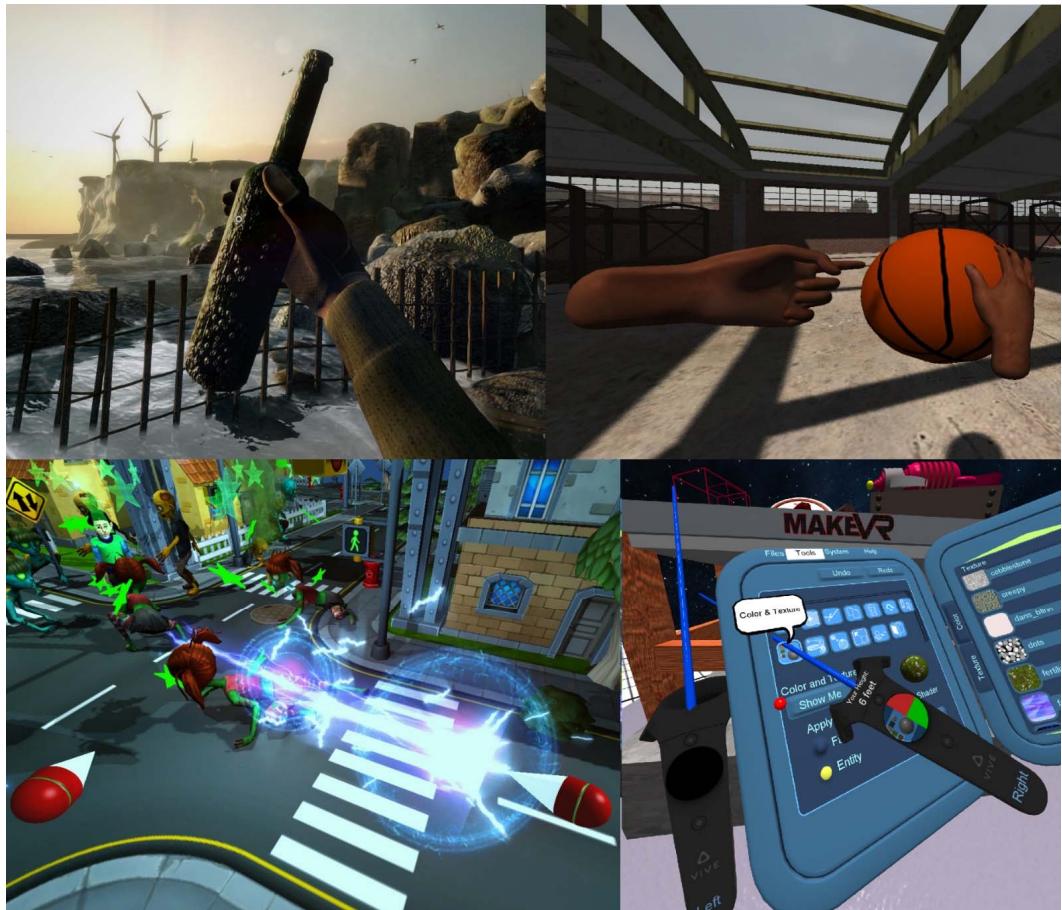


Figure 2. A realistic hand from The Gallery by Cloudhead Games (upper left), semi-realistic hands from Train Wreck by NextGen Interactions (upper right), non-realistic hands (3D cursors) from Zombie Apocalypse by Digital ArtForms (lower left), and no hands with controllers and panel held in the left hand from MakeVR by HTC, Sixense, Digital ArtForms, and NextGen Interactions (lower right).

## VIEWPOINT-CONTROL PATTERNS

*Viewpoint control* is the task of manipulating one's perspective (position, orientation, and/or scale). Viewpoint-Control Patterns include the Walking Pattern, Steering Pattern, 3D Multi-Touch Pattern, and Automated Pattern.

### The Steering Pattern

The *Steering Pattern* is continuous control of motion without movement of the feet. Example steering techniques include navigation by leaning, gaze-directed steering, torso-directed steering, one or two-handed flying, dual-analog stick steering, world-grounded steering devices, and virtual steering devices.

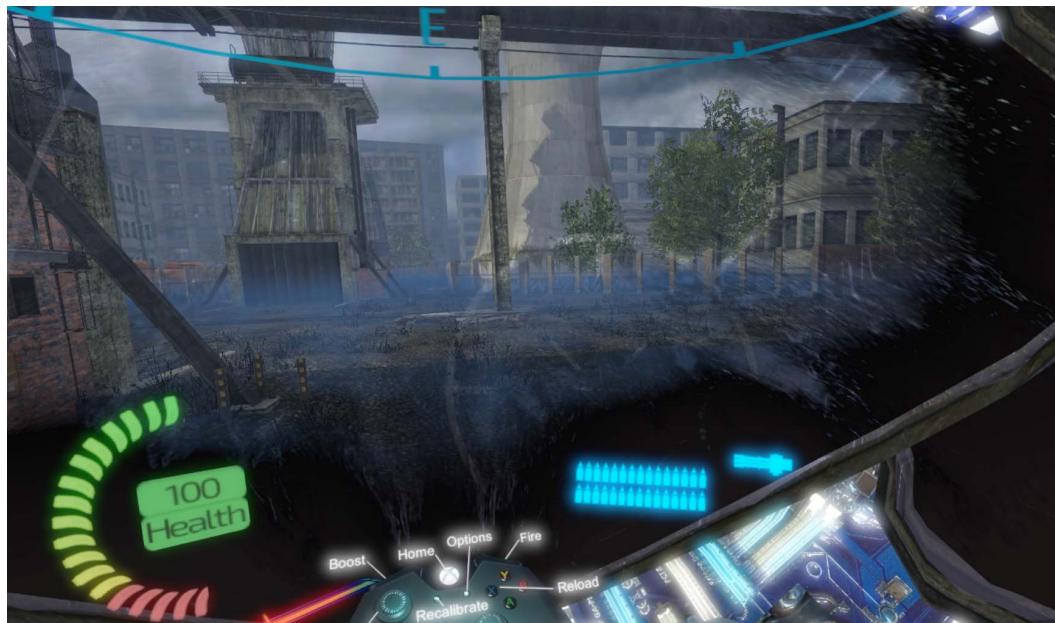
Steering is appropriate for traveling great distances when users don't want to exert much physical energy and when continuous control is important. Steering techniques should be designed to be simple, so the user can focus on spatial knowledge acquisition and information gathering. Steering works best when travel is constrained to some height above a surface and acceleration/deceleration is limited.

Unfortunately, steering can cause visually-induced motion sickness resulting from visual-vestibular conflict. This can be solved by using a motion platform so that physical motion is more congruent with visual motion. Unfortunately, motion platforms can be expensive. Thus, the Steering Pattern is usually not appropriate for those who are new to VR and motion-sensitive users.

Although there are no perfect solutions to eliminating motion sickness for steering without the use of motion platforms, there are ways to reduce sickness. For example, the field of view (FOV) in a HMD can be dynamically reduced during provocative motion.<sup>11</sup> This removes optical flow in peripheral vision, resulting in less visual-vestibular conflict and decreased motion sickness.

Rest frames can also be used to increase comfort while steering. A rest frame is a portion of the virtual environment that remains fixed in relation to the real world and does not move as the user physically or virtually moves. Rest frames can feel quite natural in some situations (e.g., sitting in a stabilized cockpit while navigating through a virtual world) but can reduce situational awareness due to the rest frame occluding parts of the virtual world. The concept of dynamic FOV and rest frames can be combined to create dynamic rest frames—rest frames that are invisible when the user is not virtually moving, but fade in as provocative motion occurs.<sup>12</sup>

At NextGen Interactions, we designed and built the game VR Apocalypse that utilizes dual-analog stick steering along with a dynamic FOV, a cockpit that serves as a static rest frame, and a dynamic rest frame to reduce motion sickness (<http://store.steampowered.com/app/554940>). The dynamic rest frame consists of fogging up the windows on the sides and top/bottom of the cockpit's windshield with a dark oily material (Figure 3). The material contains texture so that it not only reduces optical flow but also serves as a dynamic rest frame that contains visual structure on the cockpit's windows. To help maintain presence, this fogging up of the windows was made to be a part of the story that is explained in the 'contractor training' starting tutorial where the player learns to pilot his drone through a futuristic VR telepresence system. One difference of this reduced FOV from Fernandes and Feiner<sup>11</sup> and other VR games is that the FOV is only reduced outside the rest frame cockpit since no visual-vestibular conflict occurs inside the cockpit. Because the dynamic rest frame and dynamically reduced FOV only changes on the outside of the windows, they are less noticeable.



**Figure 3. Static and dynamic rest frames implemented in the game VR Apocalypse by NextGen Interactions. The dark oily material fades in on the sides and top/bottom of the cockpit's windshield during provocative virtual motion, which serves as a dynamic rest frame to reduce motion sickness.**

## The 3D Multi-Touch Pattern

The *3D Multi-Touch Pattern* enables users to simultaneously adjust the position, orientation, and scale of the world. Similar to 2D multi-touch on a touch screen, translation via 3D multi-touch is obtained by grabbing and moving space with one or both hands. A common way to use 3D multi-touch is to “crawl” with the hands by alternating the grabbing of space with each hand (like pulling on a rope but with hands typically wider apart). Initially this seems like a lot of work. However, by first scaling the world smaller (grabbing space with both hands and bringing the hands closer together), the user can crawl long distances quite efficiently. Rotation of the world is accomplished by grabbing space with both hands and rotating about a point (typically implemented as either about one hand or the midpoint between the hands).

Controlling one’s viewpoint is inversely equivalent to moving, rotating, or scaling the world. For example, moving the viewpoint to the left is equivalent to moving the world to the right, or scaling oneself to be smaller is equivalent to scaling the world to be larger. Users can think of 3D multi-touch as either 1) manipulating the world as an object (world motion), or 2) moving and scaling oneself through the world (self-motion). One’s mental model can affect motion sickness, and those that think of manipulating the world as an object may experience less motion sickness than those that think about using their hands to transform themselves through space. However, when observing avatars of other 3D multi-touch users, one cannot help but perceive those avatars as moving and scaling themselves since the world doesn’t change for the observer just because other users modify their own perspective.

3D multi-touch works well for non-realistic exploration of large and small areas of interest from arbitrary viewpoints, visualizing and manipulating abstract or scientific data, and placing assets. Like any other interaction pattern, 3D multi-touch is not appropriate for all situations, e.g., when the user is confined to the ground.

3D multi-touch can be challenging to implement as small nuances can affect usability. If not implemented well, 3D multi-touch can be frustrating to use and sickness inducing. Constraints, such as limiting scale, not allowing roll and pitch rotations, or even disabling rotations completely can be added for novice users and for those who are more prone to motion sickness. In addition, techniques can be added to further reduce motion sickness, in similar ways that can be done with the Steering Pattern. Without strong depth cues (most notably occlusion and stereoscopic cues), scaling can be confusing since it can be difficult to distinguish between a small nearby object and a larger object that is further away. Simple additions such as adding a ‘spindle’<sup>13</sup> marker at the center of scale/rotation significantly helps with planning, enhancing depth perception, and setting expectations of how motion will occur.

Even if implemented well, there can be a learning curve on the order of several minutes for some users.<sup>13</sup> By teaching from within VR via an in-app tutorial, we have reduced the time to learn 3D multi-touch to under three minutes in the immersive modeling application MakeVR (<http://ViveForMakers.com>). Once learned, 3D multi-touch is quite intuitive and works well when navigation and selection/manipulation tasks are frequent and interspersed, since viewpoint control and object control are nearly identical other than pushing a different button (i.e., cognitive load is reduced through the same concept of grabbing objects and grabbing the world as an object). Although 3D multi-touch can appear quite cumbersome to an observer, this pattern provides the ability to place objects and regions of interest into personal space where the hands can work comfortably in the lap or at the waist via position, rotation, and scaling operations. As a result, 3D multi-touch can result in less gorilla arm than other spatial interfaces and users have worked for hours without reports of fatigue.<sup>14</sup> In addition, physical hand motions are non-repetitive by nature and thus are not as subject to repetitive stress due to the lack of a physical planar constraint as is the case with a mouse.

Although 3D multi-touch has rarely been used in the past, it is now becoming more widely accepted, as it is implemented in applications such as MakeVR, the Unity VR Editor (<https://github.com/Unity-Technologies/EditorVR>), and the Unreal Engine VR Editor (<https://docs.unrealengine.com/latest/INT/Engine/Editor/VR>).

## COMPOUND PATTERNS

*Compound Patterns* combine two or more patterns into more complicated patterns. Compound Patterns include the Pointing-Hand Pattern, Multimodal Pattern, and World-in-Miniature Pattern.

### The Multimodal Pattern

The *Multimodal Pattern* integrates different sensory/motor input modalities together. The classic example technique of the Multimodal Pattern is the *Put-That-There Interaction Technique*<sup>15</sup> that uses a combination of the Pointing Pattern (to select the ‘that’ and there’) and the Non-Spatial Control Pattern via voice (to select the verb ‘put’). Neely et al.<sup>16</sup> implemented an immersive form of a put-that-there interaction style of region definition. With this technique, the user verbally names and defines the vertices of a polygonal region on a terrain via pointing and speaking. An example is the user speaking “Make Target Zone Alpha from here (pointing gesture) to here (pointing gesture) and to here (pointing gesture).” For manipulating a single object, a that-moves-there interface can be more efficient than a put-that-there interface as object-action sequences require less mental effort than action-object sequences.<sup>17</sup> An example of a that-move-there interface would be “Waypoint Gamma move here (pointing gesture).”

### The World-in-Miniature Pattern

The *World-in-Miniature Pattern* combines selection and manipulation patterns with the Proxy Pattern. A world-in-miniature is an interactive live 3D map—an exocentric miniature graphical representation of the virtual environment one is simultaneously immersed in.<sup>18</sup> A world-in-miniature not only enhances situational awareness of a small or remote area, but also enables the user to reach inside the world-in-miniature and move proxy objects that also move the objects in the larger surrounding virtual environment. In some implementations, the user can also move an avatar or ‘doll’ proxy representing the self within the world-in-miniature. The user then gives a command to teleport himself (a form of the Automated Viewpoint Control Pattern) to that location in the larger world. The Unity VR editor contains an example implementation of this World-in-Miniature Pattern called the MiniWorld.

At Digital Artforms, we implemented a form of the World-in-Miniature Pattern that also includes the Volume-Based Selection Pattern and 3D Multi-Touch Pattern that we call the Viewbox.<sup>19</sup> The user creates a cube and then stretches that cube’s axii about the volume of interest with a two-handed interface.<sup>20</sup> Once the volume of interest is defined, the Viewbox can be selected and manipulated like any other object; it can be scaled or placed anywhere in the world, or carried along with the user by attaching it to a hand-held panel of tools or to the torso-reference frame (e.g., as if attached to the user’s belt). In addition, the user can reach inside the ViewBox and manipulate the space within it via 3D multi-touch in the same way that she can manipulate the larger surrounding space.

## CONCLUSION

None of these interaction patterns and techniques are meant to be plug and play. Quality spatial interaction design requires a human-centered iterative development approach more so than for any other field. Do not assume that if a technique does not work on its first implementation that it is necessarily a bad fit for the project’s needs. It is important to experiment, rapidly prototype, evaluate and get feedback from users/stakeholders, and tweak the design until it works well or it is obvious that no amount of adjustment will make it a good fit for the project. Most importantly, spatial interaction designers and developers can learn from the attempt and build up expertise for future work.

## ACKNOWLEDGMENTS

This article is largely based on *Chapter 28 Interaction Patterns and Techniques of The VR Book: Human-Centered Design for Virtual Reality*<sup>2</sup> co-published by Morgan & Claypool Publishers and ACM Books. More information on the book is available at <http://TheVRBook.net>.

## REFERENCES

1. J. LaViola et al., *3D User Interfaces: Theory and Practice*, Addison Wesley, 2017.
2. J. Jerald, *The VR Book: Human-Centered Design for Virtual Reality*, Morgan & Claypool Publishers and ACM Books, 2015; <http://TheVRBook.net>.
3. K. Sedig and P. Parsons, “Interaction Design for Complex Cognitive Activities with Visual Representations: A Pattern-Based Approach,” *AIS Transactions on Human-Computer Interaction*, vol. 5, no. 2, 2013, pp. 84–133; <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1057&context=thci>.
4. F. Steinicke et al., *Human walking in virtual environments*, Springer, 2013.
5. S. Razzaque, Z. Kohn, and M. Whitton, “Redirected Walking,” *Eurographics*, 2001, pp. 289–294.
6. J.N. Templeman, P.S. Denbrook, and L.E. Sibert, “Virtual locomotion: Walking in place through virtual environments,” *Presence: teleoperators and virtual environments*, vol. 8, no. 6, 1999, pp. 598–617.
7. R.P. McMahan, R. Kopper, and D.A. Bowman, “Principles for Designing Effective 3D Interaction Techniques,” *Handbook of Virtual Environments*, CRC Press, 2014; doi.org/http://dx.doi.org/10.1201/b17360-6.
8. I. Poupyrev et al., “The Go-Go Interaction Technique: Non-linear Mapping for Direct Manipulation in VR,” *ACM Symposium on User Interface Software and Technology*, 1996, pp. 79–80.
9. D.A. Bowman and L. Hodges, “An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments,” *ACM Symposium on Interactive 3D Graphics*, 1997, pp. 35–38.
10. I. Poupyrev, S. Weghorst, and S. Fels, “Non-isomorphic 3D Rotational Techniques,” *SIGCHI Conference on Human Factors in Computing Systems*, 2000, pp. 540–547.
11. A.S. Fernandes and S.K. Feiner, “Combating VR sickness through subtle dynamic field-of-view modification,” *IEEE 3D User Interfaces*, 2016.
12. Z. Cao, J. Jerald, and R. Kopper, “Visually-Induced Motion Sickness Reduction via Static and Dynamic Rest Frames,” *Submitted to IEEE Transactions on Visualization and Computer Graphics*, 2018.
13. U. Schultheis et al., “Comparison of a Two-Handed Interface to a Wand Interface and a Mouse Interface for Fundamental 3D Tasks,” *IEEE Symposium on 3D User Interfaces*, 2012, pp. 117–124.
14. J. Jerald et al., “MakeVR: A 3D World-Building Interface,” *Symposium on 3D User Interfaces*, 2013, pp. 197–198.
15. R.A. Bolt, ““Put-That-There”: Voice and Gesture at the Graphics Interface,” *Proceedings of the 7th annual conference on Computer graphics and interactive techniques (SIGGRAPH 80)*, 1980, pp. 262–270.
16. H.E. Neely et al., “Multimodal Interaction Techniques for Situational Awareness and Command of Robotic Combat Entities,” *Proceedings IEEE Aerospace Conference*, 2004, pp. 3297–3305.
17. R.P. McMahan and D.A. Bowman, “An Empirical Comparison of Task Sequences for Immersive Virtual Environments,” *IEEE Symposium on 3D User Interfaces*, 2007, pp. 25–32.
18. R. Stoakley, M.J. Conway, and R. Pausch, “Virtual Reality on a WIM: Interactive Worlds in Miniature,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 95)*, 1995, pp. 265–272.
19. P. Mlyniec et al., “iMedic: A Two-Handed Immersive Medical Environment for Distributed Interactive Consultation,” *Studies in Health Technology and Informatics*, vol. 163, 2011, pp. 372–378.

20. A. Yoganandan, J. Jerald, and P. Mlyniec, “Bimanual Selection and Interaction with Volumetric Regions of Interest,” *IEEE Virtual Reality Workshop on Immersive Volumetric Interaction*, 2014.

## ABOUT THE AUTHOR

**Jason Jerald** is co-founder and principal consultant at NexGen Interactions, a company that helps companies create quality VR applications. Jason is also adjunct faculty at Duke University, and serves on advisory boards of companies that focus on VR. Jason has worked on over 70 VR projects with over 40 organizations over the last 20+ years. He has authored numerous publications, most notably *The VR Book: Human-Centered Design for Virtual Reality*.<sup>2</sup> Jason earned a PhD in Computer Science on perception of motion and latency in head-mounted displays from the University of North Carolina at Chapel Hill. He can be reached at [Jason@NextGenInteractions.com](mailto:Jason@NextGenInteractions.com).

# ColorMoves: Real-time Interactive Colormap Construction for Scientific Visualization

**Francesca Samsel**  
University of Texas at Austin

**Sebastian Klaassen**  
University of Vienna

**David H. Rogers**  
Los Alamos National  
Laboratory

**Editor:**  
Mike Potel, Wildcrest  
Associates  
[potel@wildcrest.com](mailto:potel@wildcrest.com)

This article presents ColorMoves, an interactive tool that promotes exploration of scientific data through artist-driven color methods in a unique and transformative way. We discuss the power of contrast in scientific visualization, the design of the ColorMoves tool, and the tool's application in several science domains.

The visualization of scientific data is both a science and an art, in which many powerful tools are used to explore, discover and communicate the information within the data. This process is increasingly difficult, as the size and complexity of data is constantly advancing.

As a result, scientists are constantly searching for tools and workflows that help them better understand their data. Color is a potent tool in scientific data visualization, and has been well studied.<sup>1</sup> However, color's full potential for communication and discovery remains untapped. Effective use of color requires a depth of understanding and experience employing color and color relationships to direct attention as well as the tools to translate that knowledge into scientific visualization workflows.<sup>2</sup>

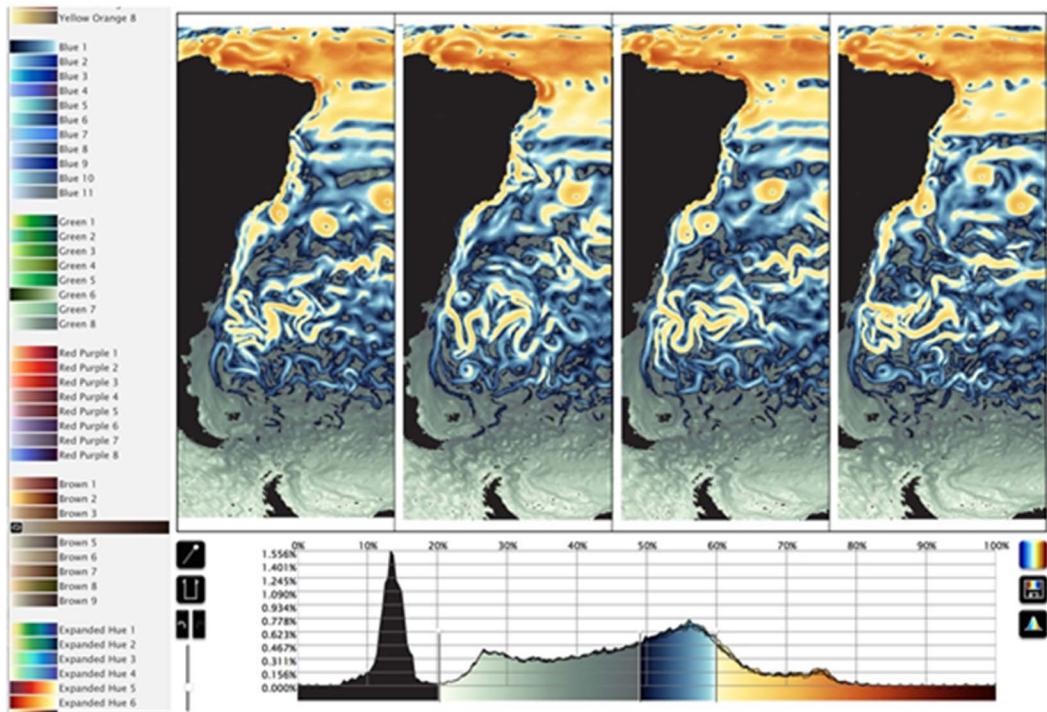


Figure 1. Four time steps of an MPAS–Ocean kinetic energy simulation is shown in the ColorMoves interface. ColorMoves is an interface that enables scientists to construct colormaps with contrast applied to regions of greatest interest. It also promotes interactive exploration of data through color. The ability to load several images into the interface window enables scientists to construct colormaps that are effective across time ranges, camera views, variables and more.

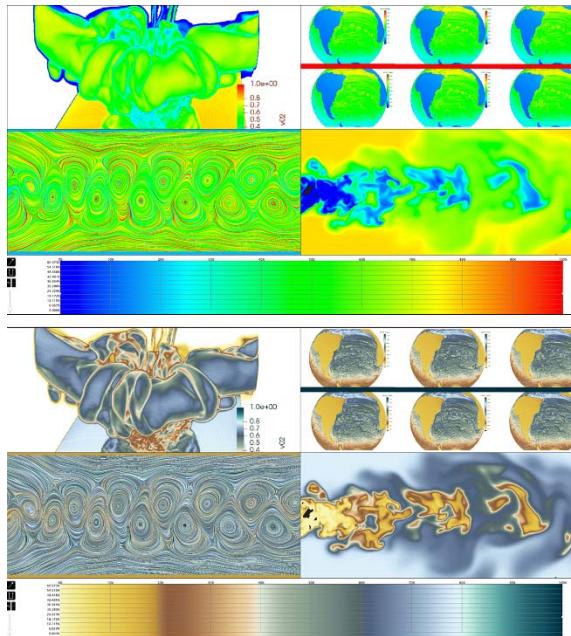
To address this, we present ColorMoves, an online tool offering artist-constructed color scales and palettes in a real-time interactive tool, providing the scientific community with an easy-to-use method of exploring the impacts of changing color encoding on their data. With ColorMoves, a scientist can highlight detail and focus attention on specific regions of interest within their data, and construct exportable colormaps that can be used in most any visualization tool. Though we designed the tool as a means of more effectively encoding and communicating data, we discovered that the domain scientists using the tool were most interested in using the tool as a new capability for data exploration.

The reader can explore ColorMoves, shown in Figure 1, online at <http://sciviscolor.org/ColorMoves>. This paper also includes a discussion on using contrast to communicate effectively, and several examples of using the tool on different science domains.

## EXPLORATION AND COMMUNICATION WITH COLORMOVES

No single colormap is optimal for all science domains, data characteristics and tasks. Detail and nuance within data often goes unnoticed or unexplored because default colormaps do not align the contrast with the areas of interest. Though work has been done to address which colormaps work with different kinds of data,<sup>1</sup> this provides at best a good starting point for coloring data.

For example, Figure 2 shows how the traditional rainbow colormap can obscure data. By applying a colormap lower in saturation but with multiple luminance ranges, we see much greater detail in all of the rendering. Two principles from artistic color theory are applied: maximizing value contrast and avoiding simultaneity of color.<sup>3</sup> We will review the other principles of artistic color theory and how to apply them for scientific benefit.



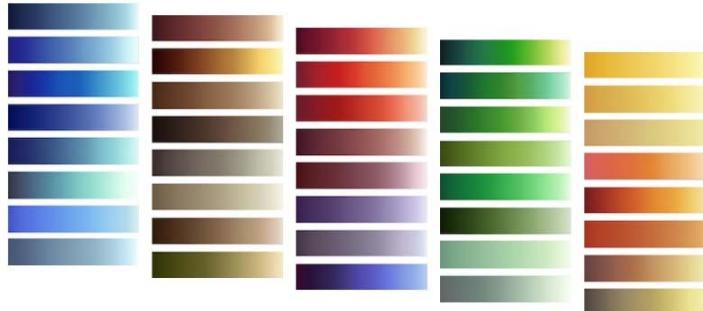
**Figure 2.** Four different data sets are compared using the rainbow colormap and a colormap with a more-muted palette. The shortcomings of the rainbow colormap are readily apparent. ColorMoves includes defaults such as the muted map shown here but also enables creation of alternatives tuned to the specifics of a dataset and task.

While many colormap creation tools are available,<sup>1,4,5</sup> we created ColorMoves to put artistic principles and experience at the fingertips of scientists. Despite the large number of colormaps available in current scientific visualization tools, limited luminance distributions combined with a lack of guidance hindered our scientists' ability to use these effectively. To

achieve our goal, we needed to provide scientists with *effective alternatives*, not just more color options.

Our work transforms color encoding of data into an interactive process enabling scientists to create colormaps customized to their data, its statistical distribution, and areas of importance, combined with intuitive domain-and subject-appropriate color scales.<sup>6</sup>

Default colormaps are designed to provide an overview of a dataset. They are useful for a general understanding of the data and for comparison to previously rendered data.<sup>7</sup> However, once a scientist has that overview, most default colormaps are not very useful for deeper exploration. ColorMoves is designed to enable a more detailed exploration by giving scientists the means of assigning color contrast where it provides the greatest value. It enables intuitive, organized colormaps fitting the domain and/or subject matter. Figure 3 shows the range of color scales that can be combined to suit the topics and structures of the data.

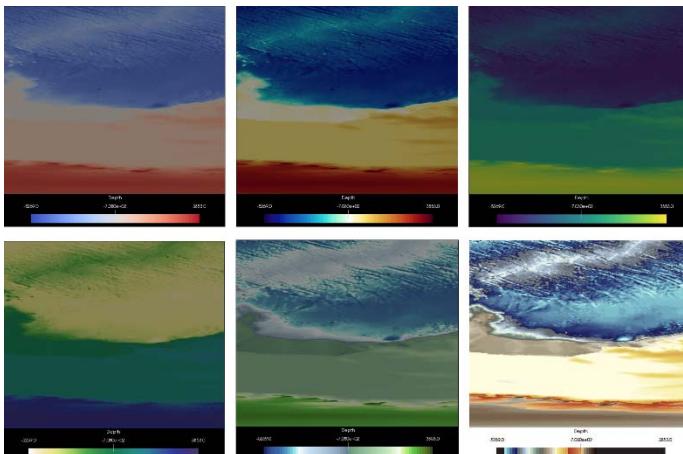


**Figure 3.** ColorMoves provides a wide range of color scales varying in hue range, saturation level and value range. These provide the building blocks needed to create focused attention, hierarchy of interest, categorical and, intuitive association as well as communicative visualization.

It is worth noting domain scientists aren't always looking for one final "optimal colormap." The scientist often needs to look at multiple aspects of the data, and through interactive exploration may discover unexpected properties of the data.

## The Importance of Allocating Contrast

Colormaps reveal data through contrast, and there are multiple types and levels of contrast.<sup>8</sup> The scientist's ability to focus attention and discriminate detail is driven by the type and degree of contrast, not by the specific hues of the colormap.<sup>9</sup> Thus our system focuses on allocation of contrast, specifically luminance and saturation, rather than on contrast of hues. Through application of multiple spans of luminance, ColorMoves can deliver higher degrees of discriminatory power that respond to specific data distributions. Figure 4 illustrates the critical role of luminance distribution in revealing detail. Here, three standard colormaps (top) and two alternative luminance distribution colormaps (lower left and middle) are compared with a map constructed in ColorMoves (lower right), an example of a customized luminance distribution. Another example, in Figure 1, demonstrates the value of focusing attention using color scales spanning in levels of saturation. The ocean eddies are the most important features and thus are rendered in the most saturated color scale: yellow-to-orange. Next, the blue mid-level saturation color scale shows more contextual data and finally a muted green renders areas of minimal importance.



**Figure 4.** Here we show an atmospheric simulation from Pacific Northwest National Laboratory rendered in six colormaps. The lower right version illustrates how the ability to control the luminance distribution is critical to optimizing discriminatory power.

A fundamental tenet of color contrast theory, supported by color perception research, says that our cognitive

interpretation of color is strongly influenced by the adjacent colors (the context).<sup>9</sup> One of the rainbow colormap's problems arises from the abutting saturated hues. Saturated colors, when abutting one another, cause a subtle visual vibration known as simultaneity of color that lowers our ability to assess the details.<sup>3</sup> Generally one can lower the distracting interaction by using muted tones as is shown in Figure 2. The rainbow colormap is in full saturation, therefore there is no control over the luminance distribution. In contrast the wave colormap used in the bottom image employs muted tones, lowering the dissonance between the saturated hues while providing greater discriminatory power via the multiple ranges through luminance. For a more in-depth review of color contrast theory application please see tutorials at <http://sciviscolor.org> or other excellent writing on the subject.<sup>9,10</sup>

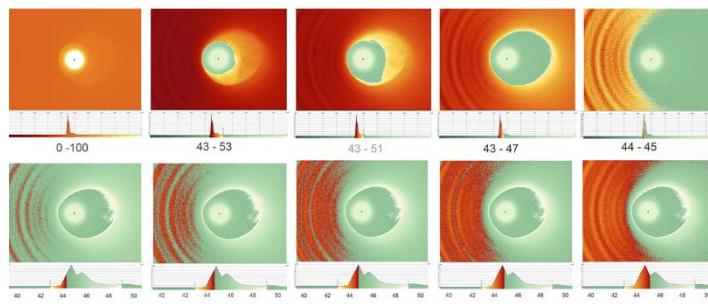
## Exploration

We expected ColorMoves to enable scientists to create visualizations exposing a greater amount of detail in order to communicate their findings more clearly. What came as a surprise is that several of the domain scientists identified ColorMoves' interactive exploration capability as its most significant benefit. Being able to zoom into the histogram and slide color scales across the data range quickly and easily enabled scientists to scan their data for unexpected properties often hidden within their ever-expanding data sets. Our sessions with the scientists that we expected to take fifteen minutes—more than enough time for the scientist to construct and export a colormap for their data—lasted as long as an hour. The scientists spent time experimenting with the placement of up to ten color scales on their data, and watching the changes within each time step. They also experimented by adjusting the specific location of each color scale change. These manipulations are simply not possible in other scientific visualization tools, and are a clear benefit of using ColorMoves.

The linear color scales shown in Figure 3 are the foundation of ColorMoves. Starting with color scales, scientists can apply specific scales to specific data ranges, creating maps addressing the data distributions and task. The color scales come in five sets of hue ranges. Within each set, the scales vary by level of saturation, length of hue range and length of value range. Each of these characteristics aligns with specific functions needed within complex colormaps. To get a closer look at detail within the data, wave colormaps—colormaps that span more than two ranges of luminance—are a good choice. ColorMoves also offers the ability to construct colormaps using domain and or subject appropriate hues,<sup>6</sup> specify areas of data to be highlighted, and create hierarchies via the use of color theory principles to direct focus and control attention.<sup>8,11</sup>

ColorMoves also includes examples of structured colormaps such as wave colormaps and asymmetrical configurations as recommended starting points. These examples include useful alternate colormap structures not typically seen in default lists of colormaps. There are also color sets, which form a cohesive group consisting of either color scales only or discrete colors matched to an underlying color scale.

The visualization series in Figure 5 shows the impact of precisely controlled placement of the color scales within the data range. Here, William Daughton from Los Alamos National Laboratory is interested in the details of bowfronts within a simulation of the magnetic field around Ganymede. On the top left, the red-to-yellow color scale spans the full data range. Moving right, the red-yellow color scale decreases down to spanning only one percent of the data range. By narrowing the span of the color scale, the bowfront waves come into focus. The bottom row narrows in on the precise applied range that most clearly presents the bowfront features. Here the color scale change in feature representation is driven by a less than one percent a percentage shift in coverage. Interactive, precision adjustments of this type are not easily achieved in other scientific visualization tools.



**Figure 5.** Multiple images of the Ganymede magnetic field simulation from William Daughton, Los Alamos National Laboratory. The version in the top row, left, shows the red color scale span the whole data range, top row, right, an inset covering one-percent of the

data. Bottom row varies as a less than one percent change in location or data span. This illustrates the value of being able to interactively adjust the data range precisely while seeing the resulting visualization.

## Communication

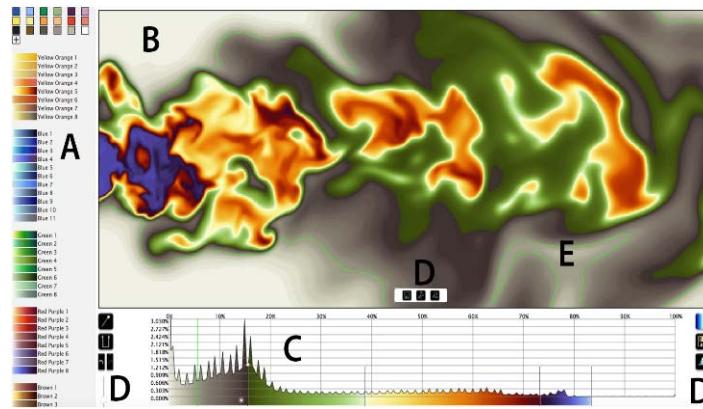
Communication involves describing the relationships and findings within the data. Having control on the hues and saturation levels enables a scientist to create intuitive hierarchies, so that the viewer can easily focus on specific data ranges and be directed via hue.

In Jacqueline Chen's combustion simulation, Figure 6, saturation is used to focus attention in the important areas. The middle section of Figure 6 is comprised of four color scales in descending levels of saturation. The red-to-yellow color scale is applied to the area of most interest. The blue and green colormaps highlight the disparity between the prevalence of material indicated by each section within the data. The brown, being the lowest in saturation provides context without demanding equal attention.

We note that low saturation or muted color scales play an important role in constructing adaptive color encodings. They provide a surprising amount of detail within the limited saturation range and more importantly enable focus to be directed away from these areas.

## Features of ColorMoves

Collaborations with domain scientists drove the ColorMoves interface design shown in Figure 6. Decisions on layout, features and functions developed over iterative consultations. ColorMoves' primary features are outlined below. We encourage you to test the program for yourself as it is through the interaction that its value becomes apparent.



**Figure 6. ColorMoves UI:**  
**A:** selection menu of color scales and discrete colors,  
**B:** viewing window,  
**C:** data distribution histogram,  
**D:** (left) pin splitter, color scale inserter, undo, redo, number of histogram bins,  
**D. (right):** color menu, export, opacity controls,  
**D: (center)** save, fit to screen, resize to original,  
**E:** histogram pin location highlight.

The first step to working with data in ColorMoves is to render your data as a float image, which encodes floating point data values using the red, green and blue color channels of a .png image. Instructions can be found at <http://sciviscolor.org/home/colormoves/data-float-instructions>. This image is the basic data format for ColorMoves. Just drag and drop it into the Viewing Window, and you are ready to experiment with the application of color scales using the features below. Refer to Figure 6 to see the location of each feature within the tool.

**A. Color Sources.** On the left side of the interface is the Color Sources panel. The top area contains discrete color choices as well as a color picker for selecting a specific color. Below the discrete color selection are the color scales divided into hue ranges, each containing multiple hue, saturation and value spans to facilitate creating the attention direction and levels of detail desired. The single hue color scales are followed by a set of linear color maps that span several hues, providing a linear continuity containing greater discriminatory power. The bottom set of colormaps contain inset discrete colors. These provide starting points for those needing to combine discrete colors and color scales. Importing one's own colormaps into ColorMoves is simply a matter of dragging an .xml file describing a colormap into the color scale section (examples are available on the ColorMoves website). The colormap will load into ColorMoves and appear at the bottom of the color selection panel.

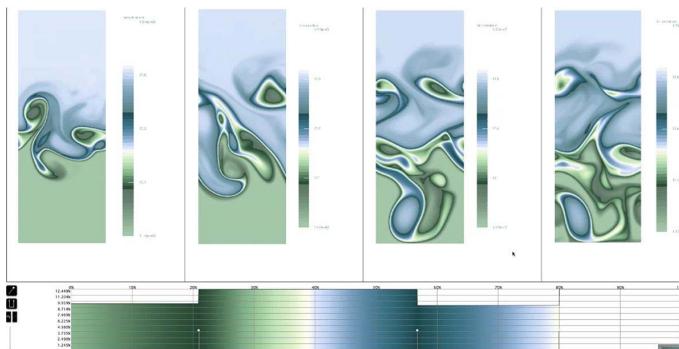
**B. The Viewing Window.** This shows the data colored by the colormap in the Histogram View. An important feature is the ability to load multiple images into the viewing area. This allows the user to work on several images at once, to explore a range of time steps, camera angles, variables or other datasets. The ability to verify that the colormap being developed works across comparisons has proven invaluable in seeing color's impact across an entire dataset, or several different runs of a simulation.

**C. The Histogram View.** This view shows the distribution of the data across the data range. Currently it is distributed by location within the full dataset but plans are underway to include the data range notations as well. If there are multiple images in the viewer then there will be multiple data distributions shown in the histogram. See Figure 6 for a detailed look at this portion of the interface.

**D. Control Icons.** There are three areas of controls associated with the Histogram View. The first set is shown in detail on the left side of Figure 6. The top icon is the pin splitter, which splits the colormap wherever it is dropped, allowing the user to add color scales. Below the pin splitter

is a U-shaped icon we call the nest, indicating the color scales can be applied on top of the existing color scale covering a specific range without impacting the underlying scale. This feature allows a scientist to place detailed color scales in a specific area while maintaining uniformity in the underlying colormap. Both of these are activated by a drag and drop onto the lower edge of the histogram. Next there is a pair of buttons to go back and go forward. Lastly there is a slide that enables one to change the magnitude of the histogram binning.

The second set of controls, on the right side of the histogram, has three icons: the top exposes the color scale selections; the middle enables you to export the new colormap created in the histogram in .xml, .json, or .png format; the third shows the opacity controls. The opacity of each color scale can be independently adjusted as is illustrated in Figure 7.



**Figure 7.** The opacity of the individual color scales is being adjusted in this baroclinic MPAS-Ocean data in a time-varying sequence in order to control the emphasis while maintaining a continuity of hues.

the bottom on the View Window. These enable you to (from left to right) save the file, resize the image to fit the available window, and resize the image to its original size.

**E. Other Features.** Hover over the histogram with a pin activates a green highlight in the image window identifying the data location. A scientist can zoom in on both the histogram and the image in order to precisely adjust the pin.

Other features are context-sensitive, and appear by hovering in specific locations. The trash can for removing pins and nests appears when one hovers over the pin. One simply drags the pin into the trash can to delete it. Radio buttons appear if you hover on a color scale with in the histogram. Sliding a radio button outside of the color scale crops the color scale.

## EXAMPLES FROM THREE SCIENCE DOMAINS

Over the past three years, we have had myriad interactions with domain scientists in a quest to create more effective scientific visualizations through a more directed application of color contrast. We have conducted formal assessment interviews, informal discussions, training sessions and working sessions as we helped them to build data-specific color encodings. Underlying themes emerging in many of these discussions included: 1) the current lack of meaningfully different colormapping schemes, 2) the lack of interactive tools and the length of the time lag between choosing a colormap and seeing it on the data, and 3) the difficulty of extracting the information in these large datasets that remains hidden because of current color approaches. One scientist described his current workflow as: “look at the list of colormaps, pick one, run the script, render, pick another, run the script, render, pick another, etc. In the end, they all look the same.” He described this as a cumbersome iterative approach bringing him no closer to understanding the data. The interactive ColorMoves approach, including the set of color building blocks and the UI, specifically addresses this problem.

To begin exploring data, the first step is to place pins into the Histogram View aligning the pins with the data ranges of interest. Multiple pins can be inserted thus enabling the application of multiple color scales that drive the resolving power of the colormaps. By sliding the pins scientists can focus on and explore multiple features across the data no matter how narrow the data range. The scientists were particularly interested in the ability to zoom into specific regions of the Histogram View, enabling precise placement of color scale boundaries.

The third set is accessed by hovering at

## Climate Science

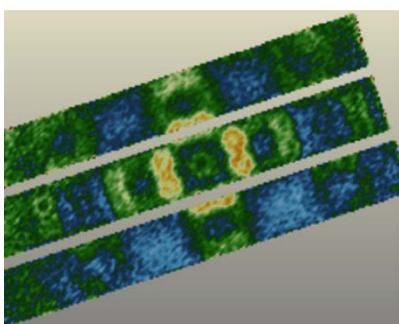
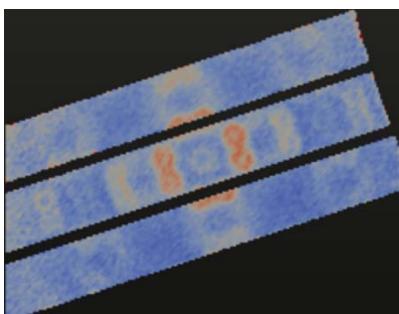
We have worked with multiple climate scientists and modelers, in particular Phillip Wolfram and Mark Peterson of the Climate, Ocean, and Sea Ice Modeling (COSIM) team at Los Alamos National Laboratory. Our work with Dr. Petersen resulted in an award-winning video at Supercomputing 2015.<sup>12</sup> Dr. Wolfram's investigation of mixing properties using the opacity function on a time-varying sequence of baroclinic data from MPAS-Ocean (Model for Prediction Across Scales - Ocean) is shown in Figure 7. MPAS-Ocean data is also shown in Figure 1. To assess the impact of the new workflow, and in particular the upgraded features, we spoke again with Dr. Wolfram. The series of baroclinic time steps provide a picture of mixing patterns in the oceans. The ability to see multiple time steps within a single window allowed the scientist to see the progression and characteristics of the baroclinic mixing over time, from different locations and different variables. The scientist spoke of a paper that informed him of an important contour that existed in the data. ColorMoves allowed him to quickly find that contour and visually showed "what took five pages of dense math to describe."

ColorMoves' opacity feature, shown in Figure 7, facilitates greater control over the emphasis of specific regions. More importantly it plays a critical role when looking at volume rendered data. In addition to adding to his ability to see detail in data, the highly interactive nature of ColorMoves is a critical feature to Dr. Wolfram, who stated that being able to quickly apply and adjust color improves his scientific workflow. Communication with peers and a broader audience is also an important aspect of his work, one which is positively impacted by ColorMoves. He commented that "our primary focus is understanding the simulation and thus furthering the science but the communication aspect is just as important. Being able to apply highlights to specific regions of data is particularly important for communication. ColorMoves, enabling us to do it so quickly, has made it practical."

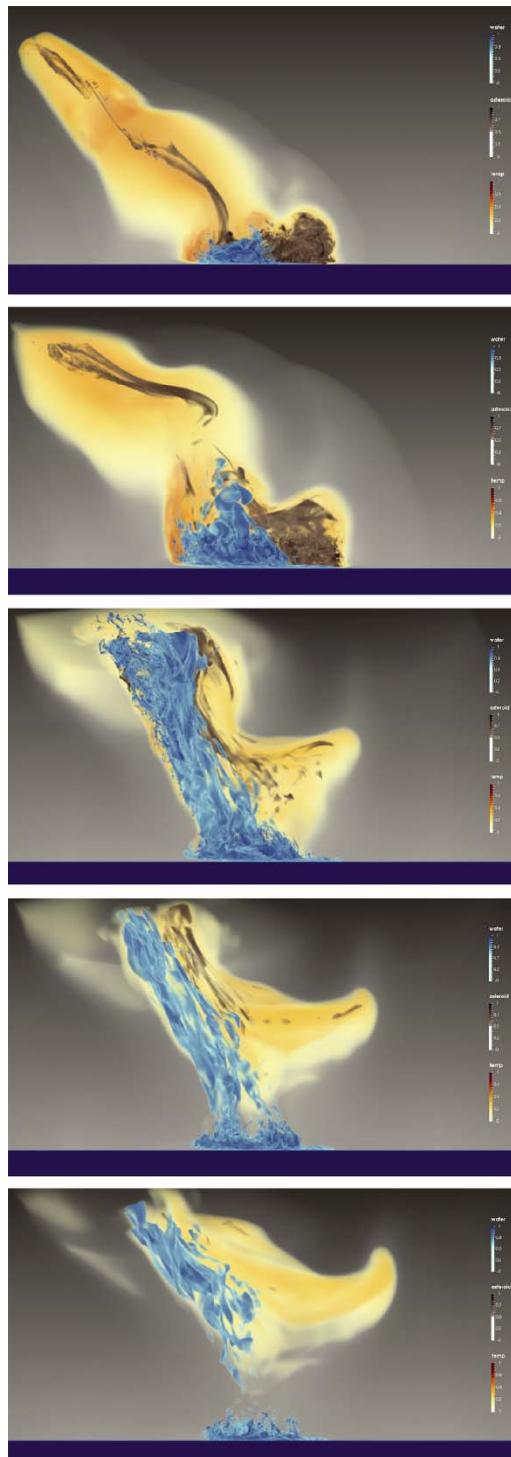
## Neutron Spectroscopy

Neutron spectroscopy data, as seen in Figure 8, has features of interest across many orders of magnitude of intensity. We worked with Dr. Garrett Granroth of Oak Ridge National Laboratory's Neutron Data Analysis and Visualization Division to bring out details within neutron scattering data. While the data analysis in this case is mostly empirical, visualization plays a key role in directing the analysis and communicating results. Using ColorMoves, he was able to distinguish and thus compare the distributions of regions of intensity. Figure 8 illustrates the level of

detail visible using a cool-warm colormap versus one created in ColorMoves. The ease of applying and adjusting color enabled Dr. Granroth to see relationships between the areas of interest that had otherwise not been apparent. The ability to export for use in ParaView, <http://www.paraview.org>, was a critical feature. Seeing the value of ColorMoves, he has shared it with many colleagues, noting that "tools like this make it approachable to get people to move away from default" colormaps.



**Figure 8.** Neutron spectroscopy data often has features of interest on many orders of magnitude of intensity. ColorMoves allowed for a custom color map to emphasize the intensity variation. The ease of adjusting the color encoding expedites the identification of features of interest in the data.



**Figure 9.** A time-varying simulation of a 250-meter asteroid impact in the ocean, rendered with a colormap developed in ColorMoves. The variables are: water fraction (blue), asteroid fraction (brown), and temperature (orange to yellow).

## ASTEROID OCEAN IMPACTS

Galen Gisler of Los Alamos National Laboratory studies asteroid ocean impacts to assist in understanding potential societal and planetary threats. Figure 9 shows a time-varying asteroid ocean impact depicting water vapor, asteroid and air pressure. Dr. Gisler ran an ensemble of simulations, varying asteroid size, angle of entry, and whether the asteroid bursts in the air or on impact with the water. These are used to study how these variables affect such things as temperature, water vapor, asteroid fraction, air pressure as well as the pressure waves below the surface created on impact. Because these are highly interdependent, it is beneficial to render multiple variables in each frame. Here we were able to test a range of combinations over the ensemble of simulations for many time steps until we found a means to render three variables clearly throughout the duration of the simulation.

Figure 9 shows the results of this work—a time series of an asteroid impact, with the three variables—water cell fraction, asteroid cell fraction, and air pressure—rendered in separate color scales.

## CONCLUSION

ColorMoves provides a means for applying artistic color expertise via an interactive, exploratory interface for scientific visualization—an approach that combines the expertise of scientists, artists and computer scientists.

One important contribution of the tool is its immediacy, and how that fundamentally changes a scientist's interaction with data and color. As soon as the scientist drops an image into ColorMoves, a scientific question suddenly appears or an answer suddenly becomes accessible in the data, or a solution to a communications need begins to emerge. For extreme data sizes such as those being generated on today's largest supercomputers, there is a need for tools that allow interactive exploration of data. Color is a critical aspect to that exploration and flexible color-mapping options are a powerful means of achieving this goal.

The ColorMoves code, color scales and supporting material can be found at  
<https://github.com/ascr-ecx/sciviscolor>.

## ACKNOWLEDGEMENTS

We would like to thank James Ahrens and the Data Science at Scale team at LANL as well as all the scientists that provided data and design feedback: G. Gisler, LANL; J. Chen, Sandia; W. Daughton, LANL; G. Granroth, ORNL; M. Petersen, LANL; P. Wolfram, LANL. This work was funded by Lucy Nowell, Program Manager, U.S. Department of Energy Office of Science, Advanced Scientific Computing Research. This publication is released under LANL LA-UR-17-29913.

## REFERENCES

1. L. Zhou and C.D. Hansen, "A Survey of Colormaps in Visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. August, 2016, pp. 2051–2069.
2. C. Ware, *Visualization Thinking for Design*, Morgan Kaufman, 2008.
3. J. Itten and F. Birren, *The Elements of Color: A Treatise on the Color System of Johannes Itten Based on His Book the Art of Color*, Van Nostrand Reinhold Company, 1970.
4. M. Harrower and C.A. Brewer, "ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps," *The Cartographic Journal*, vol. 40, no. 1, 2003, pp. 27–37.
5. S. Mittelstdt et al., "ColorCAT: Guided Design of Colormaps for Combined Analysis Tasks," *Eurographics Conference on Visualization (EuroVis) - Short Papers*, The Eurographics Association, 2015.
6. F. Samsel et al., "Intuitive Colormaps for Environmental Visualization," *Workshop on Visualisation in Environmental Sciences (EnvirVis)*, 2017; doi.org/10.2312/envirvis.20171105.
7. K. Moreland, "Diverging Color Maps for Scientific Visualization," *Proceedings of the 5th International Symposium on Advances in Visual Computing*, 2009, pp. 92–103.
8. J. Itten, *The Art of Color: The Subjective Experience and Objective Rationale of Color*, Van Nostrand Reinhold, 1961.
9. C. Ware, *Information Visualization: Perception for Design*, Morgan Kaufman, 2012.
10. T.-M. Rhyne, *Applying Color Theory to Digital Media and Visualization*, CRC Press, 2016.
11. J. Albers, *The Interaction of Color*, Yale University Press, 2009.
12. F. Samsel et al., "Visualization of ocean currents and eddies in a high-resolution ocean model," *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, 2015.

## ABOUT THE AUTHORS

**Francesca Samsel** is a Research Associate the University of Texas at Austin. Contact her at figs@cat.utexas.edu.

**Sebastian Klaassen** is a graduate student at the University of Vienna. Contact him at sebastian.klaassen@outlook.com.

**David H. Rogers** is Team Lead for the Data Science at Scale team at Los Alamos National Laboratory. Contact him at dhr@lanl.gov.

Contact editor Mike Potel at potel@wildcrest.com.

# Crosswalk Localization from Low Resolution Satellite Images to Assist Visually Impaired People

**Marcelo Cabral Ghilardi**  
PUCRS, Faculdade de Informática

**Julio C.S. Jacques Junior**  
PUCRS, Faculdade de Informática, Universitat de Barcelona, and Universitat Autònoma de Barcelona

**Isabel Harb Manssour**  
PUCRS, Faculdade de Informática

We propose a model for crosswalk detection and localization by using satellite images captured from Google Maps, for the purpose of assisting visually impaired people. The detection is performed by an SVM classifier, which is combined with Google Road Map to speed up computation time and to eliminate some possible false alarms. Experimental results indicate that the proposed model works well in low resolution images, effectively detecting and localizing crosswalks in simulated scenarios.

Automatic localization of crosswalks (also called pedestrian crossing or zebra crossing) is a very important topic of research, with a wide range of applications. One possible application to take into consideration is to assist the mobility of visually impaired people. According to data from the World Health Organization, available at [www.who.int/mediacentre/factsheets/fs282/en](http://www.who.int/mediacentre/factsheets/fs282/en), there are approximately 285 million visually impaired people in the world. Among them, 246 million have low vision (less than 30%), and 39 million are blind. Blind tracks, white canes, and guide dogs are typically used to help visually impaired people to walk outside.<sup>1</sup> However, even with these resources, mobility autonomy in outdoor environments presents a major challenge for them.

Nowadays, as mentioned in the work of Ahmetovic and colleagues,<sup>2</sup> mobile devices have shown a huge potential in supporting people with disabilities. In addition, since these devices are becoming more accessible, a wide number of assistive technologies have been proposed to support people in everyday activities. Following this trend, several solutions for crosswalks' detection

and localization are based on images acquired by smartphones, which may bring some challenges, such as the need to correctly point the camera to the crosswalk<sup>2-4</sup> or the need to turn on a sidewalk to acquire panorama images.<sup>1,5</sup> Besides that, one can take advantage of microphones and other sensors to perform automatic image acquisition,<sup>6</sup> despite being limited to very specific scenarios/conditions (for instance, considering that there are buzzers to indicate go and stop status near traffic lights).

Another possible application for using smartphones, without using their own cameras for image acquisition, is to use their GPS (embedded in many modern smartphones—some of them with good localization accuracy) to initialize the extraction of satellite images (for example, from Google Maps, as in “Path Planning Algorithm Based on Search Algorithm, Edge Detector and GPS Data/Satellite Image for Outdoor Mobile Systems”<sup>7</sup>), applied to crosswalk detection and localization problems. To prevent any misunderstandings, in this article, *crosswalk localization* is considered the relation between the detected crosswalk to the user, whilst *crosswalk detection* is related to its identification/recognition in image coordinates. Such possibility eliminates the need to worry about which direction users must point out their camera/smartphone.

To this end, the main goal of this work is to present a computer vision based model for crosswalk detection and localization, designed to provide spatial guidance to visually impaired pedestrians. In a nutshell, the proposed model can provide information about the presence (or absence) and location of crosswalks near the user. The model is initialized by a GPS coordinate (given by a smartphone, for example), which is used to initialize the image acquisition step. Then, low resolution satellite and road map images are extracted from a public dataset (that is, from Google Maps) and used in the crosswalk’s detection module. The detection is performed by a Support Vector Machine (SVM) framework. Finally, the position of the detected crosswalk is combined with the GPS coordinate to determine its localization in relation to the user, as well as to guide the user to the nearest crosswalk. However, we do not expect to guide the user to the nearest crosswalk with a centimeter precision (due to several factors, such as image acquisition problems, mapping from image coordinate to world coordinate systems, GPS accuracy, and so on), but give him/her an estimate about the nearest crosswalk. This could benefit his/her locomotion in many ways, for instance, showing that the nearest crosswalk is located on their right, at approximately 10 meters.

As stated above, the proposed model is limited to GPS accuracy. As we will address in the next section, some commercial applications (developed to help visually impaired people and based on GPS) also reported such limitation, while emphasizing that GPS can provide good accuracy most of the time. We believe that such technological limitation could be smoothed with dedicated hardware (making the solution more expensive), or even with the natural evolution of GPS technology. In addition, there are several models in the literature that use smartphone cameras to help visual impaired people (described in Related Works), that could be combined with the proposed model to fine-tune our final estimate, for instance, confirming the existence of a crosswalk, or even using our model as start point (since their initialization can be considered a crucial point, as explained next). Despite the GPS accuracy limitation, the proposed model was quantitatively evaluated (see the Experimental Results section) in three stages, as follows: first, we evaluated the accuracy of the crosswalk detection module using a dataset of crosswalk and non-crosswalk patches; during the second stage, the crosswalk detection was evaluated using satellite images extracted from Google Maps and simulated GPS coordinates; finally, the crosswalk localization was evaluated using the output from the previous evaluation. In these three evaluation scenarios, the proposed model demonstrated satisfactory results, effectively detecting and localizing crosswalks in different situations and conditions.

The main contributions of the proposed model are: (1) as far as we know there is no competitive approach that combines low resolution satellite images with a GPS system to help visually impaired people; (2) crosswalks are detected with high accuracy rates (about 96.9%), requiring low computation time (about 497 milliseconds per image); (3) crosswalks are localized with minor user intervention, with high accuracy rates (about 92.7%).

## RELATED WORK

In recent years, several works have been done regarding pedestrian crossing detection and localization to aid visually impaired people to cross the street independently and in a safer way. Many of these works were developed for images captured on the ground (that is, by a smartphone, robot or vehicle).<sup>1-6</sup> As mentioned in the work of Senlet and Elgammal,<sup>8</sup> road, building, and vegetation's detection from aerial images has been extensively studied in remote sensing. However, the use of satellite images combined with GPS data (as in "Path Planning Algorithm Based on Search Algorithm, Edge Detector and GPS Data/Satellite Image for Outdoor Mobile Systems"<sup>7</sup>) seems to be little explored in this context.

Murali and Coughlan<sup>1</sup> presented a model to provide guidance to blind and visually impaired travelers at traffic intersections. In order to estimate the user location in relation to crosswalks in current traffic intersections, a user needs to acquire a panorama image (turning in place on a sidewalk) through an Android application. Given the panorama image, the model reconstructs the aerial view (overhead) of the intersection, creating a traffic intersection's template centered on the location where the user is standing. The generated template is matched against manually segmented traffic intersection regions (built from Google Maps satellite images) to determine the user's current location in the intersection. While in this approach, the user needs to worry about acquiring the panorama image; we propose automatic image acquisition and processing.

Shangguan and colleagues<sup>6</sup> developed a new smartphone application that locates zebra patterns and guides the user along the zebra crossing while crossing the road. In this work, they used the smartphone's microphone to detect when the user is approaching the crossroad, considering that there are buzzers to indicate go and stop status near traffic lights. To avoid user intervention, the smartphone was coupled to the user's white cane, and it takes pictures oriented by short stable periods during each white cane swing cycle (tip tends to rest on the ground for a while before swinging backwards). The crosswalk detection and localization is estimated by grouping parallel lines, which are detected through Hough transformation. One drawback of such approach is that it depends on buzzers that are not always available, especially in developing countries.

Ahmetovic and colleagues<sup>4</sup> presented ZebraRecognizer software library and ZebraLocalizer application. ZebraRecognizer library is responsible for the image processing module, zebra crossing identification, as well as computing the relative position between the observer and the zebra crossing. In this work, crosswalks are detected through a line segment detection and grouping algorithm, combined with a rectification matrix computation method (used to compensate for projections' distortions of extracted features). ZebraLocalizer application acquires images from the camera, sends them to the software library, and implements the interaction paradigm that enables blind users to identify crosswalks, aligned to the best crossing position and safety on the road. Data acquired from smartphone cameras and accelerometers is combined to reach the proposed solution. Their proposed image processing module was improved in "ZebraRecognizer: Efficient and Precise Localization of Pedestrian Crossings,"<sup>2</sup> in terms of performance and accuracy, as well as to cope with different users' height. One limitation of this approach is that the user is responsible for image acquisition through the smartphone.

Regarding methodologies in which satellite images are used, Herumurti and Uemura<sup>9</sup> presented a model for urban road network extraction and zebra crossing detection from very high resolution aerial images, combined with Digital Surface Models (DSM, which gives the elevation of land surface<sup>10</sup>). In this work, the zebra crossing is first detected by a template matching technique and then refined using SURF descriptor/object detector.<sup>11</sup> The output of their zebra crossing detection is used to initialize the urban road extraction module. It is important to mention that the authors did not give details about how the region of interest, associated to the image under analysis, is obtained. This model differs from ours because it is for road network extraction in urban areas using high resolution images (12 megapixels), while we locate crosswalks from low resolution images (0.4 megapixels).

Zidek and Rigasova<sup>7</sup> proposed to combine a search algorithm with satellite image data and edge detection for trajectory planning. They use GPS data to select starting and finishing points and to acquire the actual position of the device in relation to satellite maps. The context of this work is not related to crosswalk detection/localization, but it is mentioned to motivate the use of such combination: GPS data used to initialize satellite image acquisition.

In the work of Senlet and Elgammal,<sup>8</sup> a framework to construct sidewalk and crosswalk maps from satellite images is proposed. They used a crosswalk detection algorithm to complete the connectivity of a previously estimated sidewalk map. The crosswalk is detected by a template matching scheme, using synthetic templates with different frequencies and angle orientations, followed by a global segmentation thresholding and region labeling. In order to increase computation performance, as well as to decrease possible false alarms, the detection is performed on the road or near road regions. The authors mentioned that their model achieved about 83% of precision and 80% of recall to path segmentation, but they did not report any evaluation on crosswalk segmentation.

In addition, some commercial applications have been proposed for the purpose of helping visually impaired people, such as Nearby Explorer (US\$99.00), Intersection Explorer, GetThere GPS, and Sendero GPS LookAround, which are free for download. Nearby Explorer enables independent and informative trips for visually impaired people. It not only shows surroundings and approaching streets and businesses, but also offers continually updating distance and directional information of the nearest or selected location. They mentioned that GPS provides good accuracy within a few yards from a person's actual position under optimal conditions. It means that it can be expected to achieve accuracy (most of the time) good enough to determine on which side of the street the traveler is. Intersection Explorer, offered by Google and maintained by Eyes-free group, speaks the layout of streets and intersections in neighborhoods as the user touches and drags his/her finger around the map. This helps visually impaired people to understand a neighborhood before venturing out and while on the go. In this app, as mentioned by the authors, the best accuracy achieved from the GPS was around 3 meters. Android app GetThere is offered by Lew Lasher, and it was designed specifically for blind and visually impaired people. It performs several spoken interactions with the user, providing various types of navigational information, such as the intersection that he/she will come across as he/she is walking and the distance to the next cross-street. GetThere offers navigation in most countries in the world, based on coverage from OpenStreetMap. They mentioned that the GPS reception is usually accurate enough to determine which side of the street the user is. Sendero GPS LookAround, developed for iPhone, announces the current street (with Voice Over), city, cross-street, heading and nearby points of interest. The highest quality data mentioned by the authors is in North America and Europe, but it is also available in other countries.

The aforementioned approaches have different solutions for the problem of crosswalk detection and localization: some of them require the user to take photos of the environment,<sup>1</sup> which can be considered a great challenge for visually impaired people; others need very high resolution aerial images<sup>9</sup> or are based on buzzers to indicate go and stop status<sup>6</sup> in traffic lights, which are usually not available in developing countries. Moreover, it is also desirable to locate crosswalks that are in the middle of the block and not on street corners. The mentioned commercial apps can provide spatial guidance for the user, combining GPS and spatial information from public maps, but the main difference when compared to the proposed model is that none of them propose to detect and locate crosswalks near the user. Furthermore, one of the limitations of these apps is that information returned to the user, such as businesses, banks, surrounding and approaching streets, must be previously set in a dataset, and even though they can identify the distance to the next cross-street, they do not identify crosswalks.

In the proposed model, low resolution satellite and road map images are combined with GPS coordinates for crosswalk detection and localization. The model does not depend on buzzers, the user does not need to take photos, and the location of crosswalks does not need to be previously set in a dataset. In addition, as reported in the work of Ahmetovic and colleagues,<sup>2</sup> although a universally accepted definition of crosswalk does not exist, many of currently adopted crosswalk standards are very similar. Considering that, the proposed approach could be easily adapted to other standards.

## THE MODEL

This section describes in detail the proposed model for crosswalk detection and localization. It can be divided in three main parts: (1) an Android app to get GPS information and give voice

feedback; (2) CrossLib, related to the kernel of the proposed model; and (3) a web service that provides exchange of messages between the app and CrossLib. An overview of the proposed approach is illustrated in Figure 1.

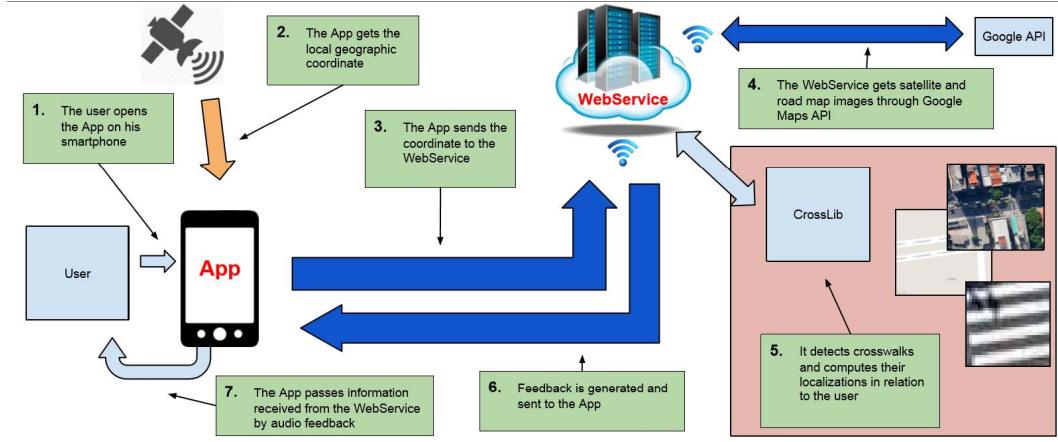


Figure 1. Overview of the proposed model.

One of the first steps of the proposed model is related to GPS coordinate acquisition. GPS coordinates are used to initialize the image extraction procedure and to make a relation between users and detected crosswalks (that is, crosswalk localization). The geographic coordinates, captured by the user's smartphone, is sent to a web service. The web service is responsible for receiving the geographic coordinates from the user (illustrated in Figure 2a by a red dot) and invokes CrossLib to initialize image extraction. To this end, Google Maps API is used to get the satellite image (Figure 2b) and the respective road map (Figure 2c). These images will be used by CrossLib during the crosswalk detection procedure (described later). The output of the crosswalk detection stage is a list of possible candidates. Such list is sent to the crosswalk localization module (described later), which gives to the user, by audio feedback, the nearest detected crosswalk location (or some other options, as described later). Next, the main steps related to CrossLib are detailed.



Figure 2. Images acquisition: (a) GPS coordinates sent by the user, illustrated by the red dot; (b) extracted satellite image from Google Maps; (c) extracted road map image from the same dataset.

## Image processing module

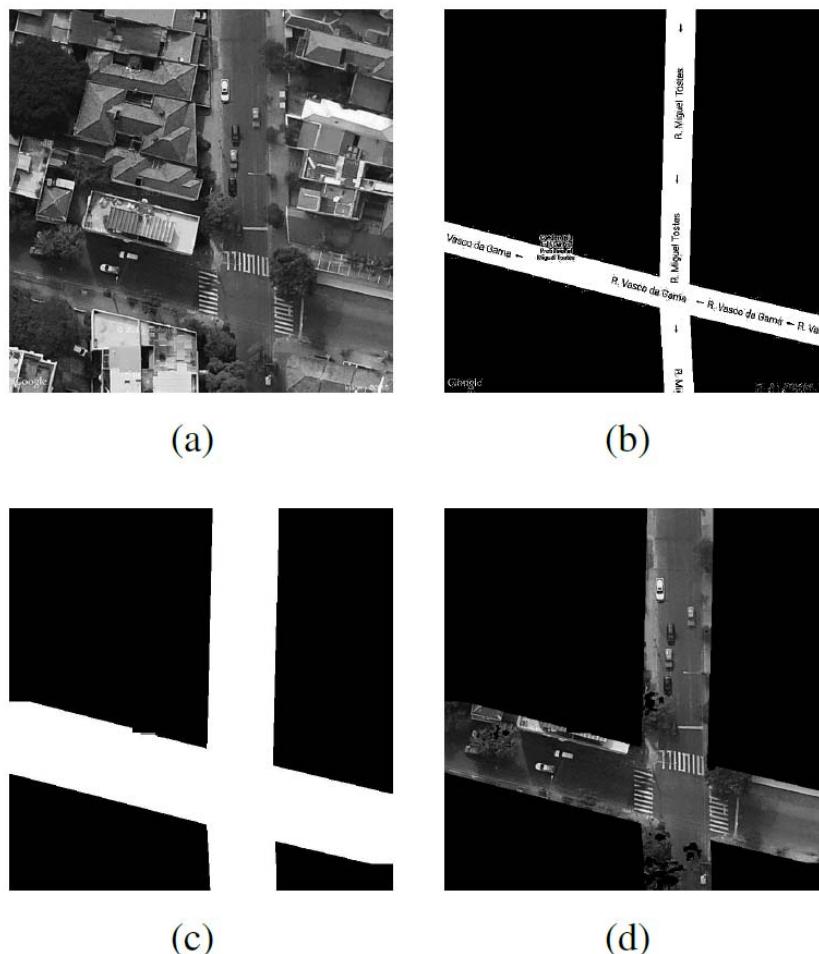
The image processing module is responsible for extracting images from Google Maps, given a reference coordinate (illustrated in Figure 2a by a red dot), as well as guiding crosswalk detection. As we are using a free version of Google Maps API, each obtained image (satellite and road

images) is a matrix with  $640 \times 640$  size of resolution, captured by a zoom of  $20\times$  (that is, the image height from the ground), corresponding to an area of  $6.890\text{m}^2$ , approximately.

The first step of the image processing module (after image acquisition) relates to image segmentation, which consists of defining a region of interest in which crosswalks could be included (on the road or near road regions). The purpose of such approach is to improve computational performance as well as to eliminate the possibility of finding false positives in unrelated areas (for instance, tops of buildings, roofs, and so on).

Initially, both the input satellite image and road image (assigned to R) are converted to grayscale (illustrated in Figure 3a). The region of interest B is generated by a simple thresholding approach. The segmentation is obtained by thresholding each pixel  $(x,y)$  of R by  $\lambda$  (where  $\lambda = 245$ , set experimentally), as defined in Equation 1. Figure 3b illustrates the output of the thresholding approach.

$$B(x,y) = \begin{cases} 0, & \text{if } R(x,y) \leq \lambda \\ 1, & \text{otherwise} \end{cases} . \quad (1)$$



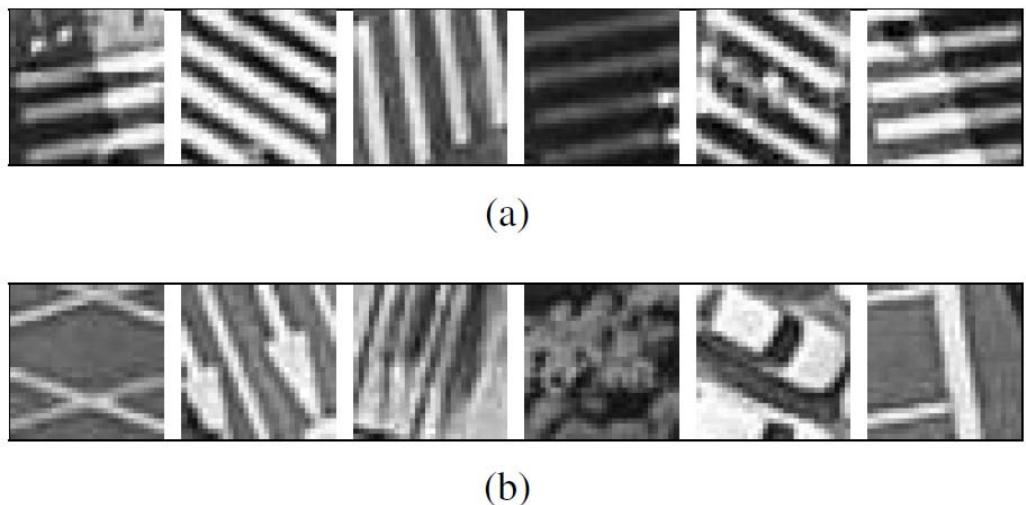
**Figure 3. Image segmentation:** (a) input satellite image (illustrated in Figure 2b) converted to grayscale; (b) initial segmentation of the road map (shown in Figure 2c); (c) output of morphological operations; (d) combining the satellite image with the region of interest.

As we can see in Figure 3b, there are several undesirable structures in the binary image (text, arrows, and so forth). By trying to eliminate such undesirable structures, a post-processing morphological operation was applied to this binary image (a combination of erosions and dilations). More precisely, we first applied 2 erosions in image B, with a squared structuring element with size  $3 \times 3$ . The resulted (eroded) image was then dilated 20 times, using the same structuring element. The goal of such procedure was to remove small artifacts found in the binary image and extend the area of interest to deal with misalignment of the road map (regarding the satellite image). The order, number, and type of morphological operations were set experimentally (these choices are strongly related to the adopted image resolution and zoom, and they can be easily changed to deal with different setups).

The output of the morphological operations, illustrated in Figure 3c, was then combined with the respective grayscale satellite image (Figure 3d). The procedure described above was used in the next stages of our model for crosswalk detection, as described next.

## SVM classifier for crosswalk detection

In this work, we propose to train a Support Vector Machine (SVM) classifier to detect crosswalks from satellite images. To this end, we built a dataset containing several positive and negative sample patches (with size  $30 \times 30$  pixels in each patch), used for cross-validation. This dataset is composed of 370 patches of crosswalks (positive samples) and 530 patches of non-crosswalks (negative samples), illustrated in Figure 4, which were manually extracted from Google Maps using its API (considering grayscale input images with  $640 \times 640$  of resolution, captured by a zoom of  $20\times$ ). As we can see in Figure 4, image patches vary according to illumination conditions (for example, captured at different times of the day, with or without shadows), and angle orientation and shapes, regarding the zebra pattern. In addition, some positive image patches are partially occluded by people, cars, and/or trees, while some negative image patches contain structured objects with parallel lines, such as directional arrows. The idea of including such features in the learning dataset was to deal with different situations/scenarios we usually find in real applications. Next, the dataset was divided into training and test sets, with 600 and 300 independent patches, respectively, on which a 10-fold cross-validation was used to optimize (hyper)parameters of SVM (described later).



**Figure 4. Manually extracted patches of (a) positive and (b) negative samples.**

We used the Local Binary Pattern (LBP) feature extraction method to train our SVM. As reported in the work of Dixit and Hegde,<sup>12</sup> the LBP texture operator was introduced as a complementary measure for local image contrast. It can be used to recognize a wide variety of texture types, in which statistical and structural methods have conventionally been used separately. We

have also tried to use the well-known Gray Level Co-occurrence Matrix (GLCM)<sup>12</sup> feature extraction, and a combination of both. Table 1 shows the evaluated methods for feature extraction under development of the proposed model.

Table 1. EVALUATED METHODS FOR FEATURE EXTRACTION.

| Feature Extraction Method       | Descriptor  |
|---------------------------------|---|
| Gray Level Co-occurrence Matrix | Contrast<br>Energy<br>Entropy<br>Homogeneity                        |
| Local Binary Pattern            | Black & White Symmetry<br>Geometric Symmetry<br>Degree of Direction |

## Crosswalk detection method

After training the SVM classifier, the next stage of the proposed model was crosswalk detection. In a practical situation, we consider an input satellite image (640×640 size of resolution), delimited by the region of interest, as illustrated in Figure 3d. This input image is divided into small cells, with size 15×15, as shown in Figure 5a (illustrated by a cropped region, highlighted in Figure 5d, for visualization purposes).

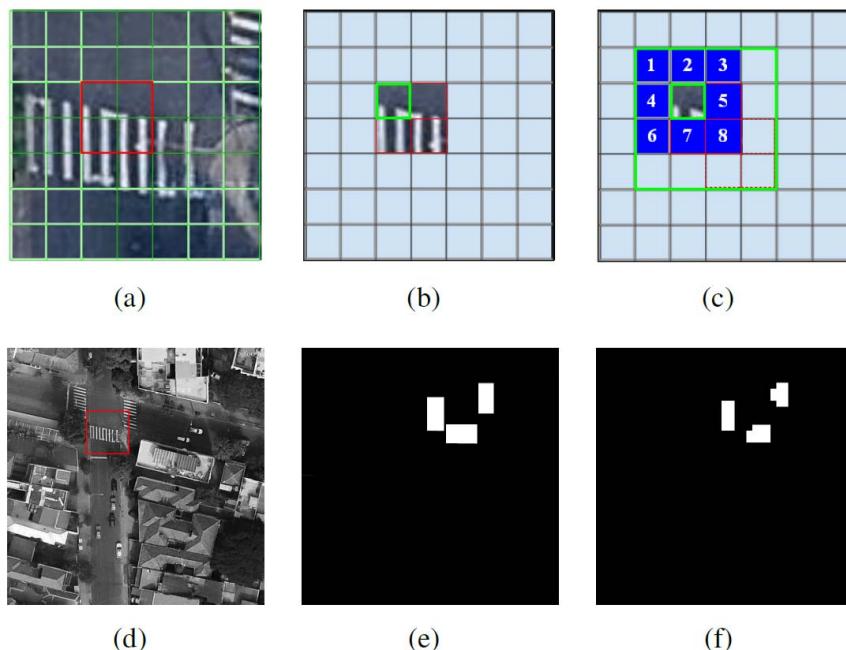


Figure 5. Crosswalk detection illustration: (a) cropped region of an image under analysis (Figure 5d) with a 30×30 reference patch marked by a red square; (b) reference patch and its 15×15 reference cell highlighted by a green square; (c) reference cells of eight patches used in the second verification; (d) input satellite image (with cropped region illustrated in the image (a) highlighted by a red square); (e) ground truth manually informed by the user; (f) final estimation, given by the proposed model.

During the classification stage, all of these  $15 \times 15$  cells, which fall in the region of interest, are visited. Then, each group of four neighboring cells (that satisfies region of interest's criteria) will form a patch with size  $30 \times 30$  (related to the same patch size used to train our SVM classifier), illustrated in Figure 5a by a red square. Each patch is composed by the reference cell (the upper left cell, illustrated in Figure 5b by a green square) and its neighboring cells (that is, the right cell, the bottom one, and the bottom right one). The goal of such procedure was to create an overlapping region between adjacent patches to deal with misalignment in the classification stage, as well as to perform a two-pass verification, as described next.

If a  $30 \times 30$  reference patch is considered a crosswalk (for instance, the patch generated from the reference cell illustrated in Figure 5b), a second verification is performed. In this second verification, eight neighboring patches ( $30 \times 30$ ) are created around the reference patch, considering their neighboring cells (the reference cells of these eight patches are illustrated in Figure 5c by blue squares). If at least one of these eight neighboring patches is also considered a crosswalk, the reference patch is then set as a crosswalk; otherwise the reference patch is discarded.

Figure 5e illustrates regions of crosswalks manually informed by a user (used as ground truth), whilst Figure 5f illustrates the overall output for a given image (Figure 5d), in which patches are detected as crosswalk (by our two-pass verification), illustrated by white blocks.

## Crosswalk localization method

Given detected crosswalks, the next stage of the proposed model relates to making a spatial relation between the user and crosswalks (that is, crosswalk localization). First, we assume the user would always send his/her GPS coordinate when facing a street and, by doing so, we can easily estimate which street he/she is by analyzing the binary image (Figure 3c) and his/her informed position (illustrated in Figure 6 by a red dot). If the user is at a corner, we assume that he/she is facing toward the corner.

Following the estimated street line in both directions (left and right), as shown in Figure 6, and also after analyzing the binary image, we can find corner intersections when facing another street (if that exists), and consequently, define the block where the user is positioned. Such information is used to discard detected crosswalks that are not connected to the block where the user is. Corner detection could also be used to give the user additional feedback (for example, when crosswalks are not detected in a block, the model could inform the user about his/her location in relation to the nearest corner).

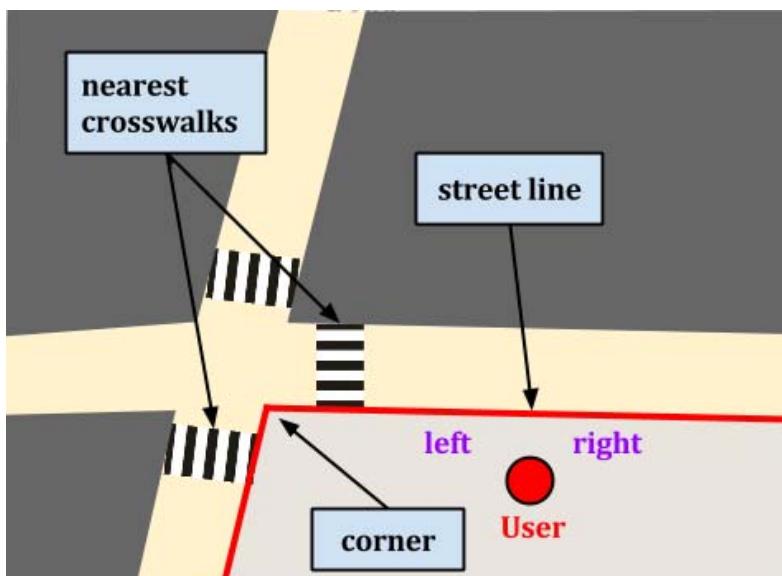


Figure 6. Illustration of street lines, corners, blocks, and crosswalks connected to the block in which the user is. All information is used in the crosswalk localization module.

Our model is developed to inform the user about the nearest crosswalk (or the two nearest in specific situations—if they are approximately at the same distance from the user). The idea is to avoid giving too much information that could confuse the user, but at the same time give him/her choices when facing two crosswalks at the same distance. To locate the nearest crosswalk (or the two nearest ones), two different situations are considered: i) the user is positioned at the corner, and ii) the user is facing the street (for example, in the middle of the block). The first case is defined just to create a reference point (that is, the corner point, or the exact street intersection point), used as a reference when informing the user the crosswalk localization (for example, there is a crosswalk on your left).

Thus, if two or more crosswalks are detected, we first compute the distance from the user to each of them, following the street line edges (to take into consideration the corners). The distance from the user to the nearest crosswalk (or corner) is computed using a very simple relation, which takes into consideration the image resolution and its real dimension, extracted from Google Maps API. Since we know that each image has 640-pixel width, related to 82.69 meters, we can assume each pixel represents 0.129203125 meters. In case two or more crosswalks have approximately the same distance in relation to the user, he/she is informed about the two nearest ones. We defined in our experiments that two crosswalks have similar distances in relation to the user when their distances differ in less than 3 meters (being a little flexible, regarding GPS accuracy). This procedure is described in Algorithm 1. Such information is used to compute the final feedback, which is sent to the user, as explained next.

```

1:  $N_c$  is the number of detected crosswalks in a block
2:  $U_p$  is the user position
3: if  $N_c = 1$  then
4:   inform the user its location
5: end if
6: if  $N_c > 1$  then
7:   Compute the distance from the user to each of them
8:   if they have different distances then
9:     inform the user the nearest crosswalk
10:    else
11:      inform the user the two nearest crosswalks
12:    end if
13: end if
```

**Algorithm 1.** Crosswalk localization.

## App feedback

The proposed app provides voice feedback to the user. In order to simplify feedback, we consider the user can be facing the corner or the street, as previously mentioned. In this way, regardless of the user's position in relation to cardinal directions, feedback is provided in relation to his/her right and/or left. The provided feedback informs the distance to the nearest crosswalk (or crosswalks) on the block the user is positioned (considering only the street in front of the user and those that intersect it, which means that if the whole block is visible in the satellite image, we can discard crosswalks located on streets far away from the user). Figure 7 exemplifies the points that can generate feedback, and Table 2 describes possible feedback. The values presented in Table 2, that is, “10 meters” and “5 meters,” are used for illustration purposes, since computed distances (in meters) may vary according to each situation. Information about “turning the corner” is provided only when the user is facing the street line (that is, in the middle of the block), and the nearest crosswalk is positioned after turning the corner.

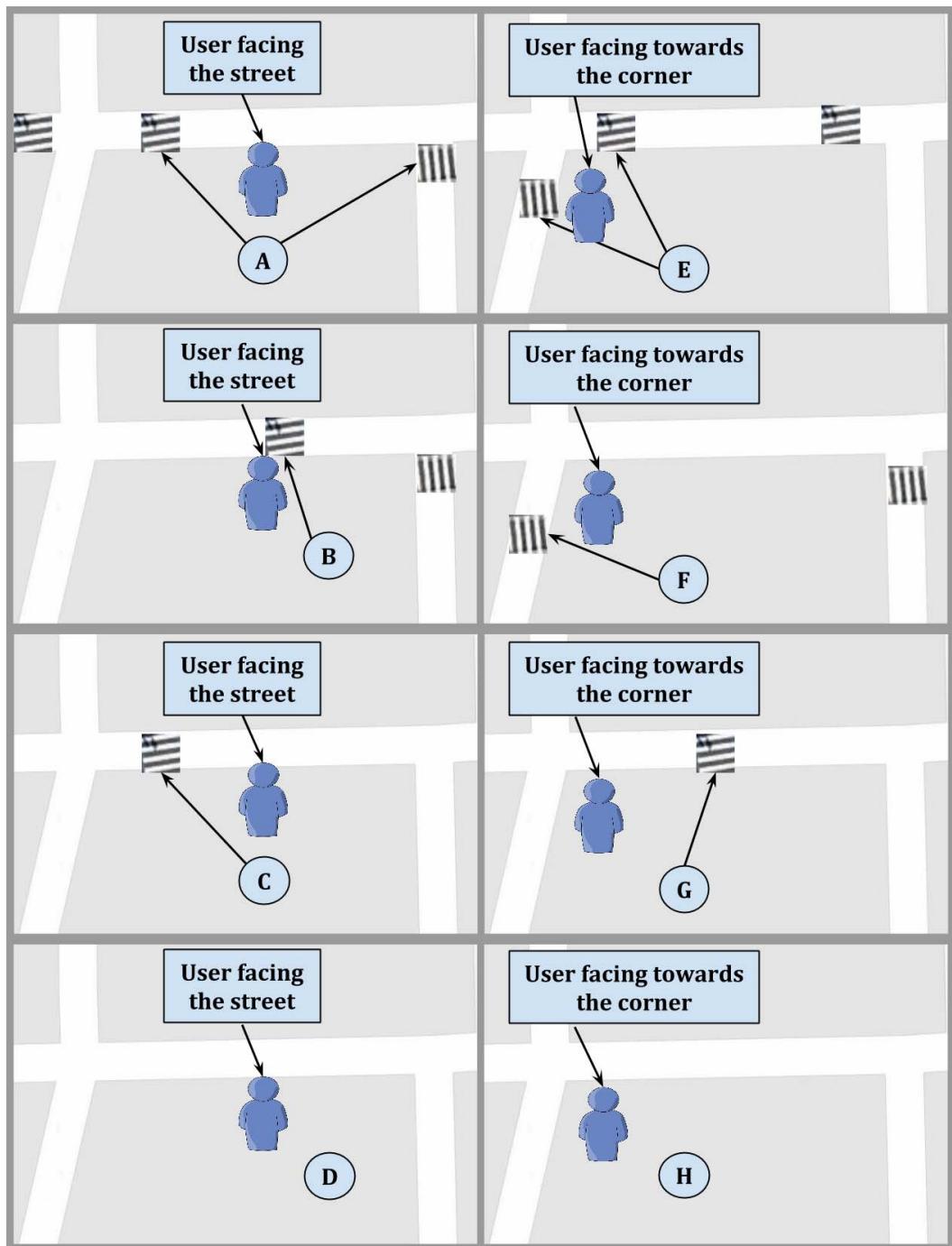


Figure 7. Examples of possible feedback.

**Table 2. FEEDBACK EXAMPLES, ILLUSTRATED IN FIGURE 7.**

|   |   |
|---|---|
| A | There are two crosswalks at 10 meters, one on your left and another on your right, just after turning the corner. |
| B | There is a crosswalk in front of you.   |
| C | There is a crosswalk at 10 meters on your left.   |
| D | There is no crosswalk in the block you are.   |
| E | You are at the corner and there are two crosswalks at 4 meters, one on your left and another on your right.       |
| F | You are at the corner and there is a crosswalk at 5 meters on your left.  |
| G | You are at the corner and there is a crosswalk at 5 meters on your right.   |
| H | You are at the corner and there is no crosswalk in the block you are.   |

## EXPERIMENTAL RESULTS

In this section, we illustrate some experimental results obtained by the proposed model. The experiments are divided in three case studies, as follows: i) feature evaluation, used to choose the best feature extraction method for texture analysis (regarding the evaluated ones); ii) crosswalk detection, used to evaluate the accuracy of the SVM classifier; and iii) crosswalk localization, which is used to evaluate accuracy of the application feedback.

### Feature evaluation

This section describes how LBP was chosen as feature extraction method. Different methods were evaluated (as previously described in Table 2), as well as some combinations of them. The experimental results are reported below. First, our database of image patches, previously described (containing a total of 900 image patches), was randomly divided in a stratified way into a training set (containing 600 samples) and an independent test set (with 300 samples). The training set was used in a 10-fold cross-validation to assess the performance of SVM for different hyper-parameters. Once the optimal hyper-parameters were found, a new model was trained using the complete training set and applied to the independent test set. The evaluation was performed in terms of sensitivity, specificity, and accuracy, defined in Equations 2, 3 and 4, respectively.

$$Sensitivity = \frac{TP}{TP + FN}, \quad (2)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4)$$

where TP, TN, FP, and FN refer to True Positive, True Negative, False Positive, and False Negative, respectively.

Considering the Gray Level Co-occurrence Matrix (GLCM) feature extraction method,<sup>12</sup> three variables must be considered: direction and distance from neighboring pixel and number of gray-scale tones. Since crosswalks can assume different directions in satellite images, we carried out the experiment using four directions (0, 45, 90, and 135 degrees), ranging the distance parameter from 1 to 4, with grayscale tones from 2<sup>1</sup> to 2<sup>7</sup>. As illustrated in Figure 8, the best result related to GLCM feature extraction (90.3% accuracy) was obtained when using distance=1 and gray-scale tones=2.

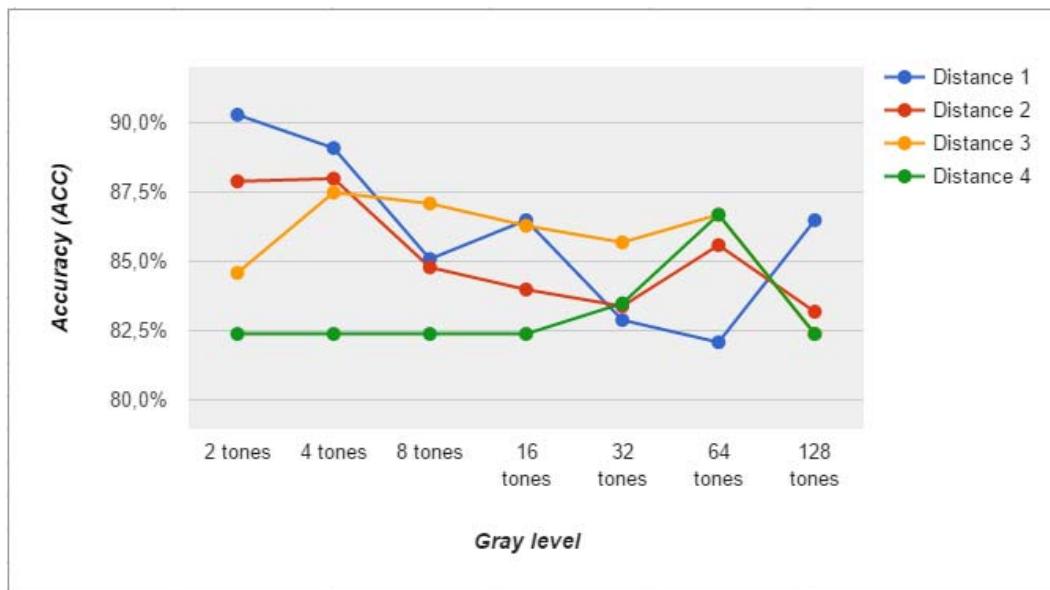


Figure 8. Accuracy evaluation of Gray Level Co-occurrence Matrix (GLCM) method.

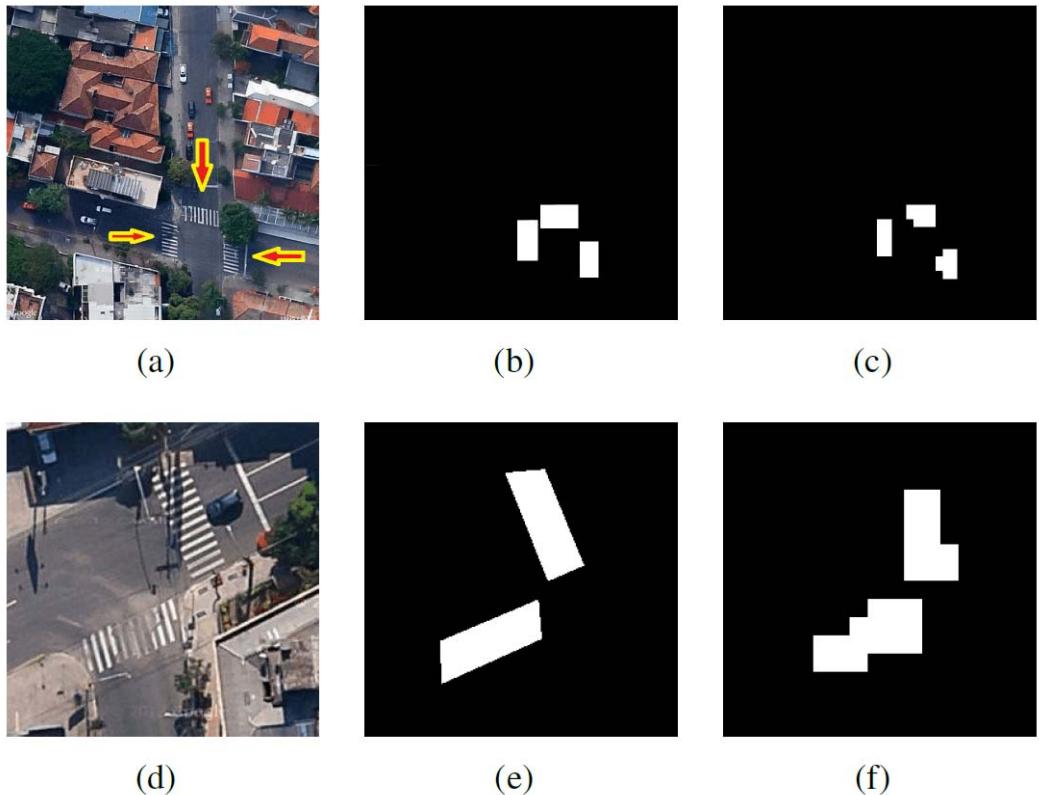
On the other hand, as we can see in Table 3, the overall best result was obtained by LBP method, reaching about 94.6% of accuracy in crosswalk detection. In LBP method, a neighborhood of  $3 \times 3$  pixels was considered, as well as the following three measures: Black & White Symmetry (BWS), which measures the symmetry between the left half and right half of grayscale's histogram; Geometric Symmetry (GS), which measures the regularity of texture form; and Degree of Direction (DD), which measures the degree of texture linearity. The other measures usually used by LBP method were not considered because they are related to texture direction extraction (and crosswalks can assume different orientations).

Table 3. COMPARISON AMONG FEATURE EXTRACTION METHODS.

| Features   | Sensitivity | Specificity | Accuracy |
|------------|-------------|-------------|----------|
| LBP        | 95.7%       | 93.9%       | 94.6%    |
| GLCM + LBP | 86.8%       | 93.6%       | 90.9%    |
| GLCM       | 89.0%       | 91.1%       | 90.3%    |

## Crosswalk detection evaluation

In this section, we describe the evaluation of the proposed crosswalk detection method in a practical situation. In this case study, several satellite images (to be more specific, 100 images) were extracted from Google Maps (considering grayscale input images with  $640 \times 640$  of resolution, captured by a zoom of  $20\times$ ), taking into account the input GPS coordinate given by the user. To simulate the GPS coordinate informed by the user, we randomly selected  $N$  (where  $N = 100$ ) coordinate positions using Google Maps API (all coordinates were extracted from locations close to streets, simulating previously mentioned conditions that the user is facing the street or a corner). For each extracted satellite image, a ground truth was manually generated by the user. The ground truth (Figure 9b) of each satellite image (Figure 9a) is defined by a binary image (with the same size) with crosswalk regions delimited by almost rectangular boxes. Such information is used for quantitative evaluation. Figure 9c illustrates the output binary image of the proposed model for a given situation, and thus we can verify how accurate the model is.



**Figure 9. Illustration of extracted satellite image in (a) and a badly painted case (zoomed for visualization purposes) in (d) with their ground truth data in (b) and (e), respectively. Images (c) and (f) illustrate the respective output binary images of the proposed model for crosswalk detection. The lower row shows a case in which the crosswalk in a bad painting condition brought some additional challenges to the crosswalk detection module.**

It is important to mention that, in this experiment, images were extracted from some state capitals of Brazil, such as São Paulo (SP), Florianópolis (SC), and Porto Alegre (RS). This choice brings some challenges. For instance, as Brazil is considered a developing country, the quality of some crosswalks can present some variations (as visually observed) regarding their painting quality (for instance, smoothed or with minor lacks); Figure 9d illustrates such situation. In addition, the proposed model is not developed to work only in such cities of Brazil, as the zebra crossing pattern is adopted in many cities around the world. However, we noticed that there will be variation regarding crosswalk patterns depending on where we are, affecting the performance of the proposed model. Figure 10 illustrates such problem, caused by different kinds of patterns found around the world. We argue that such problem could be easily addressed through different training sets.



(a) Miami, USA      (b) Phoenix, USA      (c) Tokyo, Japan

**Figure 10.** Example of different crosswalk patterns, which could affect the classifier's performance, extracted from different cities around the world: (a) Miami, USA; (b) Phoenix, USA; (c) Tokyo, Japan.

Each ground truth image (for example, Figure 9b) is confronted with the estimation given by the proposed model (for example, Figure 9c). The comparison is made in the level of patches instead of pixels, considering that ground truth images (as well as output binary ones) are discretized by patches with  $30 \times 30$  of size. We defined a True Positive patch when a ground truth patch (marked as a crosswalk region) and the output binary image patch at same location (estimated as a crosswalk region) have more than 10% of intersection. A False Positive patch is defined when the output binary image patch fails in a non-crosswalk region of the ground truth patch more than 10% of its area. True Negative and False Negative are defined similarly.

In this experiment, as shown in Table 4, the proposed model achieved an average accuracy of 96.9%, with standard deviation of 2.841, which we consider as a satisfactory accuracy rate. The average computational cost to process each image was 497 milliseconds, with standard deviation of 244, using an Intel i5 Processor, 2.27GHz and 4Gb of memory (implemented using C#.NET framework 4.5 and AForge.NET framework).

**Table 4. CROSSWALK DETECTION EVALUATION.**

| Sensitivity | Specificity | Accuracy |
|-------------|-------------|----------|
| 87.5%       | 97.8%       | 96.9%    |

It is important to mention that we are not proposing a completely novel method, since SVM and LBP feature extraction are broadly used as solutions for many classification problem. However, the proposed model achieved satisfactory performance considering specificity and accuracy values, and the novelty is based on the usability compared to existing systems, since the user does not have to worry about acquiring images. In addition, the processing was carried out automatically and with low resolution images. Because of that, it was not possible to do a direct performance evaluation comparison between our approach and other models and apps.

## Crosswalk localization evaluation

In this section, we evaluate the accuracy of the provided feedback related to the crosswalk localization module. To do so, we randomly chose a set of 100 geographical coordinates (simulating users' input), representing a wide number of situations, that is, areas with a crosswalk in front of the "user," on his/her right/left, as well as areas without crosswalks and areas with badly painted crosswalks or with partial occlusions. In this experiment, we considered that coordinates were always obtained from a sidewalk region, in which 75% of them were extracted simulating the user positioned facing the street line, whereas the other 25% simulated the user positioned at the corner.

In a second stage, we created a ground truth data associated to expected feedback for each extracted image (by visual inspection and according to Table 2, for example, “there is a crosswalk at 7 meters on your right”). In this case, we considered feedback as positive (true) when the model returned the corresponding feedback, and negative (false) otherwise. It is important to emphasize that the relative distance, measured from the user to the nearest crosswalk, was alleviated in this experiment (it means that small variations were allowed within 3m, if corresponding directions match). In addition, if there were two crosswalks at the same distance, and the provided feedback returns that there is only one crosswalk, this was considered a false result.

As shown in Table 5, the proposed model provided the expected feedback in 92.7% of the simulated cases with an average specificity of 95%. We consider these results very promising, since they include proper localization of corners and crosswalks. Lower accuracy values were observed in situations in which crosswalks were badly painted or under partial occlusions.

**Table 5. CROSSWALK LOCALIZATION EVALUATION.**

| Sensitivity | Specificity | Accuracy |
|-------------|-------------|----------|
| 91.5%       | 95.0%       | 92.7%    |

## CONCLUSION

In this work, we proposed a model for crosswalk detection and localization using satellite and road map images captured from low resolution public datasets (that is, Google Maps). The image extraction was initialized by GPS coordinates, which can be sent by the user through his/her smartphone. Crosswalks were detected by an SVM classifier and proper feedback was sent to the user based on predefined rules. The proposed model required low computational cost and had minor user intervention.

The proposed model was quantitatively evaluated (regarding crosswalks’ detection and localization procedures), using a dataset containing satellite images extracted from Google Maps and simulated GPS coordinates. Experimental results indicated that the model effectively detected crosswalks in 96.9% of cases, achieving about 92.7% of accuracy in relation to their localization. We believe the proposed model can be used to improve the quality of life of visually impaired people, helping them to get around in outdoor environments in an easier and safer way.

For future work, we intend to use other resources available on most smartphones, such as the compass and/or accelerometer, in order to increase crosswalk localization accuracy, as well as to develop a case study with visually impaired people to evaluate the real applicability of the proposed model.

## ACKNOWLEDGMENTS

The authors would like to thank Brazilian agencies FAPERGS/CAPES, Hewlett-Packard Brasil Ltda. and PUCRS for the financial support.

## REFERENCES

1. V. Murali and J. Coughlan, “Smartphone-based crosswalk detection and localization for visually impaired pedestrians,” *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–7.

2. D. Ahmetovic et al., "ZebraRecognizer: Efficient and Precise Localization of Pedestrian Crossings," *22nd International Conference on Pattern Recognition (ICPR)*, 2014, pp. 2566–2571.
3. X. Kou, Y. Wei, and M. Lee, "Vision based guide-dog robot system for visually impaired in urban system," *13th International Conference on Control, Automation and Systems (ICCAS)*, 2013, pp. 130–135.
4. D. Ahmetovic, C. Bernareggi, and S. Mascetti, "Zebralocalizer: Identification and localization of pedestrian crossings," *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)*, 2011, pp. 275–284.
5. G. Fusco et al., "Determining a blind pedestrian's location and orientation at traffic intersections," *Computers Helping People with Special Needs*, Springer, 2014.
6. L. Shangguan et al., "Crossnavi: Enabling real-time crossroad navigation for the blind with commodity phones," *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2014, pp. 787–798.
7. K. Zidek and E. Rigasova, "Path Planning Algorithm Based on Search Algorithm, Edge Detector and GPS Data/Satellite Image for Outdoor Mobile Systems," *IEEE 10th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2012, pp. 349–354.
8. T. Senlet and A. Elgammal, "Segmentation of occluded sidewalks in satellite images," *21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 805–808.
9. D. Herumurti et al., "Urban road network extraction based on zebra crossing detection from a very high resolution rgb aerial image and dsm data," *2013 International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, 2013, pp. 79–84.
10. T. Toutin and L. Gray, "State-of-the-art of elevation extraction from satellite SAR data," *Journal of Photogrammetry and Remote Sensing*, vol. 55, no. 1, 2000, pp. 13–33.
11. H. Bay et al., "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, 2008, pp. 346–359.
12. A. Dixit and N. Hegde, "Image texture analysis - survey," *Third International Conference on Advanced Computing and Communication Technologies (ACCT)*, 2013, pp. 69–76.

## ABOUT THE AUTHORS

**Marcelo Cabral Ghilardi** is a master's student in the School of Computer Science at the Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil. His research interests include all areas of computer vision, computer graphics, accessibility, and human-computer interaction. Ghilardi has a graduate degree in applied technologies for information systems from University Center Ritter dos Reis. Contact him at marcelo.ghilardi@acad.puers.br.

**Julio C.S. Jacques Junior** is a postdoctoral researcher in the Department of Mathematics and Informatics at the Universitat de Barcelona, Spain. His research interests include human behavior analysis, image segmentation, crowd analysis, object detection, and tracking. Jacques Junior has a PhD in computer science from Pontifical Catholic University of Rio Grande do Sul, Brazil. Contact him at julio.jacques@acad.pucrs.br.

**Isabel Harb Manssour** is an associate professor in the School of Computer Science at the Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil. Her research interests include all areas of visualization, computer graphics, human-computer interaction, and computer vision. Manssour has a PhD in computer science from the Federal University of Rio Grande do Sul. Contact her at Isabel.manssour@puers.br.

\

# Objective Quality Assessment and Perceptual Compression of Screen Content Images

**Shiqi Wang**  
City University of Hong Kong

**Ke Gu**  
Beijing University of Technology

**Kai Zeng**  
University of Waterloo

**Zhou Wang**  
University of Waterloo

**Weisi Lin**  
Nanyang Technological University

Screen content image (SCI) has recently emerged as an active topic due to the rapidly increasing demand in many graphically rich services such as wireless displays and virtual desktops. SCIs are often composed of pictorial regions and computer generated textual/graphical content, which exhibit different statistical properties that often lead to different viewer behaviors. Inspired by this, we propose an objective quality assessment approach for SCIs that incorporates both visual field adaptation and information content weighting into structural similarity based local quality assessment.

Furthermore, we develop a perceptual screen content coding scheme based on the newly proposed quality assessment measure, targeting at further improving the SCI compression performance. Experimental results show that the proposed quality assessment method not only better predicts the perceptual quality of SCIs, but also demonstrates great potentials in the design of perceptually optimal SCI compression schemes.<sup>1</sup>

Recently, there has been an increasing demand to enable thin-clients to enjoy the computationally intensive and graphically rich services by instantly transmitting the complicated graphical interfaces to the clients. Such time variant interface can be rendered as a screen content image (SCI), which is a mixture of pictorial and computer generated textual/graphical regions. The

quality of the SCIs directly determines the user experience of the screen remoting system. Therefore, an image quality assessment (IQA) model that can predict the perceptual quality of SCIs is desirable, which serves as a benchmark for monitoring, adjusting and optimizing the performance of the screen remoting systems.

In the past decades, there has been significant progress in the field of objective IQA.<sup>1,2,3,4,5</sup> However, most existing methods are designed and validated based on natural images, which do not always share the same properties of SCIs. Typically, the discontinuous-tone computer generated image is featured by repeated patterns, sharp edges and thin lines with few colors, while natural images usually have continuous-tone, smoother edges, thicker lines and more colors. Moreover, the acquisition of natural images may introduce noise due to the physical limitations of imaging sensors, while the screen content is usually noise free as they may be purely generated by computers. In view of these distinct properties of SCIs, in<sup>6</sup> a screen image quality assessment database (SIQAD) was created, which contains 20 reference and 980 distorted SCIs in total. The distorted images are generated by different distortion types including Gaussian noise, Gaussian blur, motion blur, contrast changing, JPEG, JPEG2000 and layer segmentation based coding. The reported low correlations between the scores of subjective and objective measures suggest that there is still large room to improve for SCI quality assessment.<sup>6</sup> In other words, IQA methods that suffice to provide useful quality evaluation of SCIs are largely lacking.

In this work, we study the characteristics of the SCIs and propose an IQA method that predicts SCI quality by incorporating viewing field adaption and local information content weighting. As widely hypothesized in computational vision science, the major task of the human visual system (HVS) when viewing a real scene is to act as an optimal information extractor, or an efficient coder.<sup>7</sup> This motivates us to evaluate the quality of SCIs with the strategy of local information content weighting. Another psychology finding regarding the perception of screen images is that the extent of the visual field used to extract useful information is much larger in pictorial than in textual regions.<sup>8</sup> A possible reason accounting for such observation is that the textual content is richer in salient stimuli. These observations further inspire us to introduce spatial adaptation in the local quality assessment approach.

In contrast to the numerous recent efforts in developing high efficiency SCI compression techniques, little has been dedicated to visual perception based SCI compression. This is due to the lack of trusted SCI IQA models that can provide essential guidance in optimizing advanced SCI coding schemes. Given our newly proposed SCI IQA method, we further incorporate it into a High Efficiency Video Coding (HEVC) screen content codec, targeting at improving the coding efficiency of SCIs. Specifically, we propose a novel perceptual SCI compression scheme inspired by the design philosophy of the divisive normalization transform,<sup>9</sup> which has been shown to be a useful framework that better accounts for the spatially varying distortion sensitivities of the HVS.

## OBJECTIVE QUALITY ASSESSMENT OF SCIs

### Characteristics of SCIs

We find that two statistical features are useful in differentiating the characteristics of pictorial and textual regions in an image, and also in the development of meaningful IQA and compression methods for SCIs.

### Frequency Energy Falloff Statistics

It has long been discovered in the literature of natural scene statistics that the amplitude spectrum of natural images falls with the spatial frequency approximately proportional to the  $1/f_s^p$  law,<sup>10</sup> where  $f_s$  is the spatial frequency and  $p$  is an image dependent constant. By contrast, typical textual images generated by computers appear somewhat “unnatural.” This inspires us to further examine such property on SCIs. Examples of natural and textual images are decomposed using

Fourier transform, as demonstrated in Figure 1. It is observed that the energy falloffs against spatial frequency for natural images are approximately straight lines in log-log scale, which is consistent with the  $1/f_s^p$  relationship. However, for textual images there are peaks at mid and high frequencies. It is also interesting to observe that larger characters push the peak frequency towards lower frequencies, which further demonstrates the close relationship between the peak and the width and spacing of the strokes. Although such properties are not explicitly taken advantage of in the design of the proposed IQA method, these observations suggest that the statistical properties of textual images differ from natural images, motivating us to distinguish them in the design of quality assessment method.

We propose a perceptual two-pass rate control scheme for High Efficiency Video Coding (HEVC). The target bits are optimally allocated by hierarchically constructing a perceptual uniform space derived based on an SSIM-inspired divisive normalization mechanism for each group of pictures (GoP), frame, and coding unit (CU). The Lagrange multiplier  $\lambda$ , which controls the trade-off between perceptual distortion and bit rate, is adopted as the frame level complexity measure. After the first pass compression, Laplacian based rate and perceptual distortion models are established to adaptively derive  $\lambda$ , and the target bits are dynamically allocated by maintaining an uniform Lagrange multiplier level through  $\lambda$ -quantization. Within each GoP, rate control is further performed at frame and CU levels in the perceptually uniform space. Extensive simulations verify that, the proposed scheme can achieve high accuracy rate control and superior rate-SSIM performance.

*Index Terms*— Two-pass rate control, divisive normalization, SSIM index, High Efficiency Video Coding

## 1. INTRODUCTION

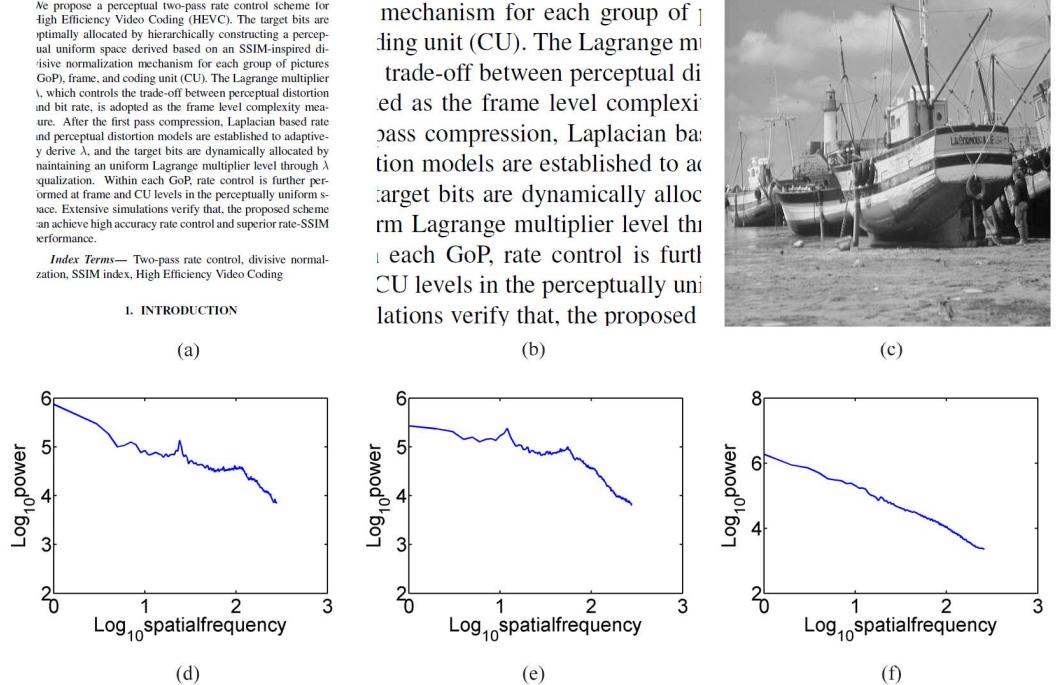


Figure 1. Examples of frequency energy falloffs of textual and natural images in log-log scale. (a) & (b) Textual images at different scales; (c) A natural image; (d) & (e) Frequency energy falloffs of textual images in (a) & (b); (f) Frequency energy falloff of the natural image (c).

## Information Content of SCIs

An effective information content model<sup>11</sup> is obtained by locally modeling the input signal with a Gaussian source that is transmitted through a Gaussian noise channel to the receiver.<sup>3</sup> As such, the mutual information between the input and received signals is the amount of the perceived information content, which can be quantified by

$$\omega = \log_2 \left( 1 + \frac{\sigma_p^2}{\sigma_n^2} \right) \quad (1)$$

where  $\sigma_p^2$  is the variance within a local window  $x$ , and  $\sigma_n^2$  is a constant parameter accounting for the noise level in the visual channel. An example of the local information maps computed using (1), together with the corresponding original images are shown in Figure 2, which provide a useful indicator about how perceptual information is distributed over space and how the distributions are different in textual and pictorial regions. In particular, since the local variances around high contrast edges are usually significant, higher information content can be observed from the information content map. As such, textual regions that contain abundant high contrast edges typically have higher local information content than pictorial regions. This is also consistent with

the recent findings regarding the saliency of webpages in,<sup>12</sup> in which it is shown that SCIs contain richer information in textual regions.



**Figure 2.** Examples of SCIs and the corresponding local information content maps (brighter indicates higher information content). (a)(c) SCIs; (b)(d) Corresponding information content maps.

## Quality Assessment Model

The local quality prediction of SCIs is based on the structural similarity (SSIM) index,<sup>1</sup> which has been demonstrated to be an effective quality measure that achieves a good compromise between quality prediction accuracy and computational efficiency. Given two local image patches  $\mathbf{x}$  and  $\mathbf{y}$  extracted from the original and distorted images, respectively, the SSIM index between them is evaluated as

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

where  $\mu_x$ ,  $\sigma_x$ , and  $\sigma_{xy}$  are the mean, standard deviation and cross correlation within a local window of size  $l \times l$ , respectively.  $C_1$  and  $C_2$  are positive constants used to avoid instability when the means and variances are close to zero, which are set to be

$$C_1 = (K_1 L)^2 \quad C_2 = (K_2 L)^2 \quad (3)$$

where  $L$  denotes the dynamic range of the pixel values. Parameters  $K_1$  and  $K_2$  are constants and are selected to be 0.01 and 0.03, respectively.

The distinct characteristics of SCIs described in Section 2.1 suggest that it is useful to differentiate the pictorial and textual content, such that their perceptual distortions can be evaluated in different ways. Another interesting property regarding the perception of the screen content is the extent of visual field when viewing SCIs. Monica Castelhano and Keith Rayner observed that the perceptual span in reading textual content is clearly smaller than that in natural scene perception or visual search.<sup>8</sup> This further motivates us to adapt the window size when accessing the local quality of textual and pictorial content.

In this work, instead of performing image segmentation that divides the image into large segments of textual and pictorial regions, we propose a block-classification approach by making use of the information content map, as shown in Figure 2. Based on our analysis earlier, the textual regions are richer in saliency stimuli and typically have higher local information content. Therefore, we classify each  $4 \times 4$  block by applying a threshold  $T_f$  on the sum of the information content in the block. Subsequently, the overall quality of the textual and pictorial regions  $\Omega_T$  and  $\Omega_P$ , denoted by  $S_T$  and  $S_P$ , respectively, are computed by applying spatially adaptive weighted pooling to access the relative weight of the local content within the textual or pictorial content,

$$S_T = \frac{\sum_{i \in \Omega_T} \text{SSIM}_i \cdot \omega_i^\alpha}{\sum_{i \in \Omega_T} \omega_i^\alpha} \quad (4)$$

$$S_P = \frac{\sum_{j \in \Omega_P} \text{SSIM}_j \cdot \omega_j^\alpha}{\sum_{j \in \Omega_P} \omega_j^\alpha}$$

where the parameter  $\alpha$  is used to adjust the strength of weighting. The parameters  $T_f$  and  $\alpha$  are selected empirically to be 30 and 0.3, respectively. Since textual content is perceived with smaller extend of visual field than pictorial regions, the local SSIM value is calculated by employing different sizes of Gaussian windows with different standard deviations (std), denoted by  $k_T$  and  $k_P$ , respectively. The local information  $\omega_i$  and  $\omega_j$  are calculated with their respective windows, within which the SSIM indices are computed. It is also worth mentioning that textual content is not the only difference between the natural images and SCIs, for example, SCIs often contain large flat areas. However, text is the most dominant characteristic in SCIs. It conveys meaningful information and meanwhile produces high perceptual contrast. Fortunately, this is captured by the information content measure that is used as a weighting factor in the proposed method.

The final SCI quality index (SQI) is given by a weighted average of  $S_T$  and  $S_P$ , which computes the relative weight between textual and pictorial region as a whole,

$$\text{SQI} = \frac{S_T \cdot \mu_T + S_P \cdot \mu_P}{\mu_T + \mu_P} \quad (5)$$

where  $\mu_T = \frac{1}{|\Omega_T|} \sum_{j \in \Omega_T} \omega_{u,j}^\alpha$  and  $\mu_P = \frac{1}{|\Omega_P|} \sum_{j \in \Omega_P} \omega_{u,j}^\alpha$ , respectively. These quantities measure the relative density of information content of the textual and pictorial regions, and for fair comparison, the Gaussian window size, denoted by  $k_U$ , used to compute  $\omega_u$  needs to be uniform in both regions, and should be a compromise between  $k_T$  and  $k_P$ .

The window size parameters  $k_T$ ,  $k_P$  and  $k_U$  are determined empirically. Based on our discussions earlier, it is a natural choice to let  $k_T < k_P$ , and  $k_U$  in-between. In our current implementation, we set  $k_T = 0.5$ ,  $k_P = 2.5$  and  $k_U = 1.5$ , respectively.

## Validation

The SIQAD database was designed specifically for SCI quality assessment. It contains 980 images that are corrupted by seven distortion types. Full reference IQA algorithms including PSNR, SSIM,<sup>1</sup> IW-SSIM,<sup>2</sup> GSIM,<sup>13</sup> FSIM,<sup>4</sup> VSI,<sup>5</sup> and VIF<sup>3</sup> are used for comparison. In previous tests using subject-rated IQA databases, these state-of-the-art IQA algorithms have been repeatedly proven to achieve high correlations with the mean opinion scores (MOS) of natural images.<sup>14</sup> Moreover, the specifically developed IQA measure SCI Perceptual Quality Assessment (SPQA)<sup>6</sup> is compared as well. Three evaluation metrics are employed to assess the performance of these IQA methods, including Pearson linear correlation coefficient (PLCC), Root mean-squared error (RMSE) and Spearman rank correlation coefficient (SRCC).

The PLCC is computed after a nonlinear mapping between the subjective and objective scores to evaluate the prediction accuracy. Given the raw objective scores, a logistic regression function is employed to generate the mapped scores, and PLCC is obtained by computing the correlation between the subjective and mapped objective scores. RMSE is subsequently calculated by comparing the subjective and objective scores after nonlinear mapping. SRCC is nonparametric rank order-based correlation metrics to assess prediction monotonicity. A better objective IQA measure should have higher PLCC and SRCC, but lower RMSE values.

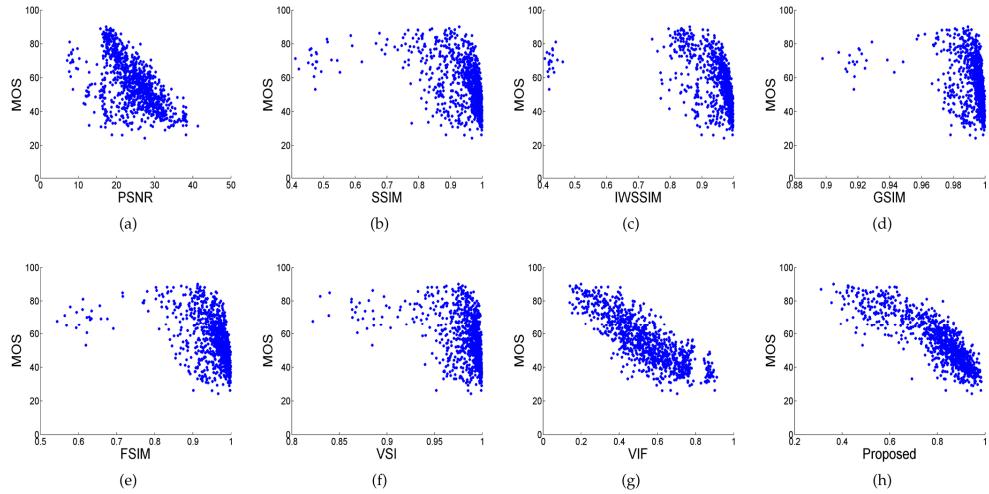
As illustrated in Table 1, when all the test images are included in the evaluation, the proposed method clearly outperforms state-of-the-art quality assessment algorithms in terms of both prediction accuracy and monotonicity. Moreover, we examine the breakdown prediction performance for individual distortion types. The breakdown performance in terms of SRCC and PLCC are provided in Table 2, where in most cases, the proposed method is among the best. Scatter plots of human ratings versus raw objective predicted quality scores before nonlinear mapping for each FR methods are shown in Figure 3. It can be observed that the proposed method can accurately predict the MOS scores.

**Table 1. Performance Comparison with State-of-the-Art FR Algorithms Based on the SIQAD Database.**

| IQA Methods | PSNR    | SSIM    | IWSSIM  | GSIM    | FSIM    | VSI     | VIF    | SPQA   | SQI           |
|-------------|---------|---------|---------|---------|---------|---------|--------|--------|---------------|
| SRCC        | 0.5608  | 0.5836  | 0.6546  | 0.5483  | 0.5819  | 0.5381  | 0.8069 | 0.8416 | <b>0.8548</b> |
| PLCC        | 0.5869  | 0.5912  | 0.6536  | 0.5686  | 0.5902  | 0.5568  | 0.8206 | 0.8584 | <b>0.8644</b> |
| RMSE        | 11.5898 | 11.5450 | 10.8329 | 11.7750 | 11.5552 | 11.8904 | 8.1795 | 7.3421 | <b>7.1982</b> |

**Table 2. Distortion Type Breakdown for IQA Performance Comparisons.**

|        | Gaussian Noise |               | Gaussian Blur |               | Motion Blur   |               | Contrast Change |               | JPEG          |               | JPEG2000      |               | Layer Coding  |               |
|--------|----------------|---------------|---------------|---------------|---------------|---------------|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|        | PLCC           | SRCC          | PLCC          | SRCC          | PLCC          | SRCC          | PLCC            | SRCC          | PLCC          | SRCC          | PLCC          | SRCC          | PLCC          | SRCC          |
| PSNR   | <b>0.9052</b>  | 0.8790        | 0.8603        | 0.8577        | 0.7044        | 0.7132        | 0.7528          | 0.6828        | 0.7696        | 0.7569        | 0.7893        | 0.7746        | 0.7809        | 0.7928        |
| SSIM   | 0.8657         | 0.8495        | 0.8530        | 0.8445        | 0.7464        | 0.7443        | 0.8106          | 0.7251        | 0.6921        | 0.6910        | 0.7037        | 0.6937        | 0.6601        | 0.6515        |
| IWSSIM | 0.8882         | 0.8743        | 0.9082        | 0.9060        | 0.8417        | 0.8421        | <b>0.8411</b>   | <b>0.7540</b> | 0.7998        | 0.7981        | 0.8041        | 0.7988        | 0.8155        | 0.8214        |
| GSIM   | 0.8511         | 0.8428        | 0.8830        | 0.8797        | 0.7750        | 0.7765        | 0.8169          | 0.7322        | 0.6765        | 0.6803        | 0.7244        | 0.7123        | 0.7231        | 0.7172        |
| FSIM   | 0.8848         | 0.8705        | 0.8208        | 0.8221        | 0.7284        | 0.7239        | 0.8222          | 0.7146        | 0.6640        | 0.6637        | 0.7046        | 0.6851        | 0.7034        | 0.7055        |
| VSI    | 0.8836         | 0.8655        | 0.8504        | 0.8495        | 0.7657        | 0.7658        | 0.7734          | 0.6459        | 0.7149        | 0.7196        | 0.7498        | 0.7299        | 0.7457        | 0.7419        |
| VIF    | 0.9015         | <b>0.8890</b> | 0.9101        | 0.9062        | 0.8503        | 0.8504        | 0.7083          | 0.5269        | 0.8005        | 0.7932        | 0.8221        | 0.8151        | <b>0.8414</b> | <b>0.8495</b> |
| SQI    | 0.8829         | 0.8602        | <b>0.9202</b> | <b>0.9244</b> | <b>0.8789</b> | <b>0.8810</b> | 0.7724          | 0.6677        | <b>0.8218</b> | <b>0.8189</b> | <b>0.8271</b> | <b>0.8169</b> | 0.8310        | 0.8432        |



**Figure 3. MOS versus model predictions.**

Generally, a good IQA measure should be tolerant to small changes in parameter values. Therefore, the sensitivities of the parameters  $T_f$  and  $\alpha$  on the IQA performance are examined. In particular,  $T_f$  varies from 20 to 40 with an interval of 5 and  $\alpha$  varies from 0.1 to 0.5 with an interval of 0.1. The results are tabulated in Table 3 and Table 4, from which we can observe that the proposed method achieves considerably stable performance.

**Table 3. Parameter sensitivity testing with the variation of  $T_f$ .**

| $T_f$ | <b>20</b> | <b>25</b> | <b>30</b> | <b>35</b> | <b>40</b> |
|-------|-----------|-----------|-----------|-----------|-----------|
| SRCC  | 0.8580    | 0.8577    | 0.8548    | 0.8489    | 0.8365    |
| PLCC  | 0.8630    | 0.8645    | 0.8644    | 0.8620    | 0.8533    |
| RMSE  | 7.2318    | 7.1946    | 7.1982    | 7.2558    | 7.4640    |

**Table 4.** Parameter sensitivity testing with the variation of  $\alpha$ .

| $\alpha$ | <b>0.1</b> | <b>0.2</b> | <b>0.3</b> | <b>0.4</b> | <b>0.5</b> |
|----------|------------|------------|------------|------------|------------|
| SRCC     | 0.8592     | 0.8586     | 0.8548     | 0.8498     | 0.8438     |
| PLCC     | 0.8617     | 0.8648     | 0.8644     | 0.8620     | 0.8582     |
| RMSE     | 7.2629     | 7.1866     | 7.1982     | 7.2569     | 7.3469     |

## PERCEPTUAL SCI COMPRESSION

### Divisive Normalization Based Video Coding

In the predictive video coding framework, previously coded frames are used to predict the current frame, and only the residuals after prediction are transformed and coded. In divisive normalization transform based video coding scheme,<sup>15</sup> the discrete cosine transform (DCT) coefficient of a residual block  $C_k$  is normalized by a positive perceptual normalization factor  $f$  to transform the DCT coefficients into a perceptually uniform domain:

$$C(k)' = C(k) / f \quad (6)$$

Subsequently, given the predefined quantization step  $Q_s$ , the quantization process of the normalized residuals is formulated as

$$\begin{aligned} Q(k) &= \text{sign}\{C(k)\} \text{round} \left\{ \frac{|C(k)|}{Q_s} + p \right\} \\ &= \text{sign}\{C(k)\} \text{round} \left\{ \frac{|C(k)|}{Q_s \cdot f} + p \right\} \end{aligned} \quad (7)$$

where  $p$  is the rounding offset in the quantization.

Correspondingly, at the decoder, the de-quantization and reconstruction of  $C(k)$  is performed

$$\begin{aligned} R(k) &= R(k)' \cdot f = Q(k) \cdot Q_s \cdot f \\ &= \text{sign}\{C(k)\} \text{round} \left\{ \frac{|C(k)|}{Q_s \cdot f} + p \right\} \cdot Q_s \cdot f \end{aligned} \quad (8)$$

As such, the transform coefficients are converted into a perceptually uniform space by adaptively adjusting the quantization parameters for each coding unit (CU). The normalization factor  $f$ , which accounts for the perceptual importance, is derived from the SSIM index in DCT domain.<sup>16</sup> Specifically, given the reference and the reconstructed blocks, denoted by  $x$  and  $y$ , respectively, the DCT domain SSIM can be calculated as,

$$\text{SSIM}(x, y) = \left( 1 - \frac{(X(0) - Y(0))^2}{X(0)^2 + Y(0)^2 + N \cdot C_1} \right) \times \left( 1 - \frac{\sum_{k=1}^{N-1} (X(k) - Y(k))^2}{\sum_{k=1}^{N-1} X(k)^2 + Y(k)^2 + C_2} \right) \quad (9)$$

where  $X$  and  $Y$  represent the DCT coefficient of  $x$  and  $y$ , respectively. Parameter  $N$  denotes the size of the block, and  $C_1, C_2$  are constants according to the definition of the SSIM index.<sup>1</sup> Assuming each CU contains  $l$  DCT blocks, the normalization factors for AC coefficients are given by

$$f_{ac} = \frac{\frac{1}{l} \sum_{i=1}^l \sqrt{\frac{\sum_{k=1}^{N-1} (X_i(k)^2 + Y_i(k)^2)}{N-1} + C_2}}{E(\sqrt{\frac{\sum_{k=1}^{N-1} (X(k)^2 + Y(k)^2)}{N-1} + C_2})} \quad (10)$$

Practically, only the original block is used because the distorted one cannot be accessed before the actual encoding, and  $f_{ac}$  is applied to derive the quantization parameter (QP) offset for each CU.

Following the divisive normalization process, rate distortion optimization (RDO) is performed by minimizing the perceived distortion  $D$  with the bit rate  $R$  subject to a constraint  $R_c$ . This can be converted to an unconstrained optimization problem by

$$\min\{J\} \quad \text{where } J = D + \lambda \cdot R \quad (11)$$

where  $J$  denotes the rate-distortion (RD) cost and  $\lambda$  is known as the Lagrange multiplier which controls the tradeoff between  $R$  and  $D$ . Specifically, the distortion  $D$  is defined by computing the sum of squared difference (SSD) between the normalized original and distorted coefficients, which is given by

$$D = \sum_{i=1}^l \sum_{k=0}^{N-1} (C_i(k) - R_i(k))^2 = \sum_{i=1}^l \sum_{k=0}^{N-1} \frac{(C_i(k) - R_i(k))^2}{f_{ac}^2} \quad (12)$$

As the divisive normalization is performed to transform the DCT coefficients into a perceptually uniform space, the Lagrangian multiplier  $\lambda$  in rate distortion optimization is untouched in the encoder.

## Perceptual SCI Compression

The main difference between SSIM and SQI may be well accounted for by the window adaption and information content weighting process. Specifically, in analogies to the SQI method, block type classification is firstly performed by evaluating the local information content in each block and then compare it with a predefined threshold. Subsequently, based on the design philosophy of SQI, the normalization factor for a textual block is given by

$$f_t = f_{ac} / g_t \quad (13)$$

where  $g_t$  denotes the relative importance of the local block in terms of the information content:

$$g_t = \sqrt{\frac{2 \left( \frac{1}{l} \frac{1}{N} \sum_{i=1}^l \sum_{k=1}^N \omega_{i,k}^\alpha \right) \cdot \mu_T}{\left( \frac{1}{|\Omega_T|} \sum_{j \in \Omega_T} \omega_j^\alpha \right) \cdot (\mu_T + \mu_p)}} \quad (14)$$

where  $k$  is the spatial location index within a block of size  $N$ , and  $i$  is the block index within each CU.

Similarly, the normalization factor for a pictorial block is given by

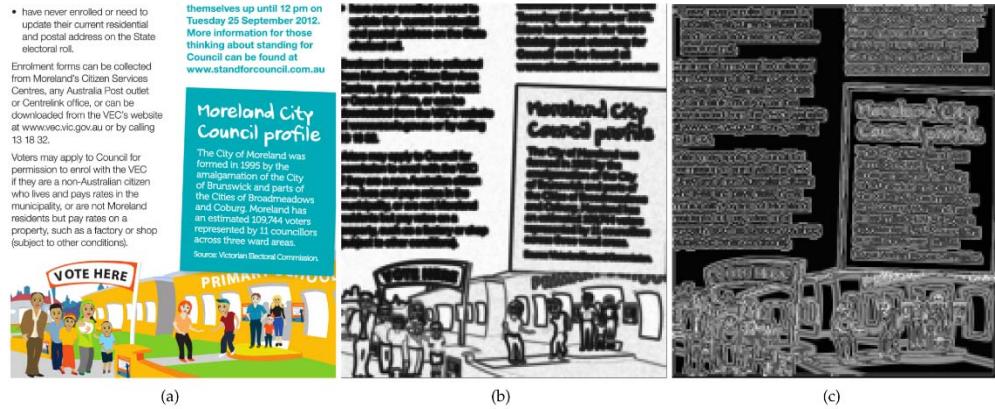
$$f_p = f_{ac} / g_p \quad (15)$$

where

$$g_p = \sqrt{\frac{2 \left( \frac{1}{l} \frac{1}{N} \sum_{i=1}^l \sum_{k=1}^N \omega_{i,k}^\alpha \right) \cdot \mu_p}{\left( \frac{1}{|\Omega_p|} \sum_{j \in \Omega_p} \omega_j^\alpha \right) \cdot (\mu_T + \mu_p)}} \quad (16)$$

It is worth noting that the local information here is computed with Gaussian window of larger size  $k_P$ .

The divisive normalization factors derived from SSIM and SQI for a typical SCI are given in Figure 4, where the divisive normalization factors are computed within each 4x4 block for better visualization. The results demonstrate that with the proposed method, we can assign smaller normalization factors to textual regions with high contrast edges, which are more sensitive to the HVS compared to the pictorial regions. Consequently, with the divisive normalization approach that is specifically designed for SCI compression, we are able to adapt the bit allocation process to improve the overall SCI quality.



**Figure 4.** Visualization of spatially adaptive divisive normalization factors for a typical SCI (darker pixels indicate higher normalization factors). (a) Original SCI. (b) Normalization factors derived from SSIM. (c) Normalization factors derived from SQI.

## Experimental Results

We incorporate the proposed perceptual SCI coding approach into the newly developed HEVC screen content coding extension codec. The test images are in YUV4:4:4 format from both the SIQAD database and HEVC test sequences (the first frame of each sequence). They include common scenarios in screen image processing, such as web browsing, office software editing and video-conferencing. The R-D performance gain (BD-Rate) between the original HEVC encoder (anchor) and the proposed approach in terms of SQI is given in Table 5. It is observed that significant bit rate saving is achieved, which further demonstrates the effectiveness of the proposed quality measure in potential applications such as encoder optimization.

**Table 5.** RD Performance for Different SCIs in terms of SQI.

| Image              | BD-Rate |
|--------------------|---------|
| Webpage            | -4.6%   |
| Digital magazine   | -4.3%   |
| PPT DOC XLS        | -5.0%   |
| Programming        | -4.2%   |
| Video Conferencing | -8.3%   |
| Word Editing       | -9.7%   |

We further carried a subjective test to verify the proposed perceptual SCI coding scheme. The subjective test is based on a two-alternative-forced-choice (2AFC) method, which has been

widely adopted in comparing the subjective quality of two video sequences.<sup>15,17</sup> Specifically, in each trial a subject is forced to choose the one he/she thinks to have better quality from a pair of compressed SCIs. We selected four pairs of SCIs, and each pair is repeated four times in random order. In total, 14 subjects were invited in the subjective test. The coding bits, SQI and the results of the subjective tests are reported in Table 6, where the percentage by which the subjects are in favor of the anchor against the proposed scheme are demonstrated. As can be observed, the SCIs are compressed at similar bit rates, and the subjects are inclined to select the proposed method to have better quality. These results provide useful evidence that the proposed method improves the coding performance in terms of subjective quality.

**Table 6. Subjective Test Configurations and Results.**

| <b>SCI</b>          | <b>Anchor</b> |            | <b>Proposed</b> |            | <b>Percentage<br/>(In favor of<br/>Anchor)</b> |
|---------------------|---------------|------------|-----------------|------------|--|
|                     | <b>bpp</b>    | <b>SQI</b> | <b>bpp</b>      | <b>SQI</b> |  |
| PPT DOC XLS         | 0.260         | 0.8856     | 0.258           | 0.8937     | 23.21%   |
| Program-ming        | 0.225         | 0.9286     | 0.230           | 0.9402     | 8.93%  |
| Video Con-ferencing | 0.378         | 0.9363     | 0.374           | 0.9454     | 14.29%   |
| Word Edit-ing       | 0.265         | 0.8748     | 0.261           | 0.8983     | 5.36%  |

## CONCLUSION

We propose an objective quality assessment method for SCIs and then employ it to optimize the encoding process of SCI compression. The quality assessment method differentiates textual and pictorial blocks, and applies different parameters in the computation of the structural similarity for local quality assessment. A local information content weighting scheme is further adopted to derive the optimal perceptual weights for spatial pooling. Experimental results show the superior performance of the proposed method in predicting the quality of SCIs, and also demonstrate its potential in improving the performance of SCI compression.

## ACKNOWLEDGEMENT

Partial preliminary results of this work were presented at IEEE International Conference on Image Processing, Quebec, Canada, Sep. 2015. This work was also supported in part by National Natural Science Foundation of China under Grants 61703009.

## REFERENCES

1. Z. Wang et al., “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004, pp. 600–612.
2. Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 5, 2011, pp. 1185–1198.

3. H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, 2006, pp. 430–444.
4. L. Zhang, L. Zhang, and X. Mou, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, 2011, pp. 2378–2386.
5. L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, 2014, pp. 4270–4281.
6. H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, 2015, pp. 4408–4421.
7. E.P. Simoncelli and B.A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, no. 1, 2001, pp. 1193–1216.
8. M.S. Castelhano and K. Rayner, "Eye movements during reading, visual search, and scene perception: An overview," *Cognitive and Cultural Influences on Eye Movements*, Psychology Press, 2009.
9. M.J. Wainwright and E.P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," *Proceedings of the 1999 Conference on Advances in Neural Information Processing Systems (NIPS 99)*, 1999, pp. 855–861.
10. D.J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vision research*, vol. 37, no. 23, 1997, pp. 3367–3383.
11. Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," *IEEE International Conference on Image Processing*, 2006, pp. 2945–2948.
12. C. Shen, X. Huang, and Q. Zhao, "Predicting eye fixations on webpage with an ensemble of early features and highlevel representations from deep network," *IEEE Transactions on Multimedia*, vol. 17, no. 11, 2015, pp. 2084–2093.
13. A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Transactions on Image Processing*, 2012, pp. 1500–1512.
14. W. Lin and C.-C. Jay Kou, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, 2011, pp. 297–312.
15. S. Wang et al., "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Transactions on Image Processing*, vol. 22, no. 4, 2013, pp. 1418–1429.
16. S.S. Channappayya, A.C. Bovik, and R.W. Heath Jr., "Rate bounds on SSIM index of quantized images," *IEEE Transactions on Image Processing*, vol. 17, no. 9, 2008, pp. 1624–1639.
17. S. Wang et al., "SSIM-motivated rate-distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 4, 2012, pp. 516–529.

## ABOUT THE AUTHORS

**Shiqi Wang** received a PhD in computer application technology from Peking University, Beijing, China, in 2014. He was a Postdoc Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. His research interests include image/video quality assessment, compression and analysis. Contact him at sqwang1986@gmail.com.

**Ke Gu** received a PhD in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2015. He is currently with Beijing University of Technology. His research interests include quality assessment, contrast enhancement, and visual saliency. Contact him at guke@ntu.edu.sg.

**Kai Zeng** received a PhD in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2013, where he is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering. His research interests include

image and video quality assessment and corresponding applications. Contact him at kzeng@uwaterloo.ca.

**Zhou Wang** received a PhD in electrical and computer engineering from The University of Texas at Austin, in 2001. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications, and biomedical signal processing. Contact him at zhou.wang@uwaterloo.ca.

**Weisi Lin** received a PhD from King's College London, London, U.K., in 1993. He is currently a Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling. Contact him at wslin@ntu.edu.sg.

# Efficient C<sup>2</sup>-Weighting for Image Warping

**Chuhua Xian**

South China University of Technology

**Shuo Jin**

The Chinese University of Hong Kong

**Charlie C. L. Wang**

Delft University of Technology

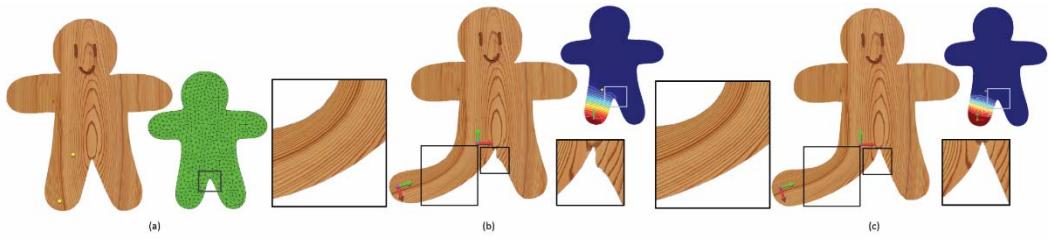
Handle-driven image warping based on linear blending is widely used in many applications because of its merits on intuitiveness, efficiency, and ease of implementation. In this paper, we develop a method to compute high-quality weights within a closed domain for image warping. The property of C<sup>2</sup> continuity in weights is guaranteed by the carefully formulated basis functions. The efficiency of our

algorithm is ensured by a closed-form formulation of the computation for weights. The cost of inserting a new handle is only the time to evaluate the distances from the new handle to all other sample points in the domain. A virtual handle insertion algorithm is developed to allow users to freely place handles within the domain while preserving the satisfaction of all expected criteria on weights for linear blending. Experimental examples for real-time applications are shown to demonstrate the effectiveness of this method.

Image warping based on linear blending has a wide popularity, as it is intuitive, effective, and easy to implement in many different scenarios of applications. By inserting handles within an image domain, users can bind an image  $\Omega$  with the handles and then manipulate their locations and orientations to drive the deformation of  $\Omega$ . Specifically, each handle  $H_i$  with  $i = 1, \dots, m$  is defined as a local frame with its origin  $\mathbf{h}_i \in \Omega$ . After defining an affine transformation  $\mathbf{T}_i$  for each handle  $H_i$ , the deformation of  $\Omega$  is realized by computing the new position of each point  $\mathbf{p} \in \Omega$  through a linear blending of affine transformations  $\mathbf{T}_i \mathbf{p}$ . The linear blending is weighted by fields  $w_i : \Omega \rightarrow \mathbb{R}$  associated with handles  $H_i$ . Basically, to achieve an intuitive and high-quality deformation, a few criteria on the weights are demanded, including smoothness, non-negativity, partition-of-unity, locality/sparsity, and nonlocal maxima/minima.<sup>1</sup>

The recent advancement of technology focuses on computing weights of blending for warping with a discrete form of domain (meshes are employed to determine piecewise linear fields of weights). Weights are computed on mesh nodes by minimizing some discrete differential energies (biharmonic, triharmonic, and quatriharmonic forms<sup>2</sup>). After incorporating hard constraints in equality-form or inequality-form according to different criteria, weights are determined on

mesh nodes with the help of non-linear optimization solvers, which is usually time-consuming. This leads to the difficulty in the realization of real-time insertion of new handles as new routines of non-linear optimization need to be taken. Moreover, the determined weights are mesh-dependent, which indicates an implicit requirement on mesh quality. For a poorly meshed domain, the artificial distortion caused by elements in poor shape could be serious (as illustrated in Figure 1). Although the artifacts can be reduced by remeshing the domain to improve mesh density and quality, this assumes high efficiency of weight generation. Otherwise, a very slow computation process will be disappointing to users. Ideally, the distribution of weights should be affected only by the domain to be warped and the locations of handles, which indicates mesh-independence. The existing mesh-independent approaches in literature for handle-driven deformation<sup>3–5</sup> can only satisfy subsets of the demanded properties on weights. These factors motivate us to investigate an efficient and high-quality method to determine weights, which can be applied to both meshed and meshless representations of an image domain.



**Figure 1.** Image warping based on linear blending. (a) Handle-driven deformation based on linear blending is an intuitive method for the interactive shape manipulation. (b) Artifacts are found with the weights computed by the mesh-dependent method<sup>1</sup>, which requires solving an equation system over the whole underlying mesh. (c) We propose a new framework to generate  $C^2$ -continuous weights for linear blending based deformation with high efficiency. Our approach inherits the merits of mesh-dependent weighting schemes while bringing the weighting to the resolution of infinity.

In this paper, we formulate the evaluation of weights in a closed-form so that the deformation framework gains the benefit of flexibility—the response of inserting new handles is real-time. Specifically, the time cost of inserting a new handle is linear to the number of samples used to represent the domain of computation. The properly chosen basis function according to the defined criteria can guarantee the properties of smoothness, non-negativity, partition-of-unity, locality/sparsity, and nonlocal maxima/minima, all of which are necessary to ensure a high-quality deformation.

Our main results are as follows:

- We present an efficient method to determine linear blending weights with  $C^2$ -continuity for real-time image warping. The weights are formulated in a closed-form of basis functions centered at the handles (see the Weighting Formulation section). After decomposing the domain to be deformed by the Voronoi diagram of handles, the aforementioned criteria for image warping are all ensured.
- A virtual handle insertion algorithm is proposed to guarantee the locality and sparsity of weights, so that a deformation interpolates all transformations defined on handles. The virtual handles are added to let the supporting region of the basis function defined on a handle not cover the origins of any other handles (see the algorithm in the Virtual Handle Insertion section).
- After constructing the Voronoi diagram of all handles (including user-input and virtual ones), its dual-graph gives a connectivity of the handles. We compute harmonic fields on the graph to determine the transformations of virtual handles according to the ones specified on the user-input handles (see the Virtual Handle Insertion section). It is proved that the deformation determined in this way leads to a natural manner that follows the intention of user input well.

- For different discrete representations of the domain to be warped, we show that our framework is efficient and flexible for handle-driven image warping (see the Implementation Details section). The experimental results in the Results section demonstrate the performance of our approach.

## RELATED WORK

Shape deformation is an important research area in image manipulation and geometric modeling. There is a large amount of existing approaches in literature. The purpose of this section is not to give a comprehensive review. We focus on discussing only the handle-driven deformation approaches.

Mesh-based techniques for discrete geometry modeling and processing have been widely explored in the past decade. Typical approaches include variational surface deformation, Poisson deformation, Laplacian editing, and other linear variational surface deformation approaches.<sup>6</sup> Volumetric information and rigidity are also incorporated to enhance shape-preservation.<sup>7,8</sup> One common drawback of these approaches is that the positions of vertices on a model need to be determined by solving a system of linear equations after every update of handles, which becomes a bottleneck of computation. One recent development<sup>1,2,9</sup> transfers the workload from online optimization to offline. In another new technique, the handles are elements of a simplified mesh.<sup>5</sup> Although this strategy is more efficient than the deformation methods based on online optimization, it still cannot avoid solving large linear systems, which slows down the response of deformation after inserting new handles. Moreover, the results of deformation also suffer from the artificial distortions caused by mesh problems, such as a mesh that is too coarse for a fine deformation, a mesh with needle and cap triangles, and the problem of symmetry. Our approach solves these problems by providing closed-form formulas to generate weights, preserving all the demanded properties for producing deformations with high quality in real time.

Another thread of research focuses on mesh-independent approaches. Different handles are employed for shape manipulation. Some techniques use points,<sup>10</sup> while others employ curves as handles.<sup>3</sup> Cage-based deformation<sup>11,12</sup> can be considered an improved generalization of grid-based deformation,<sup>13</sup> where weights can be found by a closed-form in terms of the handles. However, the construction of cages is usually not automatic, and the manipulation on cages instead of a model itself is indirect.

The moving least squares (MLS) strategy is employed to interpolate similarity/rigid deformations at handle points.<sup>10</sup> A closed-form solution is provided in their approach to determine the transformation matrix on every point in an MLS manner. The transformations in the whole domain need to be computed when any handle is moved. There is no explicit determination on the influence region of each handle. Contrary to this MLS approach, our approach belongs to the category of linear blending-based deformation. When the property of sparsity is preserved on weights, the deformation at a point is affected by only the nearby handles that can easily be predicted by users. Moreover, the deformation determined by our approach is resolution-independent, which is very important for image manipulation.

The work of generating weights for linear blending also relates to the research of scattered data interpolation, where radial basis functions (RBF) are widely used. In one technique, the deformation is governed by global RBFs that lead to a dense linear system to be solved.<sup>14</sup> The weights are determined by the dense (or global) data interpolation approach's lack of sparsity. Therefore, every point in the domain is changed when any handle is updated, even if it is far away. Although the compactly supported radial basis functions (CSRBF) can help introduce the sparsity, it does not provide closed-form formulas like our approach. One improved approach with interior RBF is able to obtain natural deformation on various models, which consists of a precomputation phase and an online deformation phase.<sup>15</sup> By minimizing an energy functional to determine the coefficients of the IRBFs, the shape can be deformed in a shape- and volume-controlled manner. When adding or removing handles during manipulation, the precomputation phase needs to be redone, which could be slow if the shape is complex.

One recent work raises a new deformation framework to design linear deformation subspaces, which generally unifies linear blend skinning and generalized barycentric coordinates into an

identical variational form.<sup>16</sup> The deformation following the movement of user handles is formulated as the minimization of a quadratic energy subject to constraints from the handles on discrete meshes. The computational bottleneck of this approach falls on the factorization when new handles are inserted, which is overcome by using advanced numerical schemes. Our method particularly focuses on a linear blending scheme, giving a closed-form solution to compute weighting, which is applicable to any type of domain representation. It shares a problem with one technique,<sup>1</sup> which is the specification of affine transformations at handle points. The framework of another technique<sup>16</sup> has a lower barrier on this, as less manipulation needs to be taken on handles.

## WEIGHTING FORMULATION

The key idea of the linear blending is to determine the new position of a point  $\mathbf{p} \in \Omega$  by a linear combination of the transformations  $\mathbf{T}_i$  defined by users on handles  $H_i$  as

$$\mathbf{p}' = \sum_{i=1}^m w_i(\mathbf{p}) \mathbf{T}_i \mathbf{p} \quad (1)$$

with  $w_i(\cdot)$  being a scalar field of weights to be determined. ( $\mathbf{T}_i$  is a homogeneous matrix, and  $\mathbf{p}$  is represented by a homogeneous coordinate.) The origin of a handle  $H_i$  is denoted by  $\mathbf{h}_i$ . This linear blending-based formulation is as popular as it is fast and easy to implement. However, carelessness in assigning weights can lead to visible artifacts in results, meaning the quality of deformation result depends on the properties of  $w_i(\cdot)$  (see the analysis of our weighting in the Properties of Weights section). Our formulation, introduced below, guarantees the generation of a  $C^2$ -continuous scalar field of weights in a closed domain, leading to a high-quality deformation based on linear blending.

## Intrinsic Distance

*Definition:* All points on an image form a bounded domain  $\Omega$ . For any two points  $\{\mathbf{p}_s, \mathbf{p}_e\} \in \Omega$ , if there exists a curve (a path)  $\mathcal{C} \subset \Omega$  connecting  $\mathbf{p}_s$  and  $\mathbf{p}_e$ , we define the intrinsic distance of  $\{\mathbf{p}_s, \mathbf{p}_e\}$  along the curve  $\mathcal{C}$  as

$$d(\mathbf{p}_s, \mathbf{p}_e; \mathcal{C}) = \int_{\mathcal{C}} f ds .$$

Then the intrinsic-distance of  $\{\mathbf{p}_s, \mathbf{p}_e\}$  in  $\Omega$  is defined as

$$d(\mathbf{p}_s, \mathbf{p}_e) = \inf_{\mathcal{C}} \int_{\mathcal{C}} f ds .$$

If  $\mathbf{p}_s$  and  $\mathbf{p}_e$  are not located in a connected region of  $\Omega$ , then there is no path connecting these two points. In this case, the intrinsic distance is defined as  $d(\mathbf{p}_s, \mathbf{p}_e) = \infty$ . It is obvious that the intrinsic-distance has the following properties:

- **Existence:**  $d(\mathbf{p}_s, \mathbf{p}_e; \mathcal{C})$  is always calculable once  $\mathcal{C}$  is determined, which is the length of a curve segment in  $\Omega$ . Therefore,  $d(\mathbf{p}_s, \mathbf{p}_e)$  always exists for  $\Omega$  when  $\mathbf{p}_s$  and  $\mathbf{p}_e$  are located in the same connected region.
- **Uniqueness:**  $d(\mathbf{p}_s, \mathbf{p}_e)$  is uniquely determined while the corresponding paths may be multiple.

*Intrinsic distance approximation:* For a graph  $\mathcal{G}$  with a set of samples  $\mathcal{S} \in \Omega$  as its nodes, we represent the shortest distance on  $\mathcal{G}$  between  $\mathbf{p}_s \in \mathcal{G}$  and  $\mathbf{p}_e \in \mathcal{G}$  as  $d(\mathbf{p}_s, \mathbf{p}_e; \mathcal{S})$ . For a given  $\varepsilon > 0$ , if for any two points  $\{\mathbf{p}_s, \mathbf{p}_e\} \in \mathcal{S}$ , we always have  $|d(\mathbf{p}_s, \mathbf{p}_e; \mathcal{S}) - d(\mathbf{p}_s, \mathbf{p}_e)| \leq \varepsilon$ , then the sampling  $\mathcal{S}$  is called an  $\varepsilon$  error-bounded sampling of  $\Omega$ , and  $\mathcal{G}$  is its corresponding graph.  $d(\mathbf{p}_s, \mathbf{p}_e; \mathcal{S})$  is treated as an approximation of  $d(\mathbf{p}_s, \mathbf{p}_e)$ .

In this work, the computation of weights in practice relies on the approximate intrinsic distance. When a domain is convex, the intrinsic distance is equal to the Euclidean distance between any

two points in it. It must be clarified that an error-bounded sampling  $\mathcal{S}$  usually does not sufficiently indicate  $\mathcal{S}$  is a high-quality approximation of the domain, which also requires that the sampling be dense to a certain extent.

## Formulation

Each handle  $H_i$  is equipped with a compactly supported basis function with a support size  $r_i$  as  $\phi_i(d(\mathbf{p}, \mathbf{h}_i)/r_i)$ , where  $\mathbf{h}_i$  is the location of  $H_i$  and  $d(\cdot, \cdot)$  returns the intrinsic distance between two points inside  $\Omega$ . The scalar field of weights for  $H_i$  is then defined as

$$w_i(\mathbf{p}) = \frac{\phi_i(d(\mathbf{p}, \mathbf{h}_i)/r_i)}{\sum_{j=1}^m \phi_j(d(\mathbf{p}, \mathbf{h}_j)/r_j)}. \quad (2)$$

To ensure the  $C^2$ -continuity over the whole domain  $\Omega$ , we need to properly select the basis functions according to the following criteria:

1.  $\phi_i(0) = 1$  and  $\phi_i(t) = 0$ ,  $\forall t \geq 1$ ;
2.  $\phi'_i(t) < 0$ ;
3.  $\phi''_i(t)$  is continuous for  $0 < t < 1$ ;
4.  $\phi'_i(0) = \phi'_i(1) = \phi''_i(0) = \phi''_i(1) = 0$ .

A fulfillment of the above requirements on basis functions guarantees the  $C^2$ -continuity in the domain to be warped, leading to a smooth deformation following the transformations of the handles. When the intrinsic distance is used to generate the input parameter  $t$  for the basis functions, linear blending-based deformations driven by these basis functions behave in a shape-aware manner. Another problem to be solved is how to determine the support size  $r_i$  of each basis function. As a basic requirement of handle-driven deformation based on linear blending, every point  $\mathbf{p} \in \Omega$  should be influenced by at least one handle. To be shape-aware, a point  $\mathbf{p}$  should be mostly affected by its closest handle in  $\Omega$ . Voronoi diagram sited at the origins of handles  $\{\mathbf{h}_i\}$  provides an intrinsic decomposition of  $\Omega$  according to these observations (see Figure 2(a)), where the intrinsic distance in  $\Omega$  is used as the metric for generating the Voronoi diagram. We denote the cell that corresponds to  $\mathbf{h}_i$  by  $(\mathbf{h}_i)$ . Two metrics with regard to a handle  $H_i$  can be defined as follows (see Figure 2(b) for an illustration):

- Size of a Voronoi cell:

$$r_d(\mathbf{h}_i) = \sup_{\mathbf{q} \in v(\mathbf{h}_i)} d(\mathbf{q}, \mathbf{h}_i);$$

- Separation to other sites:

$$r_h(\mathbf{h}_i) = \inf_{\mathbf{h}_{j(j \neq i)}} d(\mathbf{h}_i, \mathbf{h}_j).$$

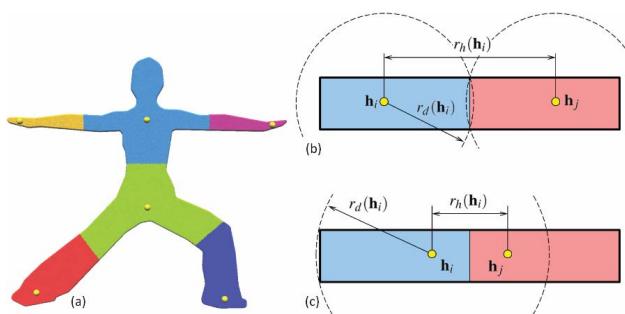


Fig. 2. Voronoi diagram based method to determine the size of local support. (a) The Voronoi diagram of handles can decompose  $\Omega$  into smaller pieces. (b) The illustration of  $r_h(\mathbf{h}_i)$  and  $r_d(\mathbf{h}_i)$  in the Voronoi diagram. (c) Very close handles can lead to  $r_h(\mathbf{h}_i) < r_d(\mathbf{h}_i)$ .

To let the basis function  $\phi_i(t)$  centered at  $H_i$  cover all points in  $\mathcal{V}(\mathbf{h}_i)$  and to ensure the handle interpolation property, it should have

$$r_d(\mathbf{h}_i) \leq r_i \leq r_h(\mathbf{h}_i). \quad (3)$$

The support size can be  $r_i = (1-\alpha)r_d(\mathbf{h}_i) + \alpha r_h(\mathbf{h}_i)$  with  $\alpha \in [0,1]$  being specified by users as a shape factor affecting the rigidity of deformation. For most of the examples in this paper,  $\alpha = 1$  is used. It is possible to have two handles too close to each other so that  $r_h(\mathbf{h}_i) < r_d(\mathbf{h}_i)$  (see Figure 2(c) for an example); in other words, the condition in Equation 3 could be violated. To solve such cases, we will exploit a virtual handle insertion algorithm (presented later).

## Properties of Weights

We analyze the properties of our weights for the handle-driven warping based on linear blending.

- **Smoothness:** The scalar field of weights must be smooth to avoid visual artifacts (discontinuity). The criteria we set on the basis functions ensure the  $C^2$ -continuity of our weights.
- **Interpolation:** The final transformations determined by the linear blending must interpolate the transformations at the handles. Specifically, the weight at each handle  $H_i$  is one at its origin, while basis functions centered at other handles give zero at this point. This is guaranteed by the locality and sparsity in our formulation.
- **Consistency:** When applying the same transformation  $\mathbf{T}$  on all handles, all points in  $\Omega$  must be consistently transformed by  $\mathbf{T}$ . This is enforced by the partition-of-unity property in our formulation. Another consistency requirement is about direction. The region influenced by a handle should not change in the inverse direction of the transformation assigned on the handle. We ensure this by the property of non-negativity.
- **Shape-awareness:** A shape-aware warping by the transformations of handles indicates that the influence of handles should be conformal to the shape features of the image, rather than the Euclidean distances to handles. As we use intrinsic distance to compute the weights on points by linear blending, shape-awareness is naturally guaranteed in our approach. All points will behave in a conformal manner with their relative handles.
- **Non-negativity:**  $\phi_i(t) \geq 0$  so that  $\forall \mathbf{p} \in \Omega$ ,  $w_i(\mathbf{p}) \geq 0$ . Moreover, when  $r_i \geq r_d(\mathbf{h}_i)$  is ensured for all handles, every point in  $\Omega$  should be covered by at least one handle's support. In other words,

$$\sum_{j=1}^m \phi_j(\cdot) \neq 0 .$$

- **Partition-of-unity:** This has been enforced by the formulation in Equation 2. That is,

$$\sum_{i=1}^m w_i(\mathbf{p}) = \sum_{i=1}^m \frac{\phi_i(d(\mathbf{p}, \mathbf{h}_i) / r_i)}{\sum_{j=1}^m \phi_j(d(\mathbf{p}, \mathbf{h}_j) / r_j)} \equiv 1 .$$

- **Locality/Sparsity:** This is preserved by  $\forall t \geq 1$ ,  $\phi_i(t) \equiv 0$  and the condition given in Equation 3. The transformation at a point coincident with a handle is only determined by the handle itself.  $\forall i \neq j$ ,  $\phi_j(\mathbf{h}_i) \equiv 0$ .
- **Nonlocal maxima/minima:** The global maximum of a weight  $w_i$  only occurs at the origin of handle  $H_i$  and the regions solely covered by the support of  $H_i$ . The global minima of a weight  $w_i$  only occurs in the regions not covered by the support of  $H_i$ . For the intersected regions of the supports of  $H_i$  and other handles, we observe this phenomenon in all our experiments. (This property is proved experimentally in all our tests. More results of weight distribution with isocurves are shown in the supplementary material.)

In short, our method preserves all the merits of prior methods for linear blending-based warping<sup>1,2,9</sup> while introducing new benefits of flexibility and efficiency.

The current weighting formulation has a major defect. The interpolation property cannot be preserved when the distance between two handles is too small while the regions to be covered by

either handle are large. Specifically, the interpolation of handles becomes an approximation when  $r_d(\mathbf{h}_i) \leq r_h(\mathbf{h}_i)$  in Equation 3 cannot be satisfied, which is not expected. To tackle this problem, we propose a virtual handle insertion algorithm in the following section.

## VIRTUAL HANDLE INSERTION

Our virtual handle insertion algorithm allows users to freely set handles within the domain of an image, with the help of which the interpolation property is always guaranteed and the shape-awareness of deformation is improved.

### Insertion algorithm

When  $r_d(\mathbf{h}_i) > r_h(\mathbf{h}_i)$ , there are points in the Voronoi cell  $\mathcal{V}(\mathbf{h}_i)$  whose distances to  $\mathbf{h}_i$  are larger than the minimal distance from  $\mathbf{h}_i$  to other handles. It is easy to observe that inserting new sites at the points  $\mathbf{h}_d \in (\mathbf{h}_i)$  with  $d(\mathbf{h}_d, \mathbf{h}_i) = r_d(\mathbf{h}_i)$  can reduce  $r_d(\mathbf{h}_i)$  while keeping  $r_h(\mathbf{h}_i)$  unchanged. Based on this observation, we develop a greedy algorithm for virtual handle insertion. Define  $\mathcal{H}$  as the set of handles and  $\delta(\cdot) = r_d(\cdot) - r_h(\cdot)$ . When there exists  $\mathbf{h}_i \in \mathcal{H}$  with  $\delta(\mathbf{h}_i) > 0$ , new virtual handles are inserted to resolve this problem by reducing

$$\max_{\mathbf{h}_i \in \mathcal{H}} \{\delta(\mathbf{h}_i) / r_d(\mathbf{h}_i)\}.$$

The pseudo-code for the algorithm **Virtual Handle Insertion** is described as follows:

1. **Input:** the set  $\mathcal{H}$  of real handles
2. **Output:** the expanded set  $\mathcal{H}$  with virtual handles
3. **while**  $\exists \mathbf{h}_i \in \mathcal{H}, \delta(\mathbf{h}_i) > 0$  **do**
4.     Find the handle  $\mathbf{h}_m = \arg \max_{\mathbf{h}_i \in \mathcal{H}} \delta(\mathbf{h}_i) / r_d(\mathbf{h}_i)$ ;
5.     Find a point  $\mathbf{p} \in (\mathbf{h}_m)$  with  $d(\mathbf{p}, \mathbf{h}_m) = r_d(\mathbf{h}_m)$ ;
6.     Insert a new virtual handle located at  $\mathbf{p}$  into  $\mathcal{H}$ ;
7.     Update the values of  $r_d(\cdot)$  and  $r_h(\cdot)$  on all handles;
8. **end while**
9. **return**  $\mathcal{H}$ ;

We know that inserting new handles in a Voronoi cell  $(\mathbf{h}_i)$  with  $\delta(\mathbf{h}_i) > 0$  can reduce the value of  $\delta(\mathbf{h}_i)$ . However, inserting a new site  $\mathbf{h}_d \in (\mathbf{h}_i)$  can also affect other handles ( $\mathbf{h}_j$  with  $j \neq i$ ). In an extreme case, the original  $\delta(\mathbf{h}_j) \leq 0$  could be turned into  $\delta(\mathbf{h}_j) > 0$ . Then, more virtual handles need to be added into  $(\mathbf{h}_j)$ .

Our virtual handle insertion algorithm can be considered a variant of the farthest point sampling algorithm,<sup>17</sup> which tends to tessellate a domain into a Voronoi diagram with neighboring Voronoi cells having similar sizes. The condition of  $r_d(\cdot) \leq r_h(\cdot)$  is satisfied on all handles when this is the case. Our experimental tests also verify this observation (see Figure 3 for an example).

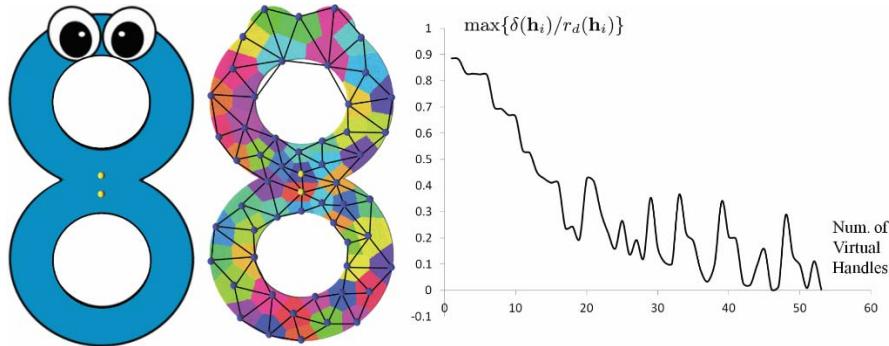
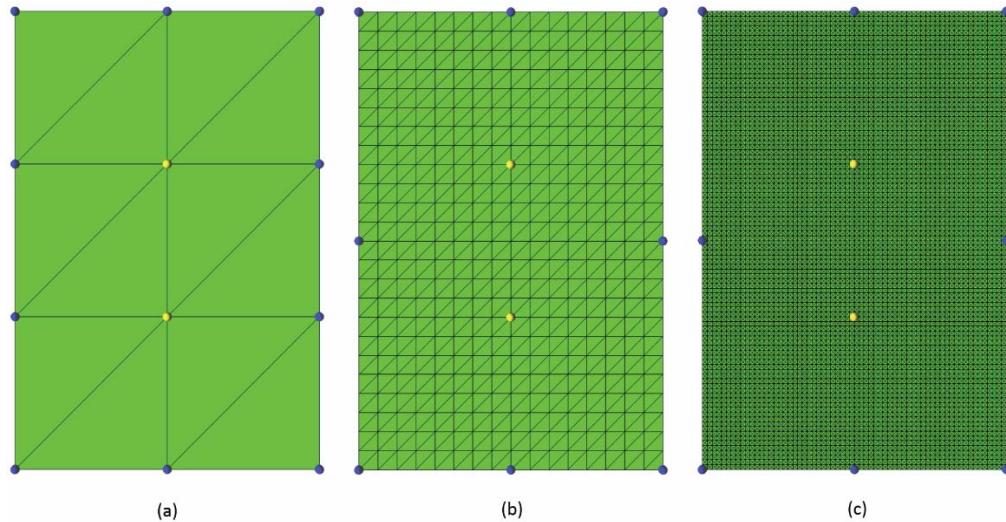


Figure 3. (Left image) When two handles are too close to each other, the condition for interpolation ( $r_d(\cdot) < r_h(\cdot)$ ) can only be satisfied after inserting virtual handles. (Middle image) The newly inserted virtual handles (in blue) tessellate the deformation domain into Voronoi cells whose areas are similar to neighboring cells. The Delaunay graph, ( $\mathcal{H} \cup \mathcal{H}^V$ ), of the Voronoi diagram is also shown—see the network linking the handles. (Right image) The score,  $\max\{\delta(\mathbf{h}_i)/r_d(\mathbf{h}_i)\}$ , of our Virtual Handle Insertion algorithm drops while inserting virtual handles.

*Remarks:*

(1) We observe the convergence of virtual handle insertion in all our experiments, and many results provided in this paper and the supplementary material show the final distribution of handles (user-input and virtual ones). The insertion process might fail if the mesh density is not compared enough with the closeness of user-input handles, especially when there is no point between two handles on the mesh (which indicates the current sampling rate gives a very poor approximation of the domain to be warped and an up-sampling step is required to refine this approximation – see an illustration in Figure 4.) We can easily detect such a case and conduct a dynamic up-sampling step in our implementation. Interestingly, it is worth pointing out that the “failure” of the insertion process here means all points in the domain will be selected as virtual handles, rather than a dead loop leading to no response in our implementation, which indicates a termination guarantee of the insertion process in a discrete sense. Meanwhile, the determination of weights on virtual handles degenerates to solve a harmonic problem over the whole underlying mesh, which will be detailed in the Transformation on Virtual Handles section.

(2) It is noted that it lacks a complete proof on the convergence of virtual handle insertion here to provide a theoretical foundation for this algorithm. More interestingly, we find this can be generalized to a spatial partition problem in  $n$  dimensional space. Therefore, we propose a conjecture to be proved in the sidebar at the end of the article, and we would like to invite future efforts on it.



**Figure 4.** The distribution of virtual handles (in blue) with two point handles (in yellow). (a) For an extremely coarse mesh, all points are selected as handles (real and virtual), and a Voronoi diagram based on handles cannot be built. (b,c) For dense meshes, the virtual handles are located on the meshes by satisfying the prescribed conditions.

## Transformation on Virtual Handles

The problem is how to determine the transformations on virtual handles according to the user-specified transformations on real handles. We denote the set of real handles as  $\mathcal{H}$  and the set of virtual handles as  $\mathcal{H}^V$ . As aforementioned, the handles of  $\mathcal{H} \cup \mathcal{H}^V$  can partition the given domain  $\Omega$  into a Voronoi diagram  $Vor(\mathcal{H} \cup \mathcal{H}^V)$ . A dual graph of  $Vor(\mathcal{H} \cup \mathcal{H}^V)$  can be constructed by 1) using the sites of every Voronoi cell as nodes and 2) linking the sites of every two neighboring Voronoi cells by lines as edges, which is a Delaunay graph. We denote the Delaunay graph as  $(\mathcal{H} \cup \mathcal{H}^V)$  and use the symbol  $H$  to represent nodes (real or virtual handles) in  $\mathcal{D}\mathcal{G}$ . The transformations of handles in  $\mathcal{H}^V$  are determined as follows:

- For each real handle  $H_i$  in  $\mathcal{H}$ , a harmonic field  $\varpi_i(\cdot)$  is computed on  $\mathcal{D}\mathcal{G}$  to assign a field value  $\varpi_i(H_g)$  to each of the other handles. Boundary conditions,  $\varpi_i(H_i)=1$  and  $\varpi_i(H_{j \neq i})=0$ , are given to compute the harmonic field  $\varpi_i(\cdot)$ . If there are  $m$  handles in  $\mathcal{H}$ ,  $m$  harmonic fields are determined on  $\mathcal{D}\mathcal{G}$ .
- After converting the transformation  $\mathbf{T}_i$  of each real handle into a rotation quaternion  $\mathbf{q}_i$  and a translation vector  $\mathbf{t}_i$ , the rotation and the translation on a virtual handle  $Hv \in \mathcal{H}^V$  can be determined by

$$(\mathbf{q}_v) = \frac{1}{\varpi_{sum}(H_v)} \sum_{H_i \in \mathcal{H}} \varpi_i(H_v) (\mathbf{q}_i) \quad (4)$$

$$\text{with } \varpi_{sum}(\cdot) = \sum_{H_i \in \mathcal{H}} \varpi_i(\cdot) .$$

- Finally, the quaternion and the translation determined on each virtual handle are converted back into a transformation matrix to be used in linear blending.

The transformations of virtual handles determined in this way improve the warping quality. As illustrated in Figures 5 and 6, deformation of the whole image driven by the transformations on both real and virtual handles is very natural. The influence of a real handle decays when the intrinsic distance to it increases.

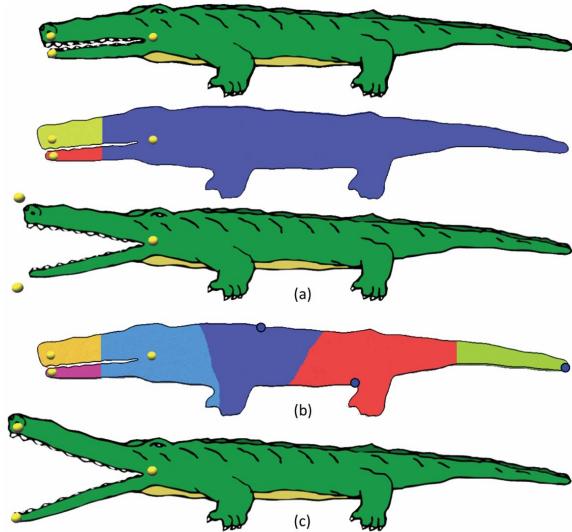


Figure 5. A handle covering a large region can affect the interpolation on its nearby handles. (a) The right handle's Voronoi cell covers the whole right part of the alligator, which leads to a value of  $r_d(\cdot)$  that is much larger than  $r_h(\cdot)$ . In this case, transformations at the left two handles cannot be interpolated. (b) Virtual handles (in blue color) are added to resolve the problem by the insertion algorithm. As a result, the domain to be deformed has been decomposed into smaller Voronoi cells with handles (real and virtual) as sites. (c) The deformation result is driven by both the real and the virtual handles, where the transformations at real handles are interpolated.

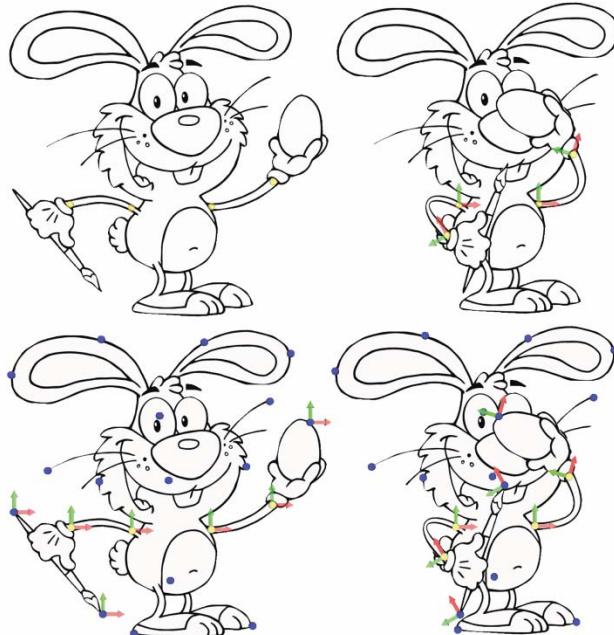


Figure 6. The deformation of a rabbit is driven by four real handles (yellow dots). The result of deformation is determined with the help of virtual handles (blue dots). The transformations at handles (both real and virtual) are illustrated by frames.

## IMPLEMENTATION DETAILS

### Basis Function

Previously, we listed the required properties for basis functions to achieve the  $C^2$ -continuity of weights. In our implementation, Bézier polynomial is employed for the function  $\phi_i(t)$  so that the constraints for  $C^1$ - and  $C^2$ -continuity at the boundary of the supporting regions can be satisfied. To ease the evaluation and analysis, each  $\phi_i(t)$  is represented as the  $y$ -component (in other words,  $\phi_i(t) = \mathbf{b}^y(t)$ ,  $t \in [0, 1]$ ) of a 2D Bézier curve with degree- $n$  ( $n \geq 5$ )

$$\mathbf{b}(t) = \sum_{i=0}^n \mathbf{b}_i B_{i,n}(t) , \quad (5)$$

where  $B_{i,n}(t)$  are the Bernstein polynomials. From the property of Bézier curves, we have

$$\mathbf{b}'(0) = n(\mathbf{b}_1 - \mathbf{b}_0), \mathbf{b}'(1) = n(\mathbf{b}_n - \mathbf{b}_{n-1})$$

$$\mathbf{b}''(0) = n(n-1)(\mathbf{b}_2 - 2\mathbf{b}_1 + \mathbf{b}_0)$$

$$\mathbf{b}''(1) = n(n-1)(\mathbf{b}_n - 2\mathbf{b}_{n-1} + \mathbf{b}_{n-2})$$

for a Bézier curve in  $n$ -th order. Incorporating the constraints in the Formulation section, we have

$$\mathbf{b}_1 = \mathbf{b}_0, \mathbf{b}_n = \mathbf{b}_{n-1}, \mathbf{b}_1 = \frac{\mathbf{b}_0 + \mathbf{b}_2}{2}, \mathbf{b}_{n-1} = \frac{\mathbf{b}_n + \mathbf{b}_{n-2}}{2} .$$

As we already set  $\mathbf{b}_i^x = i/n$  to let  $x = t$ , it is not difficult to find that  $\mathbf{b}_0^y = \mathbf{b}_1^y = \mathbf{b}_2^y = 1$  and  $\mathbf{b}_n^y = \mathbf{b}_{n-1}^y = \mathbf{b}_{n-2}^y = 0$  satisfy all these constraints. For the rest control points, we can simply assign them as 0.5 or align them along the line  $\mathbf{b}_2 \mathbf{b}_{n-2}$  uniformly.

### Intrinsic Distance Approximation

Generally, we sample the input domain  $\Omega$  (the input image) to be deformed into a dense set of points  $\mathcal{P}$ . By searching for  $k$ -nearest-neighbors of each point, a graph  $\mathcal{G}(\mathcal{P})$  spanning  $\Omega$  (in discrete form) can be established by using points in  $\mathcal{P}$  as nodes and adding links between neighboring points as edges. Note that user-specified handles should also be added into  $\mathcal{P}$  to construct the graph ( $\mathcal{H} \subset \mathcal{P}$ ). The intrinsic distance from any point  $\mathbf{q} \in \mathcal{P}$  to a handle is approximated by the distance between  $\mathbf{q}$  and the handle on the graph, which can be computed efficiently with Dijkstra's algorithm. Also, the Voronoi diagram  $Vor(\mathcal{H})$  can be obtained by Dijkstra's algorithm with multiple sources on  $\mathcal{G}(\mathcal{P})$ , where each sample is assigned to a Voronoi cell. To determine the weights on a general point  $\mathbf{p} \in \Omega$  that is not a sample in  $\mathcal{P}$ , a linear blending based on reciprocal distance weights<sup>18</sup> is employed to obtain the weight on  $\mathbf{p}$  from its  $k$ -nearest-neighbors in  $\mathcal{P}$ . One strategy has more sophisticated parameterization strategies,<sup>18</sup> which can also be applied here. With the help of this meshless parameterization, we can easily take an up-sampling step in the domain  $\Omega$  when the point set  $\mathcal{P}$  becomes sparse when applying a drastic deformation.

As a special case, intrinsic-distance evaluation for pixelized images or convex domain is simplified into the computation of Euclidean distance between two points as the whole domain is convex, which can be computed very fast. This also ensures the computational efficiency of our algorithm.

### Harmonic Field

After using the virtual handle insertion algorithm to generate a set of new handles, harmonic fields are computed on a dual graph of  $Vor(\mathcal{H})$  to determine the transformations on virtual handles. By our boundary condition, all field values are non-negative when uniform Laplacian is employed. In other words, the coefficients used in Equation 4 are non-negative. Instead of solving a linear system to compute the harmonic field, we initially assign the field values on all real

handles as one and the weights on all virtual handles are set as zero. Then we apply Laplacian operators to update their field values iteratively. The field values on virtual handles can be efficiently obtained after tens of iterations.

## Interactive Handles

The point handles can be generalized to different types of bar handles, such as line segments and polygons. Unluckily, such extension is not straightforward as 1) the full boundary of an image to be warped is often expected to be strictly fixed when applying a deformation inside the domain and 2) the transformations of points on any handles should respect the interpolation property introduced in the Properties of Weights section. It also needs to be processed carefully when a handle is very close to another. In our implementation, the generalization of handles is realized by taking advantage of the properties of the Voronoi diagram (see Figure 7). The bar handles are discretized into a dense set of bar samples to ensure that all points around the bar handles should belong to the Voronoi sites centered at those discrete bar samples, which leads to perfect interpolation on the bar handles. In practice, we determine the set of bar samples for a bar handle by uniformly dividing it into  $\lceil L_{bar}/d \rceil$  segments, where  $L_{bar}$  is its length and  $d$  refers to the minimum Euclidean distance between any handles (point and bar handles) over the whole domain of computation (refer to Figure 8(a)). The transformations of bar samples are calculated as linear combinations of the transformations at bar handles' end points.

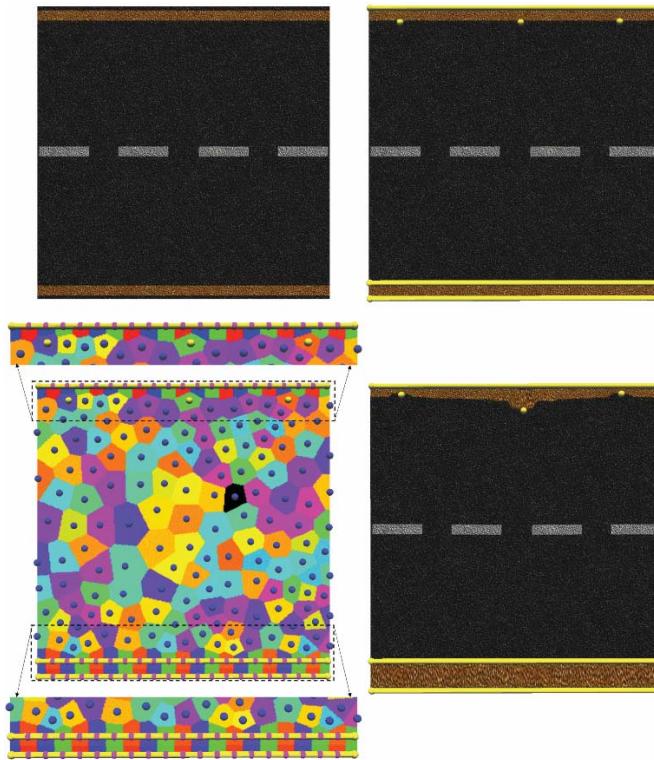
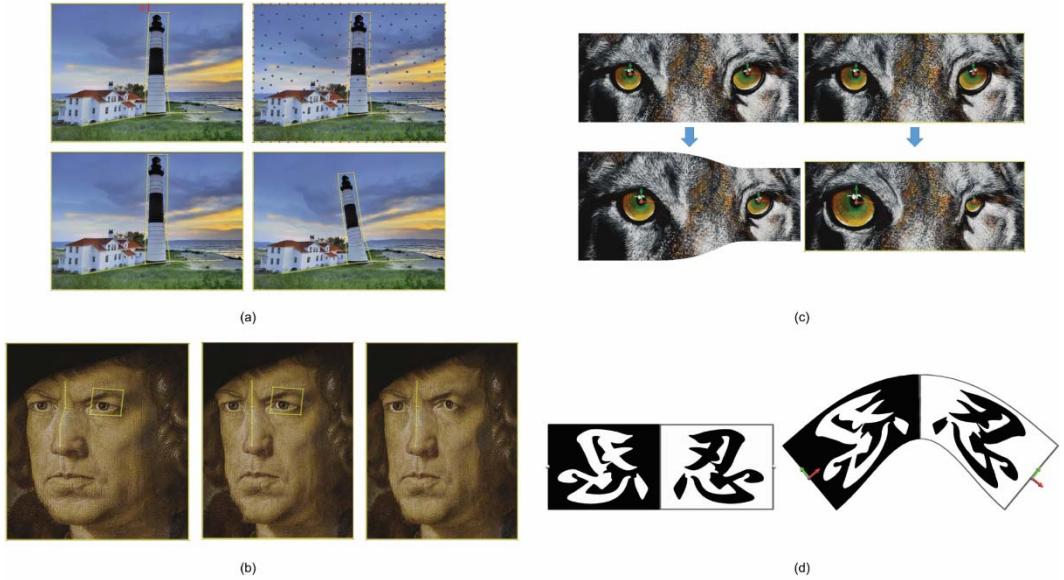


Figure 7. An example of warping an image with point and bar handles. As shown in the Voronoi diagram based on the distribution of virtual handles and the tessellation of bars, the bar handles dominate the regions around it, leading to perfect interpolation about the movement of handles.



**Figure 8. Experimental results.** (a) An example of manipulating the tower while fixing the image boundary. (b) A portrait edited by bar handles with salient feature inside the closed loop of bar handles at the left eye preserved. (c) A comparison of image editing with and without fixing its boundary. (d) An example with symmetric transformations on two symmetric handles to deform a symmetric domain.

## RESULTS

Our weighting method provides a compact tool to assign continuous weights for all points in the domain to be warped. With the help of sophisticated techniques for assigning transformations on the handles (such as the pseudo-edge method<sup>1</sup>), a natural user interface for image warping can be achieved.

We have tested this approach on a variety of examples using both the point and bar handles. Figures 5 and 6 have already demonstrated the functionality of point handles. Especially in Figure 5, the scheme of virtual handles insertion guarantees the interpolation at real handles. Figure 6 illustrates the effectiveness of our method in determining transformations on virtual handles. The example of using bar handles to warp the shape of a tower has been shown in Figure 8(a). Another example is given in Figure 8(b) to deform a portrait. Figure 8(c) shows a comparison of image manipulation with and without fixing boundary. To obtain natural bending results, we can add rotations on handles by heuristics (such as the pseudo-edge method<sup>1</sup>). The example in Figure 8(d) demonstrates the performance of our approach with a symmetric deformation. When deforming a symmetric domain by adding symmetric transformations on the symmetric handles, it is expected to get a symmetric warped result. This property is strictly preserved by our formula.

In prior mesh-based approaches, the numerical system must be solved once more when new handles are inserted. By using our weighting formulation, the time cost of adding new handles is very trivial as the weights are determined in a closed-form. Table 1 lists the statistics of our approach on different examples, where  $|\mathcal{H}_b|$  and  $|\mathcal{H}_p|$  denote the number of bar and point handles,  $|\mathcal{V}|$  stands for the number of virtual handles added in the domain, and  $|\mathcal{S}|$  represents the number of points used in the computation. The  $t_{rd}$  and  $t_w$  columns state the time used in the computation of region decomposition (including both Voronoi diagram computation and intrinsic-distance evaluation) and weights respectively. All the tests are conducted on a laptop with Intel Core i7-3740QM CPU at 2.70 GHz and 8 GB memory. Our current version of implementation only uses a single core. As shown in the table, the computational bottleneck is the computation of region decomposition that is closely related with the number of points used in the domain and the type and placement of handles. But compared with prior approaches, our method has achieved significant improvement in computational efficiency. Table 2 shows a comparison with a biharmonic

method<sup>1</sup> using the same setup as the Gingerman model in Figure 1. We fail to test the bi-harmonic method<sup>1</sup> on a mesh having 617,000 vertices with the public program provided by the authors, as there are too many vertices. To compare our technique with another approach,<sup>15</sup> both approaches require reinitialization when inserting a new handle in the computational domain. It roughly takes seconds (using Euclidean distance) to minutes (using geodesic distance) to compute interior distances.<sup>15</sup> In our experiments with a 3D bar and armadillo model shown in the accompanying video, the time costs for precomputation are about 0.1 second (using Euclidean distance) and 1.5 minutes (using geodesic distance) respectively.

**Table 1. Example statistics.**

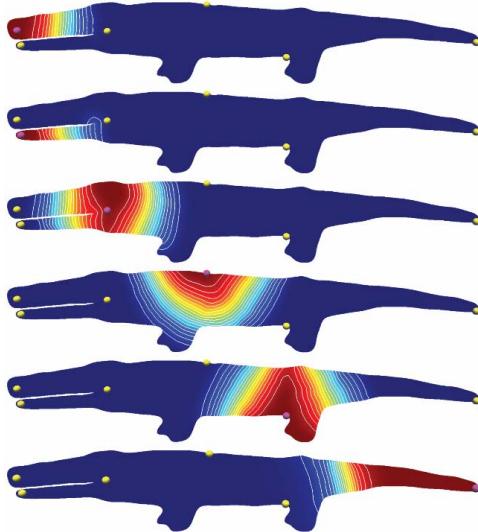
|           | $ \mathcal{H}_b $ | $ \mathcal{H}_p $ | $ \mathcal{V} $ | $ \mathcal{S} $ | $t_{rd}$ (sec.) | $t_w$ (sec.) |
|-----------|-------------------|-------------------|-----------------|-----------------|-----------------|--------------|
| Tower     | 12                | 0                 | 113             | 7,834           | 3.75            | 0.013        |
| Wolf      | 4                 | 2                 | 1               | 6,694           | 0.020           | 0.001        |
| Portrait  | 7                 | 0                 | 36              | 9,438           | 0.380           | 0.006        |
| Alligator | 0                 | 3                 | 3               | 9,456           | 0.409           | 0.002        |
| Rabbit    | 0                 | 4                 | 16              | 22,972          | 1.700           | 0.434        |
| Road      | 3                 | 3                 | 124             | 13,212          | 3.18            | 0.002        |

**Table 2. Comparison with different resolutions.**

| Number of vertices | Time of computation (sec.) |            |
|--------------------|----------------------------|------------|
|                    | Biharmonic                 | Our method |
| 627                | 0.818                      | 0.024      |
| 31,325             | 8.691                      | 0.110      |
| 619,649            | N/A                        | 20.796     |

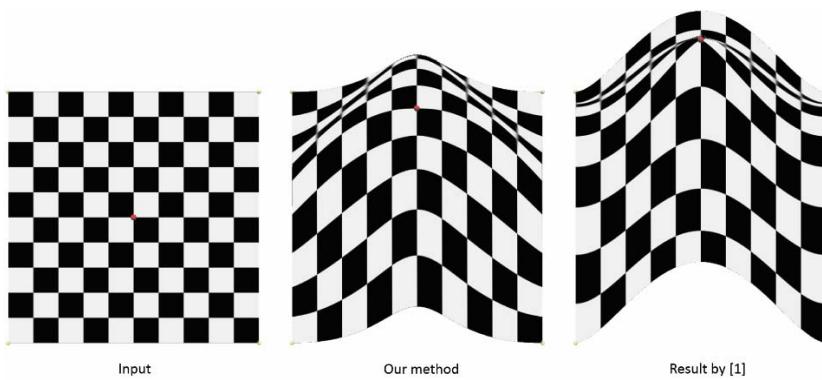
*Limitations and discussion.* When using the weighting formulation presented in this paper to deform images, sample points are adopted as the medium for realizing the computation. The error-bound of computation on this discrete representation is controlled by the density of samples. However, during the process of a sequence of deformations, the density of points could be changed dramatically. In this sense, a dynamic up-sampling step should be integrated in the framework to preserve the error-bound of intrinsic-distance computation. Super-sampling techniques or texture mapping on a mesh can be exploited in image editing applications. In our framework, the time cost of weight evaluation is trivial after resampling. The bottleneck is the computation of intrinsic distances on the sample points. Our current implementation is based on the Dijkstra's algorithm. This shortest path problem with multiple sources can be computed in parallel by using graphics accelerated hardware, which can result in a significant speed-up.

Our formulation gives global maxima at the positions of handles according to the interpolation property. For a shape-aware deformation, it is also demanded having nonlocal maxima/minima anywhere else in the domain. This has been verified in our experimental tests. We check the topology of isolines on the fields of weights (see Figure 9 for an example). If there is a closed loop formed by isolines of  $w_i(\cdot)$  at one place except the center of the handle  $\mathbf{h}_i(\cdot)$ , a local maximum/minimum is generated there. No such case is found in all our examples. More examples for verification can be found in the supplementary material. It should be admitted that a theoretical proof for nonlocal maxima/minima property is lacking here as a technical guarantee.



**Figure 9.** The verification of nonlocal maxima/minima is taken by analyzing the topology of isolines on the scalar fields of weights. The handles (real and virtual) in this example are the ones shown in Figure 5.

The deformations driven by linear blending are not always injective and therefore can generate the results with fold-overs and self-intersections. Recently, some research has been conducted in this thread to produce injective mappings,<sup>19</sup> which are mainly mesh-based. In a function-based formulation, the injectivity of a mapping can be checked by the positiveness of Jacobian. But the interpolation property is not preserved anymore in those methods trying to hold the injectivity of mappings. We argue that a simultaneous satisfaction of both the interpolation and injectivity property is paradoxical. Other methods preserving the interpolation property also violate the injectivity as our approach.<sup>20</sup> We will unavoidably see fold-overs in some warped results if the handles are dragged too much (see Figure 10 for an example).



**Figure 10.** Self-intersection can be found for the methods respecting handle interpolation.

As analyzed in the Formulation section, when inserting a new handle, the distribution of virtual handles needs to be recomputed to ensure that no violation of condition 3 happens in the whole domain, which is the current computational bottleneck of this proposed method. A dynamic mechanism can be realized in practice to perform a resampling of virtual handles only if unsatisfaction of condition 3 is detected as a trick to reduce computational overhead. As indicated in Tables 1 and 2, the time cost for determination of virtual handles and region decomposition is negligible for small-scale meshes and trivial for large-scale ones.

## CONCLUSION AND FUTURE WORK

In this work, we present a method to efficiently determine weights of linear blending for image warping. Our formulation is in a closed-form and can be easily used in a variety of applications. Equipped with the virtual handle insertion algorithm, good properties of weights generated by prior mesh-based methods can all be preserved in this approach. A variety of examples have been shown to demonstrate the effectiveness of our approach.

There are some potential improvements for our technique. First, the tessellation of bar handles is simple in the current version. As it will affect the result of warping, we are curious about finding a locally adaptive and non-uniform strategy to further strengthen our result. A similar scenario is about the density of underlying mesh. Although the validity of our method about mesh density applies to infinity, the distribution of real and virtual handles will converge when the mesh is fine enough, indicating the possible existence of a density threshold, which can help reduce the computational overhead of this approach (see Figure 4(b)(c)). Second, it is interesting to investigate how to resolve the problem when self-intersection is detected, which will be one of our future works. We may restrict the movement of handles to avoid the occurrence of fold-overs. Besides, only a linear-blending scheme in 2D is tested in the paper. Actually, our method can be extended directly to 3D models for shape deformation without any modifications (see two examples in the accompanying video). Lastly, we plan to further apply the weights generated in this approach to more advanced skinning applications, such as the blending of two rigid motions resulting in a rigid motion, which is a very important property when deformation of articulated characters is computed by the skinning methods.

## SIDE BAR: CONJECTURE

Following the definition of intrinsic distance  $d(\cdot, \cdot)$  in the Intrinsic Distance section, we consider the partition of a connected closed domain  $\mathcal{D} \subset \mathbb{R}^n$  into cells. We first give some fundamental definitions:

- Anchor points  $\mathcal{A} = \{\mathbf{h} | \mathbf{h} \in \mathcal{D}\}$  are a set of points that are predefined in the domain  $\mathcal{D}$ .
- Size of a cell  $\mathcal{V}$  centered at  $\mathbf{h}_i$  is defined as  $r_d(\mathbf{h}_i) = \sup_{\mathbf{p} \in \mathcal{V}(\mathbf{h}_i)} d(\mathbf{p}, \mathbf{h}_i)$ .
- Separation of  $\mathbf{h}_i$  to other cells is defined as  $r_h(\mathbf{h}_i) = \inf_{\mathbf{h}_{j(j \neq i)}} d(\mathbf{h}_i, \mathbf{h}_j)$ .

The domain  $\mathcal{D}$  is partitioned with regard to a set of anchor points in a Voronoi-diagram-like manner. Specifically, a point  $\mathbf{p}$  belongs to a cell centered at  $\mathbf{h}_i$  if  $d(\mathbf{p}, \mathbf{h}_i) \leq d(\mathbf{p}, \mathbf{h}_j), j \neq i$ .

Conjecture: For any given set of anchor points  $\mathcal{A}$ , there exists at least an expanded set  $\mathcal{H}$  based on  $\mathcal{A}$  ( $\mathcal{A} \subseteq \mathcal{H}$ ), with regard to which the partition of  $\mathcal{D}$  into cells satisfies the following non-intrusive condition:  $r_d(\mathbf{h}_i) \leq r_h(\mathbf{h}_i), \mathbf{h}_i \in \mathcal{H}$ .

## ACKNOWLEDGMENTS

We would like to thank Dr. Yang Liu with Microsoft Research for his insightful suggestions regarding this work. We also feel thankful for the constructive comments from all reviewers on the improvement of this manuscript. This work is partially supported by Hong

Kong RGC/GRF (No. 14207414), NSFC (No. 61300136), NSF of Guangdong (No. 2015A030313220, S2013020012795), and Doctoral Fund of Ministry of Education of China (No. 20130172120010).

## REFERENCES

1. A. Jacobson et al., “Bounded biharmonic weights for real-time deformation,” *ACM Trans. Graph.*, vol. 30, no. 4, 2011, p. 78:1.
2. A. Jacobson, T. Weinkauf, and O. Sorkine, “Smooth shape-aware functions with controlled extrema,” *Comp. Graph. Forum*, vol. 31, no. 5, 2012, pp. 1577–1586.
3. K. Singh and E. Fiume, “Wires: A geometric deformation technique,” *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques* (SIGGRAPH 98), 1998, pp. 405–414.
4. W. von Funck, H. Teisel, and H.-P. Seidel, “Vector field based shape deformations,” *ACM Trans. Graph.*, vol. 25, no. 3, 2006, pp. 1118–1125.
5. R.W. Sumner, J. Schmid, and M. Pauly, “Embedded deformation for shape manipulation,” *ACM Trans. Graph.*, vol. 26, no. 3, 2007.
6. M. Botsch and O. Sorkine, “On linear variational surface deformation methods,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 1, 2008, pp. 213–230.
7. M. Botsch et al., “Adaptive space deformations based on rigid cells,” *Computer Graphics Forum*, vol. 26, no. 3, 2007, pp. 339–347.
8. O. Sorkine and M. Alexa, “As-rigid-as-possible surface modeling,” *Proceedings of the Fifth Eurographics Symposium on Geometry Processing* (SGP 07), 2007, pp. 109–116.
9. A. Jacobson and O. Sorkine, “Stretchable and twistable bones for skeletal shape deformation,” *ACM Trans. Graph.*, vol. 30, no. 6, 2011, p. 165:1.
10. S. Schaefer, T. McPhail, and J. Warren, “Image deformation using moving least squares,” *ACM Trans. Graph.*, vol. 25, no. 3, 2006, pp. 533–540.
11. P. Joshi et al., “Harmonic coordinates for character articulation,” *ACM Trans. Graph.*, vol. 26, no. 3, 2007.
12. M. Ben-Chen, O. Weber, and C. Gotsman, “Variational harmonic maps for space deformation,” *ACM Trans. Graph.*, vol. 28, no. 3, 2009, p. 34:1.
13. S.-M. Hu et al., “Direct manipulation of ffd: efficient explicit solutions and decomposable multiple point constraints,” *The Visual Computer*, vol. 17, no. 6, 2001, pp. 370–379.
14. M. Botsch and L. Kobbelt, “Real-time shape editing using radial basis functions,” *Computer Graphics Forum*, vol. 24, no. 3, 2005, pp. 611–621.
15. Z. Levi and D. Levin, “Shape deformation via interior rbf,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 7, 2014, pp. 1062–1075.
16. Y. Wang et al., “Linear subspace design for real-time shape deformation,” *ACM Trans. Graph.*, vol. 34, no. 4, 2015, p. 57:1.
17. Y. Eldar et al., “The farthest point strategy for progressive image sampling,” *IEEE Transactions on Image Processing*, vol. 6, no. 9, 1997, pp. 1305–1315.
18. M.S. Floater and M. Reimers, “Meshless parameterization and surface reconstruction,” *Computer Aided Geometric Design*, vol. 18, no. 2, 2001, pp. 77–92.
19. C. Schüller et al., “Locally injective mappings,” *Proceedings of EUROGRAPHICS/ACM SIGGRAPH Symposium on Geometry Processing*, 2013, pp. 125–135.
20. Y. Lipman, “Bounded distortion mapping spaces for triangular meshes,” *ACM Transactions on Graphics*, vol. 31, no. 4, 2012, p. 108.

## ABOUT THE AUTHORS

**Chuhua Xian** is an associate professor in the School of Computer Science and Engineering at South China University of Technology. He has a PhD from the State Key Lab of CAD&CG at Zhejiang University. He has been a visiting scholar at the Chinese University

of Hong Kong, and his current research interests are computer graphics, geometry modeling and processing, and CAD/CAE integration. Contact him at chhxian@scut.edu.cn.

**Shuo Jin** is a PhD candidate in the Department of Mechanical and Automation Engineering at the Chinese University of Hong Kong. He has a BS in information and computing science from the Department of Mathematics at Zhejiang University. His current research interests are geometric modeling and processing and their real-world applications. Contact him at sjin@mae.cuhk.edu.hk.

**Charlie C.L. Wang** is a professor and chair of Advanced Manufacturing in the Department of Design Engineering at Delft University of Technology in the Netherlands. He has a PhD in mechanical engineering from the Hong Kong University of Science and Technology. He is an ASME Fellow, and his research interests are geometric modeling, design and manufacturing, and computational physics. Contact him at c.c.wang@tudelft.nl.

# Parametric Reshaping of Portrait Images for Weight-change

**Haiming Zhao**  
Zhejiang University

**Xiaogang Jin**  
Zhejiang University

**Xiaojian Huang**  
Zhejiang University

**Menglei Chai**  
Zhejiang University

**Kun Zhou**  
Zhejiang University

We present an easy-to-use parametric image retouching method for thinning or fattening a face in a single portrait image while maintaining a close similarity to the source image. First, our method reconstructs a 3D face from the input face image using a morphable model. Second, according to the linear regression equation derived from the depth statistics of the soft tissue in the face and the user-set parameters of weight-change degree, we calculate the new positions of the feature points. The Laplacian

deformation method is then used for non-feature points in the 3D face model. Our model-based reshaping process can achieve globally consistent editing effects without noticeable artifacts. We seamlessly blend the reshaped face region with the background using image retargeting method based on mesh parametrization. The effectiveness of our algorithm is demonstrated by experiments and user study.

Faces are essential to make a first impression, consciously or unconsciously. Facial appearance is also vital for communication. Beautiful faces are pleasurable to look upon.<sup>1</sup> Since facial shape is an important determinant of beauty, it can be desirable to modify a face to be fatter or thinner in order to be more attractive. To accomplish this, a facial weight-change simulator is needed to measure model growth and shape modification. Potential applications of this simulator are not limited to the beauty and medical industries. It also plays an important role in digital entertainment, and film and television production.

Compared to other images, processing of facial images is particularly delicate. People are also relatively good at determining the smallest differences in the appearance of a face. Photo retouching is able to present convincing adjusted faces while maintaining the natural appearance of the face (see Figure 1). However, this time-consuming work must generally be performed by a

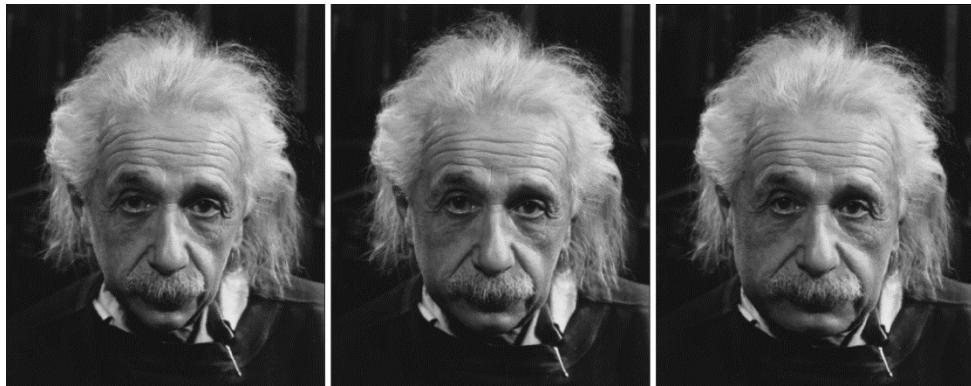


Figure 1. Our parametric facial reshaping method automatically simulates the weight-change of a 2D portrait image and generates a fatter or thinner face as intended. (Middle) is the original input image of Albert Einstein; (left) is the result of weight-change degree -2, which indicates losing weight by 2 degrees; (right) is the result of weight-change degree +2, which implies gaining weight by 2 degrees.

skilled, talented retouching artist. Since retouching is experience-based, the result relies heavily on the users' preference and effort. The process is also not parametric, which makes it especially difficult to control the degree of weight change.

The most related work to ours is proposed by Danino and colleagues who proposed a parametric 2D facial weight-change simulator based on 2D empirically knowledge.<sup>2</sup> This method can gener-

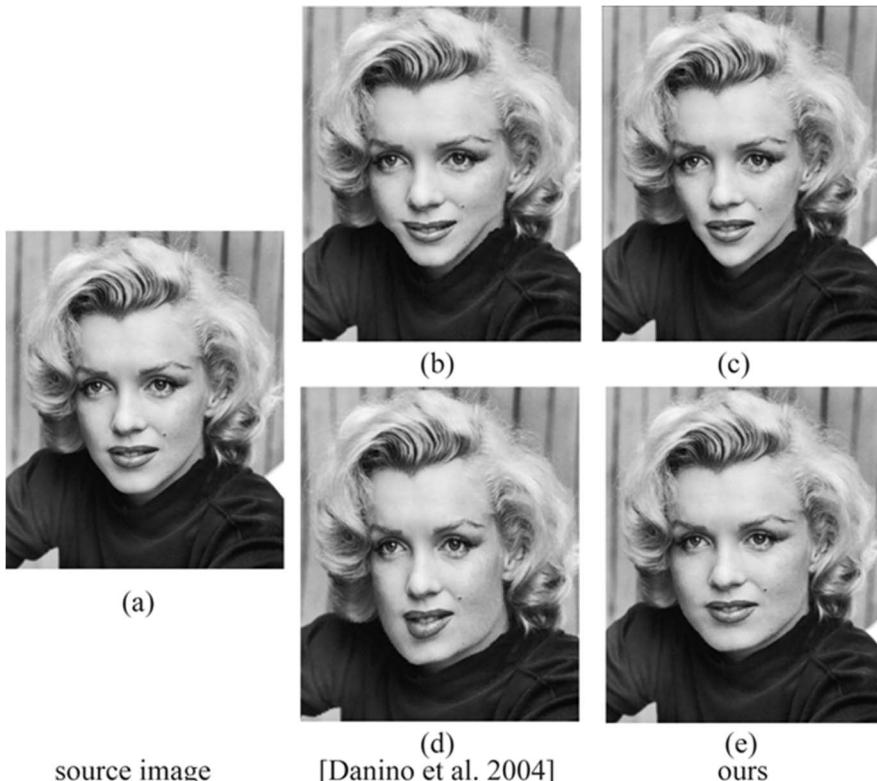


Figure 2. Comparison results. (a) is the input image. (b) and (d) are the results of Danino and colleagues' method,<sup>2</sup> while (c) and (e) are our results.

ate realistic results when the input face is frontal with neutral facial expressions. However, this method does not use the semantic information of the underlying face model, and the background is simply warped without considering the contents of the image. As a result, it may introduce obvious artifacts when the weight-change is large (see Figure 2). Another related work is introduced by Zhou and colleagues<sup>3</sup> who proposed an image retouching technique for realistic reshaping of human bodies in a single image. This model-based approach can create desired reshaping effects by changing the degree of reshaping which characterizes a small set of semantic attributes. However, it cannot deal with facial reshaping directly. Moreover, it relies on a large 3D whole-body morphable model which may limit its application.

Inspired by the work of Zhou and colleagues,<sup>3</sup> we present an image-based facial reshaping method using a linear regression equation. The weight change deformation of a face is parameterized by adjusting BMI (Body Mass Index) values.<sup>4</sup> We first reconstruct a 3D face model from the input 2D image using a morphable model,<sup>5</sup> and label the feature points on the 3D face model. Then, we calculate the deformed positions of feature points according to the weight-change degree related to BMI. After that, we generate the deformed 3D face model using the Laplacian deformation method, and project it into 2D image as the deformed face region. With the help of content-aware image retargeting approach by Guo and colleagues,<sup>6</sup> we finally blend the deformed face region and the background to obtain the reshaped 2D image.

The contributions of this work are: (i) a novel geometric weight-change simulator is presented, which is automatic, fast, and robust; (ii) parametric deformation of the face caused by varying BMI is based on a reliable face tissue depth database, which leads to a reshaped face in compliance with life experience and the repeatable reshaping process.

## RELATED WORK

### Weight-change Simulator

Few approaches to weight-change simulation have been proposed during the past decades. It first appeared in the innovative work of Blanz and Vetter.<sup>5</sup> The morphable 3D face model was built on hundreds of 3D face scans. Certain features, including weight, were manually labeled and mapped to the parameter space. Thus, weight-change simulation could be achieved by adjusting weight parameters. However, the simulated result is greatly affected by the constraints of the database. If the reshaping parameter is beyond the scope of the database, the reshaped face is probably unsatisfactory. Moreover, the hair part of the image is particularly problematic. Danino and colleagues presented a facial weight-change simulator for 2D images.<sup>2</sup> The face region is divided into regions characterized by different weight-change patterns. Its overall process is fast and robust, and the results are clear, sharp, and realistic. Nevertheless, the transformation between the original part and modified face parts is empirically defined without considering the semantic information of the underlying face model. In addition, the input images are limited to frontal face images with neutral facial expressions, and the involved warping method is not content-aware.

### 3D Face Reconstruction

Lots of facial reconstruction methods based on a single image exist. In the exceptional work of Blanz and Vetter,<sup>5</sup> a morphable face model was matched to a given 2D image by optimizing the parameters for the similarity between the 2D rendering of the morphable model and the original 2D image. Similar to the morphable head model, Chai and colleagues computed around 100 principal components for a collected head model database and fitted a 3D head model to the input image.<sup>7</sup> After that, a plausible high-resolution strand-based 3D hair model was developed for portrait manipulations, such as portrait pop-ups. Compared to previous 3D facial databases, FaceWarehouse by Cao and colleagues provided a much richer matching collection of expressions which can depict most human facial actions.<sup>8</sup> Different from these approaches, we are interested in facial reshaping based on a face image.

In 3D craniofacial reconstruction, Greef and colleagues conducted a large-scale study on how facial soft tissue thickness changes according to sex, age, and weight.<sup>4</sup> They studied 967 Caucasian subjects of both sexes, and varying ages and BMIs, and measured their facial soft tissue

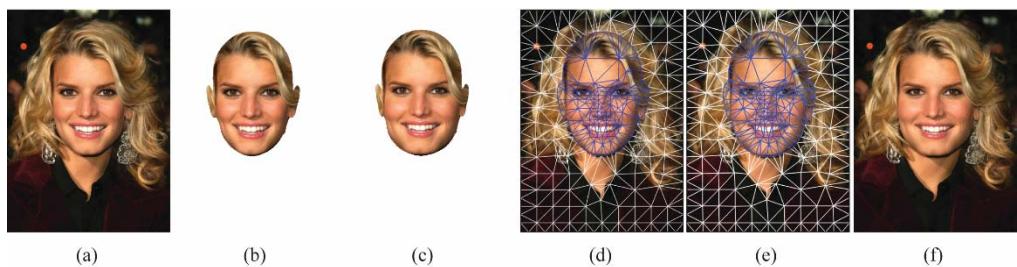
thickness on 52 facial feature points. For each factor and for both sexes separately, a multiple linear regression of thickness versus age and BMI was calculated. Our weight-change simulation is inspired by their regression equations.

## Image Resizing and Retargeting

Many content-aware image retargeting techniques have recently been proposed. Following the insightful survey conducted by Shamir and Sorkine,<sup>9</sup> the approaches fall into two categories: discrete and continuous. In discrete methods, seam carving and cropping are adopted to resize the input image. Continuous approaches optimize mapping using constraints, leading to content-aware resizing. Similar to body-aware image warping by Zhou and colleagues,<sup>3</sup> we embed the input image into a 2D triangular mesh, which is used to drive image warping to guarantee coherent resizing effects across the background. An approach to image retargeting employing mesh parametrization is proposed by Guo and colleagues,<sup>6</sup> which achieves the goals of emphasizing the important while retaining the surrounding context with minimal visual distortion. The preservation of salient object and image structure are maintained by optimizing constrained energy.

## ALGORITHM

We divide a portrait image into two regions: face region and the remaining region. For simplicity, we call the remaining region as the background region in our paper. A reshaping algorithm of a portrait image requires several steps. Figure 3 illustrates the outline of our algorithm. A 3D face model is first reconstructed using the method developed by Blanz and Vetter.<sup>5</sup> Based on forensic research results and the weight-change degree assigned by the user, deformed facial point positions are set (see 3D Face Reshaping section). Laplacian transformation is conducted afterwards (see 3D face Deformation section). Since only changing the face region is likely to introduce noticeable distortion to the background, a retargeting method is adopted (see Image Retargeting section).



**Figure 3. Algorithm overview.** Derived from input image (a), a 3D face model is calculated (b). Based on forensic research results and user-specified weight-change degree +3, Laplacian deformation is performed (c). Afterwards, a feature-relevant control mesh (d) is built on the original image. The target mesh (e) is produced by solving a mesh parametrization problem which preserves deformed face features with minimal visual distortion to the background. Standard texture mapping is finally used to render the target image, as shown in (f).

## 3D Face Reshaping

Our face reshaping algorithm is inspired by forensic research results by Greef and colleagues.<sup>4</sup> This study was focused on how sex, BMI and age influence the depths of facial soft tissue. The population in their research consisted of 457 males and 510 females of varying ages and BMIs. They selected 52 feature points where 10 points located on the midline and 21 points located bilaterally. The selection of these feature points was based on the ability to reliably locate them on the face. A multiple linear regression of soft tissue thicknesses versus BMI and age was calculated for male and female separately, as tabulated in Table 1.

**Table 1.** Linear regression equation: regression coefficients, the root mean square (RMS) errors and the significance levels. \* $p < 0.05$ ; \*\* $p < 0.01$ .

| Point numbers and descriptions | Males                |       |       |       |       |      | Females |       |     |       |     |      |     |
|--------------------------------|----------------------|-------|-------|-------|-------|------|---------|-------|-----|-------|-----|------|-----|
|                                | $b_0$                | $b_1$ | $p$   | $b_2$ | $p$   | RMSE | $b_0$   | $b_1$ | $p$ | $b_2$ | $p$ | RMSE |     |
| 1                              | Supraglabella        | 1.7   | 5.0   | *     | 104.5 | **   | 0.6     | 2.7   | 2   | 62    | **  | 0.6  |     |
| 2                              | Glabella             | 2.5   | 3.2   |       | 103.1 | **   | 0.7     | 3.4   | -2  | 77    | **  | 0.8  |     |
| 3                              | Nasion               | 3.6   | 11.9  | **    | 87.9  | **   | 1.2     | 4.8   | 15  | **    | 42  | **   | 1.3 |
| 4                              | End of nasal         | 1.8   | 2.4   |       | 37.6  | **   | 0.6     | 1.7   | -2  | *     | 36  | **   | 0.5 |
| 5                              | Mid-philtrum         | 11.4  | -37.2 | **    | 22.8  |      | 1.7     | 9.7   | -39 | **    | 39  | *    | 1.6 |
| 6                              | Upper lip margin     | 11.3  | -36.5 | **    | 17.7  |      | 2.0     | 10.6  | -18 | **    | -21 |      | 1.7 |
| 7                              | Lower lip margin     | 11.0  | -30.5 | **    | 92.0  | **   | 2.1     | 10.1  | -9  |       | 37  |      | 2.0 |
| 8                              | Chin-lip fold        | 7.4   | 11.7  | *     | 107.0 | **   | 1.3     | 7.9   | 23  | **    | 54  | **   | 1.2 |
| 9                              | Mental eminence      | 3.4   | 29.7  | **    | 238.3 | **   | 1.7     | 5.5   | 12  | **    | 174 | **   | 1.7 |
| 10                             | Beneath chin         | 1.7   | 2.4   |       | 190.8 | **   | 1.3     | 2.9   | 9   | *     | 118 | **   | 1.5 |
| 32/11                          | Frontal eminence     | 1.7   | 1.0   |       | 108.4 | **   | 0.7     | 2.3   | 1   |       | 76  | **   | 0.6 |
| 33/12                          | Supraorbital         | 1.8   | 5.1   |       | 148.8 | **   | 0.9     | 3.3   | 0   |       | 95  | **   | 0.8 |
| 34/13                          | Lateral glabella     | 5.0   | -15.0 | **    | 55.6  | **   | 1.3     | 4.9   | -15 | **    | 48  | **   | 1.1 |
| 35/14                          | Lateral nasal        | 3.4   | -9.1  | *     | 30.6  | *    | 0.6     | 3.9   | -14 | **    | 10  |      | 0.6 |
| 36/15                          | Suborbital           | 4.0   | 6.9   |       | 199.1 | **   | 2.3     | 7.7   | -29 | **    | 130 | **   | 2.2 |
| 37/16                          | Inferior malar       | 5.2   | 36.9  | **    | 452.6 | **   | 3.3     | 12.3  | 6   |       | 249 | **   | 2.8 |
| 38/17                          | Lateral nostril      | 10.2  | -29.5 | **    | 25.7  |      | 1.5     | 9.9   | -32 | **    | 14  |      | 1.3 |
| 39/18                          | Naso-labial ridge    | 12.0  | -32.3 | **    | -4.2  |      | 2.0     | 9.6   | -55 | **    | 70  | **   | 1.6 |
| 40/19                          | Supra canina         | 10.5  | -21.3 | *     | 25.8  |      | 2.0     | 10.6  | -57 | **    | 20  |      | 1.7 |
| 41/20                          | Sub canina           | 7.2   | -13.6 |       | 149.9 | **   | 1.7     | 9.2   | -31 | **    | 82  | **   | 1.5 |
| 42/21                          | Mental tubercle ant. | 4.2   | 23.4  | **    | 208.9 | **   | 1.4     | 6.6   | 8   |       | 129 | **   | 1.5 |
| 43/22                          | Mid lateral orbit    | 2.8   | -3.5  |       | 83.9  | **   | 0.7     | 4.1   | -1  |       | 42  | **   | 0.9 |
| 44/23                          | Supraglenoid         | 8.3   | -34.2 | **    | 109.7 | *    | 2.8     | 8.2   | -34 | **    | 104 | **   | 1.9 |
| 45/24                          | Zygomatic arch       | -1.2  | -5.1  |       | 315.4 | **   | 1.2     | 3.0   | -15 | *     | 194 | **   | 1.4 |
| 46/25                          | Lateral orbit        | -0.3  | -13.4 | *     | 364.9 | **   | 1.4     | 5.2   | -44 | **    | 266 | **   | 1.7 |
| 47/26                          | Supra-M2             | 12.4  | 9.6   |       | 565.5 | **   | 3.4     | 22.5  | -56 | **    | 275 | **   | 2.9 |
| 48/27                          | Mid-masseter muscle  | 6.7   | -9.1  |       | 447.0 | **   | 4.5     | 13.4  | -47 | **    | 194 | **   | 3.3 |
| 49/28                          | Occlusal line        | 8.8   | -36.0 | **    | 503.4 | **   | 2.4     | 13.1  | -58 | **    | 340 | **   | 2.0 |
| 50/29                          | Sub-M2               | 5.4   | 1.8   |       | 516.5 | **   | 3.2     | 14.2  | -27 |       | 250 | **   | 3.2 |
| 51/30                          | Gonion               | 2.0   | -2.8  |       | 547.0 | **   | 3.0     | 7.5   | -30 | **    | 340 | **   | 2.4 |
| 52/31                          | Mid mandibular angle | -4.1  | 45.9  | **    | 562.0 | **   | 2.5     | 3.8   | 12  |       | 329 | **   | 2.3 |
| 54/53                          | Pterion              | 1.7   | 1.0   |       | 108.4 | **   | 0.7     | 2.3   | 1   |       | 76  | **   | 0.6 |

Our work differs from that of Greef and colleagues<sup>4</sup> in that we add two extra control points (see Figure 4 and points 53 and 54 in Table 1) and set the limit for weight-change degrees to make them more suitable for our framework. Without feature points 53 and 54, the deformed 3D faces are likely to introduce artifacts around the pterion. The linear regression equation can be expressed as follows:

$$\mathbf{Y} = \mathbf{b}_0 + \mathbf{b}_1 \times \text{age} + \mathbf{b}_2 \times \text{BMI},$$

$$\mathbf{b}_0 = (b_0^1, b_0^2, \dots, b_0^{54}),$$

$$\mathbf{b}_1 = (b_1^1, b_1^2, \dots, b_1^{54}),$$

$$\mathbf{b}_2 = (b_2^1, b_2^2, \dots, b_2^{54}),$$
(1)

where vector  $\mathbf{Y}$  represents the tissue depths of the 54 feature points,  $\text{BMI}$  represents the body mass index, and  $\mathbf{b}_0$ ,  $\mathbf{b}_1$ ,  $\mathbf{b}_2$  represent regression coefficients, respectively.

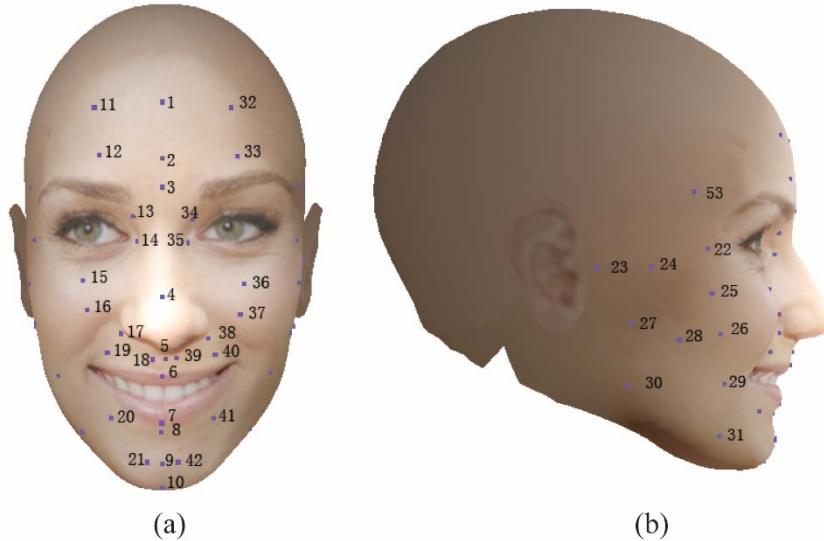
Moreover, there should be a limit for losing weight. Even someone is emaciated, the depths of soft tissue are still above 0. Therefore, we define the limit of the weight-change degree for the  $i$ th point  $T_i$  as:

$$T_i = -\frac{1}{b_2^i} (b_0^i + b_1^i \times \text{age} + b_2^i \times \text{BMI}), i = 1, 2, \dots, 54. \quad (2)$$

For a particular input image, the age of a person remains unchanged. Therefore, with varying BMIs, the updated feature point positions are only influenced by  $\mathbf{b}_2$ . We assume that the variation of facial tissue depth is along the feature point normal direction:

$$\mathbf{S}'_i = \mathbf{S}_i + \frac{db_2^i \mathbf{N}_i}{100}, i = 1, 2, \dots, 54, \quad (3)$$

where  $\mathbf{S}_i$  is the  $i$ th feature point position before deformation,  $\mathbf{S}'_i$  is the deformed  $i$ th feature point position,  $d$  is the weight-change degree, and  $\mathbf{N}_i$  is the corresponding normal of the  $i$ th point.



**Figure 4.**  
Illustration of  
feature  
points. The  
total number  
of feature  
points is 54,  
with 10  
located on  
the midline  
and 22  
located  
bilaterally.

### 3D Face Deformation

The 3D face model for the input portrait image is reconstructed using the method proposed by Blanz and Vetter.<sup>5</sup> They collected 200 head structure data using laser scans and exploited the statistics of the dataset to derive a morphable model and a parametric description of faces. Then, a fitting algorithm is developed to match the morphable model to the input 2D face image under shape and texture constraints. After that, a 3D face model conforms to the 2D face image is reconstructed. As a preparation to our algorithm, one of the generated models needs to be labeled with feature points manually. Since the topology of the morphable model mesh remains the same, we can use the pointwise correspondence to locate the feature points on other face models automatically.

After obtaining the deformed feature point positions in the 3D Face Reshaping section, various methods are capable of calculating the displacements of the non-feature points. Noh and colleagues proposed to use Radial Basis Functions to solve this problem.<sup>10</sup> A human face is full of abundant geometric details, and human perception is extremely sensitive to facial distortion. Therefore, we employ a Laplacian deformation method similar to that employed by Liao and colleagues,<sup>1</sup> which is based on the differential surface representation proposed by Sorkin and colleagues.<sup>11</sup> By utilizing Laplacian deformation, geometric details are preserved as optimally as possible.

The 54 feature points are assigned as handles, which are moved to new positions  $\mathbf{S}'_i$ . The updated positions of feature points are calculated in 3D Face Reshaping section. A better result is obtained if the handle constraints are satisfied in a least square sense. With  $c_i^x$ ,  $c_i^y$ , and  $c_i^z$  representing the  $x, y, z$  coordinates of the new position of the  $i$ th feature point respectively, the 54 handle constraints for  $x$ -coordinates are:

$$x_i = c_i^x, i \in 1, 2, \dots, 54 \quad (4)$$

Thus, all deformed face point positions  $\tilde{\mathbf{x}}$  are obtained by solving the following quadratic minimization problem:

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} (\|L\mathbf{x} - \delta_x\|^2 + \sum_{i=1}^{54} |x_i - c_i^x|^2), \quad (5)$$

where matrix  $L$  is the topological Laplacian of the face mesh,  $\mathbf{x}$  is the vector of the x-coordinate of all vertices, and  $\delta$  is the Laplacian coordinate matrix. The  $y$  and  $z$  coordinates are calculated in the same way.

The 3D face deformation results are shown in Figure 5. The negative weight-change degree indicates the decrease of BMI, which means losing some weight. On the contrary, positive weight-change degree represents the increase of BMI.

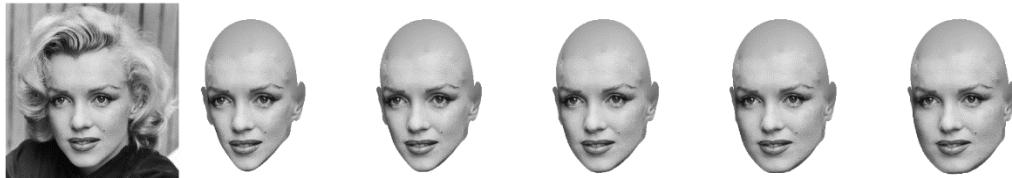


Figure 5. 3D face deformation results. The image on the left is the original image. The following images are the 3D face deformation results of weight-change degree  $-4$ ,  $-2$ ,  $0$ ,  $+2$ , and  $+4$ , respectively.

## Image Retargeting

Directly projecting a reshaped 3D face model into a 2D image will introduce visual artifacts. A content-aware image warping method is employed. Our method is based on the work of Guo and colleagues,<sup>6</sup> which avoids the distorting the salient object and retains the surrounding background with slight distortion. In their approach, a feature consistent mesh is generated using a constrained Delaunay triangulation algorithm according to the feature points extracted from the 2D input image. Several constraints, including boundary, saliency, and structure, are defined to avoid distorting salient objects in the optimization process for retargeting. After a stretch-based mesh parametrization process, the homomorphous target mesh is calculated, and the resulting image is rendered using texture mapping.

## Background Region

The control mesh should be consistent with image structure and retain uniformity of point density. The boundary of the input image is discretized first, and all of the points are set as control points. For the background part, the Canny operator is employed, and other control points are detected. Some additional points are added to keep the points well-distributed. As shown in Figure 6 (a), the blue points represent the control points in the background.

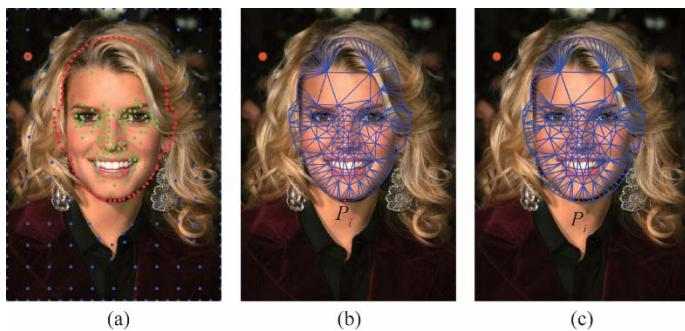


Figure 6. Control mesh comparison before and after 3D face deformation. The red dots in (a) are the control points on the contour profile as hard constraints, the green dots on the face are regarded as hard constraints, and the blue points on the background are set as soft constraints. The four points on the corners of the picture are set as fixed points. (b) is the hard constrained mesh superimposed on the source image. (c) is the comparison of deformed mesh superimposed on the source image. Since the weight-change degree is  $+3$ , the face contour expands, and the background needs to be compressed.

## Face Region

The control points on face regions are selected based on the result of the Canny operator, as well. Once we obtain the 3D face model and the deformed model in 3D Face Deformation section, a pointwise correspondence is set. Consequently, the deformed face region can be achieved easily. The  $i$ th point of the original morphable model is projected into the image space and marked as  $P_i$ .  $P'_i$  stands for the projected position of the  $i$ th point on the deformed model. The control mesh on the face region expands or shrinks with varying weight-change degrees. In Figure 6(c), the deformed constraint mesh is drawn on the source image. With +3 weight-change degree, the constrained mesh over the face expands.

## Face Contour

After 3D face deformation, the locations of vertices of 3D faces will change, and also their 2D projections in 2D image. As a result, the control points originally located on the contour profile are likely to shift away from the deformed contour profile after 3D face deformation, which will lead to noticeable artifacts after the retargeting process. Therefore, the control points along the contour profile of the face must be carefully selected. Let  $Mc$  be the set of the contour points along the source image, and  $P_i^c$  be the  $i$ th point in  $Mc$ , and  $P_0^c, P_1^c, \dots, P_n^c$  are in clockwise order along the contour. With a predefined threshold  $l$ , the control points are selected by minimizing the following energy function:

$$\min Et + \lambda Ed, \quad (6)$$

where  $Et$  is employed to distribute the control points uniformly along the contour of the face region, and  $Ed$  is employed to constrain the shifting of control points from the deformed contour. They are defined as follows:

$$E_t = \begin{cases} \infty & n = 0, \\ (d_A)^2 & n = 1, \\ \sum_{P_i^c \in Mc} ((d_{arc}(P_i^c, P_{(i+1)\%n}^c) - l)^2) & n > 1. \end{cases} \quad (7)$$

$$E_d = \sum_{P_i^c \in Mc} (d(P_i^c; B)^2 + d(P_i^{c'}; B')^2), \quad (8)$$

where  $B$  stands for the background of the source image,  $B^0$  is the deformed background,  $n$  is the number of points in set  $Mc$ ,  $d_A$  represents the length of the contour along the face region in the source image,  $\lambda$  is the weight factor balancing the influence of distance threshold constraints and location energy, which is set to 10 for our results.  $d(x; S)$  is the least distance of point  $x$  from set  $S$ . That is,

$$d(x; S) = \inf \{d(x, s) \mid s \in S\}. \quad (9)$$

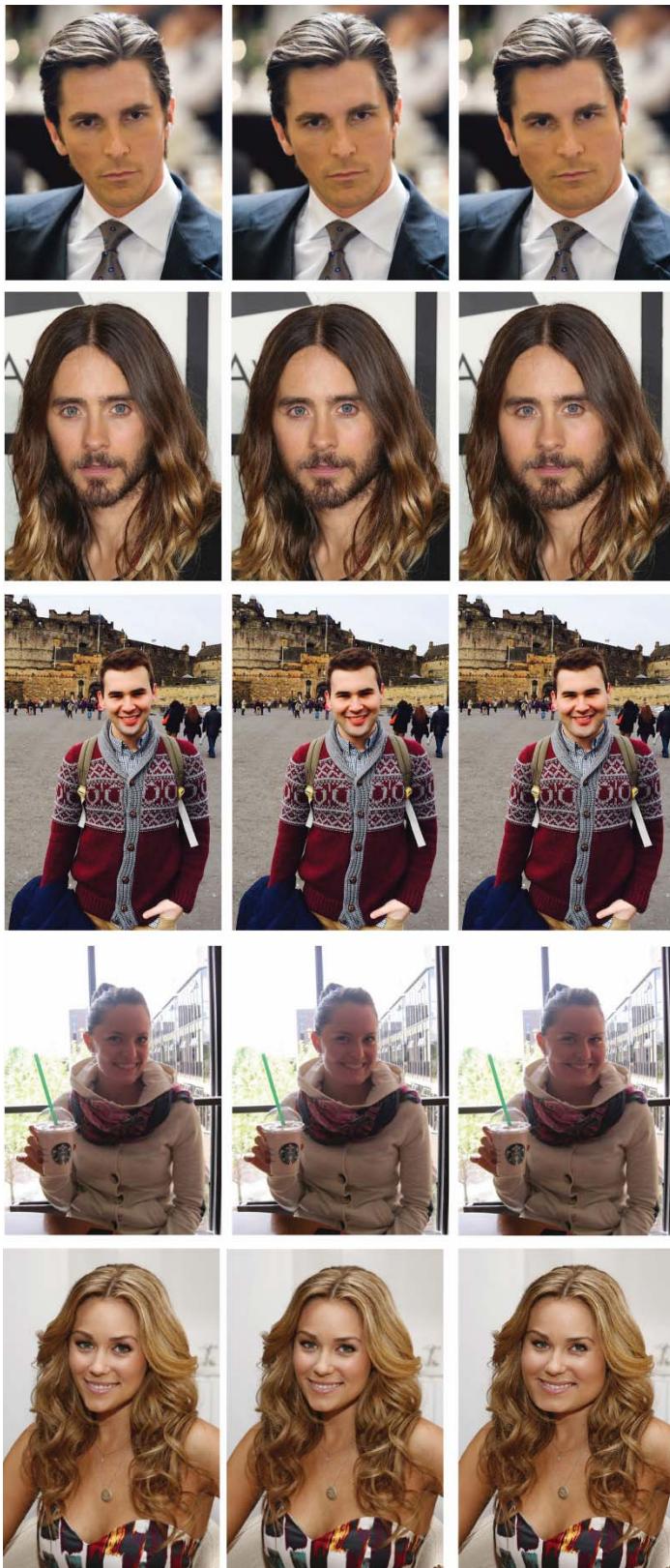
$d_{arc}(P_i^c, P_{i+1}^c)$  stands for the length of the face contour from  $P_i^c$  to  $P_{i+1}^c$  in clockwise order.

Equation 6 is minimized by adding one selected control point in  $Mc$  each time, which results in the largest decrease of energy. The selected point in each iteration is located between the adjacent dots with the longest distance along the contour. Thus, the above process in general can be efficiently implemented. The final solution is reached if adding a point does not reduce energy.

## Constrained Mesh Parametrization

Based on the control points selected from the background, face region and face contour, the constrained Delaunay triangulation algorithm is utilized to generate a feature-consistent mesh, as shown in Figure 3(d). Using the method proposed by Guo and colleageus,<sup>6</sup> the homomorphous target mesh is achieved, as shown in Figure 3(e). The background part is rendered using texture mapping, while the face region part is rendered based on the 3D deformed model. Finally, the reshaped image is obtained, as shown in Figure 3(f).

Figure 6 shows the comparison of control mesh before and after 3D face deformation. The control mesh of the face region (see Figure 6(b)) is flattened after 3D face deformation (see Figure 6(c)). In this example, the weight-change degree (Figure 6(a)) is +3 degrees.



**Figure 7. Reshaping results.**  
The left column is the weight-change degree of -2, the middle column is the original input image, and the right column is the weight-change degree of +2.

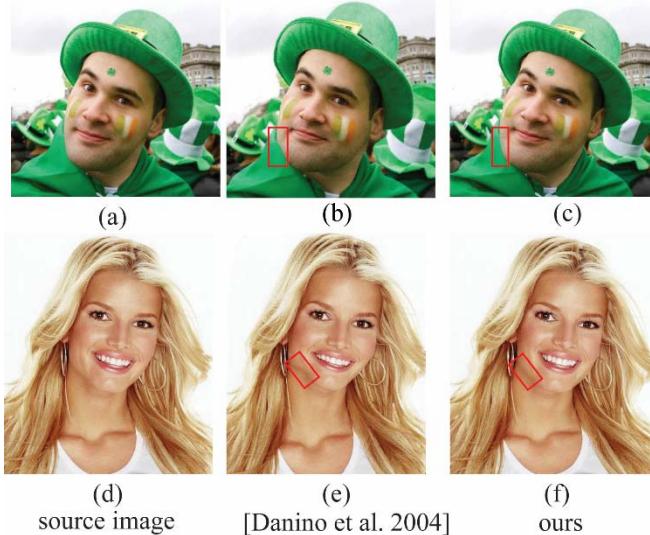
## RESULTS AND DISCUSSION

We have implemented our algorithm on a desktop PC with Intel I7 4.0 GHz CPU and 32 G memory. The average computation time is about 0.6 seconds for images with dimensions  $640 \times 480$ ; 1.2 seconds for dimensions  $800 \times 600$ ; and 1.6 seconds for dimensions  $1024 \times 768$ . We tested our method on a variety of facial images with various backgrounds and poses. Figure 7 shows some examples. For each example, the image in the middle is the input portrait, the left and right images are the reshaping results of -2 degrees and +2 degrees, respectively.

### Comparisons

One available facial reshaping method is the facial weight-change simulator proposed by Danino and colleagues.<sup>2</sup> This approach consists of the following steps. First, user marks thirteen landmarks on the portrait along the cheek and two landmarks around the neck. With user specified weight-change degree, the new locations of the landmarks are calculated based on empirically determined coefficients. After that, a thin-plate spline warping is employed to obtain the deformed facial image. To remedy the artifacts in the deformed background, a synthetic background with similar color is used to replace the actual background.

For frontal-view face images with neutral facial expressions and simple backgrounds, this method can produce realistic results. However, it may produce artifacts for non-frontal-view face images because some landmarks are hidden. Moreover, as the landmarks are mostly along the cheek regions, the nonlinear thin-plate spline warping will cause obvious distortions in other face regions (such as cheek regions as shown in Figure 8 (b) and (e)). When the weight-change degrees are large, Danino and colleagues' method<sup>2</sup> will produce obvious artifacts (see the distortions in Figure 2 (b) and (d)). For facial images with complex backgrounds, this method will generate unnatural distortions because their image warping is not content-aware. Since our approach recovers the 3D face model to simulate the weight-change of face and employs the content-aware image retargeting method, we can produce natural results with various expressions and poses.



**Figure 8. Comparison results.** (a) and (d) are input images. (b) and (e) are results of Danino and colleagues' method,<sup>2</sup> while (c) and (f) are our results. Please note the differences in the regions marked by red boxes.

We also compare our reshaping results with unprocessed camera images. We collected the pictures of some celebrities who have experienced weight change from being underweight to overweight or vice versa. Figures 10 (b-d), (g-i) are camera images and (a), (e), (f), (j) are our reshaping results. These reshaping results share a close similarity with the camera images.

### User Study

We have devised a user study to objectively verify the effectiveness of our facial reshaping method by measuring if the subjects can differentiate between our reshaping images and unprocessed camera images among various individuals of both sexes and varying BMIs.

**Examples.** We generate several reshaping images using our method described in the Algorithm section, which is called *ours*. We also collect various unprocessed images which contain human faces via the Internet, which is called *real*. The individuals shown in *real* have experienced significant changes of weight.

**Study details.** We recruit 25 subjects for this task. Each subject views 16 pairs of images of the same individual. Subjects are told to choose the most realistic image in the image pair. Two reference unprocessed images are provided in order to give users a more comprehensive impression of the person shown in the image pair. The first part of the user study is called *RT*. Ten of these pairs contain one real image and one reshaping image of the same person taken from different places. One example is shown in Figure 9 (a-d). (a) is our reshaping result and (b) is an unprocessed image. (c) and (d) are both unprocessed images, which are provided as references. The second part of this user study is called *ST*. The remaining six pairs contain one real image and the reshaping image on which it is based. One example of this is shown in Figure 9 (e-h). (f) is the original image, and (e) is the reshaping result generated from (f). (g) and (h) are provided as reference images for this pair.

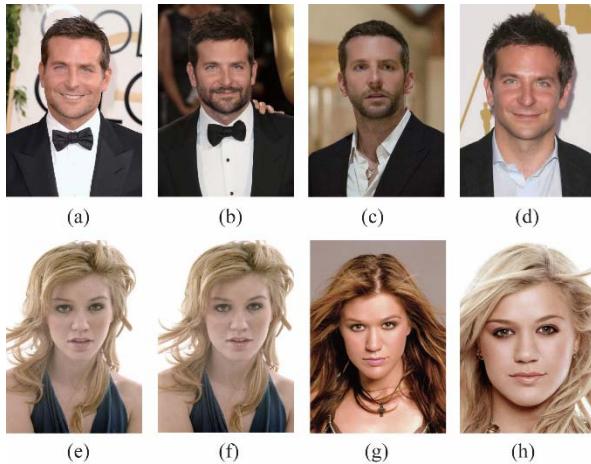


Figure 9: User study examples. (a-d) are the images used in user study *RT*, while (e-h) are used in user study *ST*. (a, e) are our reshaping results, and (b, f) are unprocessed images. (c, d, g, h) are unprocessed camera images as well, which are provided as references. In *RT*, ours (a) are compared with other camera images taken under different circumstance (b). In *ST*, ours (e) are compared with the source images (f).

Table 2. User study results. The statistics results of one-sample, two-tailed t-test for *RT* and *ST*. Test value is 0.5 (50%). CI stands for Confidence Interval of the difference.

|           | <b>Mean</b> | <b>Test Value = 0.50</b>  |              |              |
|-----------|-------------|---------------------------|--------------|--------------|
|           |             | <b>P-value (2-tailed)</b> | <b>95%CI</b> |              |
|           |             |                           | <b>Lower</b> | <b>Upper</b> |
| <b>RT</b> | 0.4960      | 0.956                     | 0.3369       | 0.6551       |
| <b>ST</b> | 0.4533      | 0.640                     | 0.2119       | 0.6947       |



Figure 10: Comparison results. (b-d) and (g-i) are camera images. (a), (e), (f), (j) are our reshaping results.

These image pairs are presented in a randomly permuted order, and the placement (left or right side) of *real* and *ours* is randomized, as well.

**Results.** We analyze the user study data of the two cases (*RT*,*ST*) separately. When the objects are asked to pick which image appeared more realistic in the *RT* test, 49.6% of the subjects choose *ours*. By performing a one-sample, two-tailed *t*-test for these 10 examples, we find out that subjects cannot find significant differences between our results and the real images (*p*-value  $\gg 0.05$ ). Therefore, the results of ours are as realistic as real to some extent. Regarding the *ST* part, fewer subjects chose ours (45.33%). Compared with the source image, subjects are able to distinguish the source image better. However, the *t*-test result of *ST* demonstrates that the difference is also not substantially obvious. Through this user study, we can conclude that our method is able to create natural reshaping results.

## Limitations

For very large weight-change degrees, our approach may generate artifacts around the cheek region and introduce noticeable distortion to the background. In our current implementation, the neck region of the input image is considered as background. As a result, the artifacts near neck regions may become obvious when the weight-change degrees are large (as shown in Figure 2).

Gaining or losing some weight will influence the appearance of the face. When gaining weight, a person's facial contours tend to expand, wrinkles seem reduced and, to some extent, a double chin emerges. When losing weight, a person's facial contours tend to shrink, wrinkles seem increased and, to some extent, a double chin disappears. Our current approach cannot simulate such wrinkle changes and "double chin" changes. As shown in the bottom-right image of Figure 7, wrinkle artifacts around the eye regions may arise when the weight-change degree is large.

## CONCLUSIONS AND FUTURE WORK

We have proposed an effective image reshaping system to thin or fatten a face based on user input weight-change degree. After we gain a 3D morphable face model, forensic data are used to parameterize the reshaping process of the 3D model. We rely on the deformed 3D model to reshape the source image. We introduce a novel approach for choosing control points along the profile of the face. The effectiveness of our parametric weight-change reshaping method is proved by examples and user study.

Our system provides a real-time solution to reshaping a camera image by simply setting weight-change degree.

We are currently working on several enhancements to our reshaping system. Although the current system allows reshaping face regions, the neck region should be added to generate more visually pleasing results. In addition, there are more extensions to render the face region with the reconstructed face morphable model, such as relighting. We are also interested in extending our approach to the mobile phone platform.

## ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their valuable suggestions and comments, which help to improve the readability of this paper. Xiaogang Jin is supported by the National Key R&D Program of China (No. 2017YFB1002600), the NSF of China (Nos. 61732015, 61472351), and the Key Research and Development Program of Zhejiang Province (No. 2017C03SA160073). Kun Zhou is supported by the NSF of China (No. 61272305), the National Program for Special Support of Eminent Professionals of China, and Lenovo's Program for Young Scientists.

## REFERENCES

1. Q. Liao, X. Jin, and W. Zeng, “Enhancing the symmetry and proportion of 3d face geometry,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 10, 2012, pp. 1704–1716.
2. U. Danino, N. Kiryati, and M. Furst, “Algorithm for facial weight-change,” *Proceedings of the 2004 11th IEEE International Conference on Electronics, Circuits and Systems (ICECS 04)*, IEEE, 2004, pp. 318–321.
3. S. Zhou et al., “Parametric reshaping of human bodies in images,” *ACM Trans. Graph.*, vol. 29, no. 4, 2010, p. 126:1.
4. S. De Greef et al., “Large-scale in-vivo caucasian facial soft tissue thickness database for craniofacial reconstruction,” *Forensic Science International*, vol. 159, 2006, pp. S126–S146.
5. V. Blanz and T. Vetter, “A morphable model for the synthesis of 3d faces,” *Proceedings of the 26th annual conference on Computer graphics and interactive techniques (SIGGRAPH 99)*, 1999, pp. 187–194.
6. Y. Guo et al., “Image retargeting using mesh parametrization,” *IEEE Transactions on Multimedia*, vol. 11, no. 5, 2009, pp. 856–867.
7. M. Chai et al., “Single-view hair modeling for portrait manipulation,” *ACM Trans. Graph.*, vol. 31, no. 4, 2012, p. 116:1.
8. C. Cao et al., “Facewarehouse: A 3d facial expression database for visual computing,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, 2014, pp. 413–425.
9. A. Shamir and O. Sorkine, “Visual media retargeting,” *ACM SIGGRAPH ASIA 2009 Courses*, 2009, p. 11:1.
10. J.-Y. Noh and U. Neumann, “Expression cloning,” *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 01)*, 2001, pp. 277–288.
11. O. Sorkine et al., “Laplacian surface editing,” *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing Pages (SGP 04)*, 2004, pp. 175–184.

## ABOUT THE AUTHORS

**Haiming Zhao** is a Ph.D. candidate at State Key Lab of CAD & CG, Zhejiang University. Zhao received his B.Sc. degree in statistics from Zhejiang University in 2012. His research interests include mesh editing, image processing and 3D printing. Contact him at haiming2zhao@gmail.com.

**Xiaogang Jin** (corresponding author) is a professor at the State Key Lab of CAD & CG, Zhejiang University. His research interests include implicit surface computing, cloth animation, crowd and group animation, texture synthesis, and digital geometry processing. Jin received his Ph.D. in applied mathematics from Zhejiang University. He received an ACM Recognition of Service Award in 2015. Contact him at [jin@cad.zju.edu.cn](mailto:jin@cad.zju.edu.cn).

**Xiaojian Huang** is a Software Engineer in Hangzhou Tuhua Technology Co. Ltd. Huang received his B.Sc. degree in computer science from Zhejiang University in 2014. His research interests include facial animation and cloth simulation. Contact him at [xjhuang0401@foxmail.com](mailto:xjhuang0401@foxmail.com).

**Menglei Chai** is working toward the Ph.D. degree at the State Key Lab of CAD & CG, Zhejiang University. Chai received his B.Sc. in computer science from Zhejiang University in 2011. His research interests include image-based modeling and interactive image manipulation. Contact him at [cmlatsim@gmail.com](mailto:cmlatsim@gmail.com).

**Kun Zhou** is a Cheung Kong Professor in the Computer Science Department of Zhejiang University, and the Director of the State Key Lab of CAD & CG. Prior to joining Zhejiang University in 2008, Dr. Zhou was a Leader Researcher of the Internet Graphics Group at Microsoft Research Asia. He received his B.S. degree and Ph.D. degree in computer science from Zhejiang University in 1997 and 2002, respectively. His research interests are in visual computing, parallel computing, human computer interaction, and virtual reality. He currently serves on the editorial/advisory boards of ACM Transactions on Graphics and IEEE Spectrum. He is a Fellow of IEEE. Contact him at [kunzhou@acm.org](mailto:kunzhou@acm.org).

# ARIES: Enabling Visual Exploration and Organization of Art Image Collections

**Lhaylla Crissaff**  
Universidade Federal Fluminense

**Louisa Ruby and Samantha Deutch**  
The Frick Art Reference Library

**Luke DuBois**  
New York University

**Jean-Daniel Fekete**  
INRIA

**Juliana Freire and Cláudio T. Silva**  
New York University

Art historians have traditionally used physical light boxes to prepare exhibits or curate collections. On a light box, they can place slides or printed images, move the images around at will, group them as desired, and visually compare them. The transition to digital images has rendered this workflow obsolete. Now, art historians lack well-designed, unified interactive software tools that effectively support the operations they perform with physical light boxes. To address this problem, we designed ARIES (ARt Image Exploration Space), an interactive image manipulation system that enables the exploration and

organization of fine digital art. The system allows images to be compared in multiple ways, offering dynamic overlays analogous to a physical light box, and supporting advanced image comparisons and feature-matching functions, available through computational image processing. We demonstrate the effectiveness of our system to support art historians' tasks through real use cases.

The field of Art History expanded rapidly with the advent of photography in the 19th century. For the first time, art historians had permanent access to images of works of art that they could use as reference tools for their writing and thinking on art. With the release in the past fifteen years of millions of images online, the field is once again experiencing a major transformation. Although a large number of works of art or even existing photographs of works of art in the

world have yet to be digitized, having digital images has revolutionized the way art historians function. Rather than having to rely on analog photographs of varying quality, available only as part of either a personal slide or image library, art historians now have instant online access to a large number of high quality images.

Art historians study artistic artifacts—paintings, drawings, sculptures and architectures—for various purposes such as defining their meaning, style, value and historical context in which they were created. For an art historian, images are the foundation of all research; they are the examples, use cases, test cases, subjects, objects, and purpose behind all of their work. Before the digital age, art historians used reproducible physical media such as prints (etchings or lithographs), photographs, or slides as memory aids for works of art they had seen and wanted to discuss. These physical objects can be moved around on tables or light boxes (Fig. 1) and organized into different piles and groupings for further examination.



**Figure 1.** In the physical world, art historians work with photographs and slides that can be moved around on tables (left) or light boxes (right).

Today, art historians mostly work with digital images, but their needs have not been adequately met: there are no software tools that support the organization, editing, use, and flow of images needed to support their work methodology. As a result, working with digital images is time-consuming, tedious, and requires ad hoc configurations of software tools to perform the basic tasks required in their field.

To address this problem, we have designed ARIES (ARt Image Exploration Space), a new interactive system that better supports art historians in their daily workflow. ARIES was developed as part of a collaboration with professional art historians from the Frick Collection, one of the richest privately-held art collections in the United States and a prestigious art museum in New York City. The system simplifies the exploration, analysis, and organization of digital image collections by allowing experts to easily manipulate digital images like they used to manipulate printed images, and combining this functionality with advanced tools for comparing images and matching features.

Our main contributions can be summarized as follows:

- We present the tasks and requirements that influenced the design of ARIES and discuss how we addressed them.
- We describe the ARIES integrated environment and its components, including the virtual light box and a set of tools that support grouping, organization, annotation and comparison of images.
- We report on the use of ARIES by art historians and discuss a set of use cases that demonstrate the effectiveness of the system.

## BACKGROUND

Tools have been designed to support the creation, sharing and integration of archives of digital objects. Fuchs and colleagues<sup>1</sup> and Smith and colleagues<sup>2</sup> presented VLMA, a peer-to-peer framework that facilitates viewing, collecting, and reusing distributed visual archives and associated metadata. By configuring VLMA on a server, a user can both share her files and obtain access to files shared by other VLMA users. Within a server, once the files are imported from other collections, it is possible to perform simple tasks such as arranging, scaling and adding annotations. Google's Cultural Institute<sup>3</sup> is a web-based collection exploration environment that allows users to create their own collections by importing fragments from other collections stored in Google's database. These include art collections from museums, galleries, and works of individual artists (for instance, street graffiti artists). Collections of images can be browsed, individual items can be examined in high resolution, shared, and it is possible to view two images side by side. Saklofske<sup>4</sup> created an exploratory tool to visualize artwork collections. In this tool, each artwork is a page in the archive, and each page is represented as a node in the visualization. Nodes can be repositioned, grouped, connected (by drawing a line between two nodes), and annotated. Like these approaches, ARIES helps users organize image collections. However, ARIES also provides operations that support analysis, exploration, and comparison of images as well as the ability to arrange multiple images freely on a screen that can be shared between users.

Ciocca and colleagues<sup>5</sup> proposed a system for browsing museum collections that employs a multi-touch table as a user interface. Images from their database can be imported into the application through physical object interactions, and then be manipulated by applying rotation, scaling and translation using finger gestures on the touch interface. Multi-touch tables systems have also been designed by Hinrichs and colleagues,<sup>6</sup> Hornecker,<sup>7</sup> and Davis and colleagues<sup>8</sup> to engage museum visitors in the exploration of museum exhibitions. In all cases, tabletops display information about artworks and visitors can select what to see, such as textual information, pictures and videos. Geller's survey<sup>9</sup> presents many successful tabletop systems used by museums. While these systems were designed for museum visitors, ARIES and its features were designed to support research in art history.

An important task required by art historians and not supported by any of these tools is the ability to compare images. In the Computer Graphics field, automatic techniques have been proposed for image comparison.<sup>10,11</sup> In these works, keypoint detection and feature vector extraction enable the identification of matches between comparable images. Images that art historians compare usually show differences of degree, not kind, a nuance lost by simple automatic image comparison algorithms, which do not behave in way that would add to any of the tasks we were aiming to address. Instead of applying automated techniques that are not useful for art historians, we opted to approach image comparison in a user-driven fashion: we provide tools that enable users to focus on (important) details present in the artworks, and guide art historians as they classify the images.

In the domain of cartography, tools have been created to compare images, either to study evolution over time or to register multiple images. Lobo and colleagues<sup>12</sup> described and evaluated the effectiveness of multiple interactive techniques to compare two maps, which are images with specific semantics. Their main task consisted of spotting changes between two images. Elias and colleagues<sup>13</sup> also proposed an interface for browsing and correlating large numbers of maps. Art historians also need to correlate images, but just like comparing images, these are only two of the many tasks they routinely perform while working on their images.

Currently, there are no tools designed specifically for art historians, working on their own, offline, and with images from their own collections. As a result, art historians are forced to use multiple tools to accomplish what the (physical) light box offered. Tools such as Pinterest and applications such as Adobe Photoshop, Adobe Bridge, Adobe Lightroom and Microsoft PowerPoint are commonly used but support only a subset of the functionality they need. Having to use multiple tools that are not properly integrated greatly hampers the experts' ability to analyze their data. ARIES focuses on the specific tasks performed by art historians and provides an integrated environment designed to support art historical research.

## REQUIREMENTS

Art historians usually work with images of paintings, drawings and sculptures taken not only from their own databases, but from sanctioned databases (so-called “special collections”), as well as the web. These images are organized into groups using criteria established for a given project, such as subject matter, artist, style, date, or location. Metadata is used to describe the work of art, such as title, date, medium, dimensions and provenance (history of locations the artwork has been and events it was part of). Images are visually compared to check for similarities and differences, duplicates, degradation or restorations, forged artworks, to see closeups of details, and so forth. Their findings are shared with colleagues from museums, galleries, schools and research centers around the world.

When we started this project, our goal was to create an environment to help art historians explore and organize art images in a virtual space. We wanted to bring the flexibility and freedom they enjoyed while manipulating images in light boxes to a digital environment.

The initial needs assessment for ARIES was done in collaboration with three professional art historians from the Frick Collection. We carried out extensive interviews to understand their workflow, how they handled art images, and what questions they need to answer in their research. Two of the art historians, co-authors of this article, also tested the system, provided feedback during the development, and used it in real projects, which we describe later.

After identifying the main tasks, an initial list of desired functionality was made based on art historians’ experience in using other computational tools such as Adobe PowerPoint, Adobe Bridge, and Pinterest. Initial mock-ups of a possible digital light box interface were developed by graduate students studying user experience and presented to the staff from the museum. Although each mock-up had some of the established required functionality, we had to combine many versions of the mock-ups to create an interface that supported multiple image manipulation functions. This iterative design phase was crucial for developing an interface focused on art historians’ needs and still maintaining its simplicity.

The interviews phase allowed us to derive two sets of requirements. The basic requirements guided the design of the workspace to manipulate digital images. The workspace should allow the user to bring several images from multiple sources into a single environment and to work on them at will; work with the metadata associated with the images; add annotations to images; group and arrange images; save work in progress for later use; and share work in progress with collaborators. These features, despite being simple and already supported by other tools, enable ARIES to provide a unified framework that supports all the tasks in the art historians’ workflow.

The second set, the specific requirements, determined the functionalities needed to explore and compare images. The system should assist experts to detect, interpret and understand works of art and their relationships. In particular, it should support the ability to: superimpose images, allowing for a quick and easy determination of physical changes to works of art that may have taken place over time due to damage or conservation; the detection of forgeries and copies; and the instantaneous comparison of the differences between draft drawings and finished works of art. The system should also support complex visual comparisons of similar compositions by being able to quickly select analogous portions of multiple works of art and placing them side by side. This allows art historians to analyze an artist’s myriad portrayals of a specific scene, be it a painting, print or drawing, or to trace the development of artistic themes over a specific period of time either within one artist’s oeuvre or across the oeuvres of several artists. In addition, it should have the ability to easily adjust images to appear in the correct size relative to one another, a function necessary for planning the accurate hanging of works of art in a particular exhibition space, recreating historical displays of art, or even producing fantasy exhibitions with works of art that no longer exist or that cannot be moved from their current locations. Combining these advanced functions with basic image editorial capabilities, such as cropping and adjusting the color and light saturation, allows art historians more easily to make use of lesser quality photographs taken in the field.

In this article, we focus on the novel tools and techniques we designed to address the specific requirements.

## SYSTEM OVERVIEW

We built ARIES to meet the requirements gathered during our needs assessment. The ARIES system consists of the following components: a user-friendly interface; the workspace management module that supports uploading images, saving and opening image groups, and managing and sharing work in progress; and tools for image exploration, interactive visual comparison, and metadata exploration.

### User Interface

Projects carried out by art historians share the common need to view works of art. When these works are unavailable for study, which is usually the case, art historians must substitute visual surrogates in the form of photographs or digital images. Projects can require from very few to a large number of images. The goals also vary. Sometimes, art historians have large collections of images that they need to browse, and by grouping them into observed categories and relationships, they can make new discoveries. In some cases, they just want to check whether two images represent the same work of art or, perhaps, similar works by the same artist or by two different artists. The ARIES interface gives art historians a place to gather these images and to explore them in many different ways, according to the needs of a given project.

The ARIES interface was inspired in the mock-ups developed by graduate students studying user experience. We created several prototypes, performed tests with actual users to evaluate the components of the interface, and iteratively refined the design. The final and functional design is shown in Fig. 2. This interface fulfills some of the basic requirements and it was the first step toward a computational tool for exploring images of works of art.

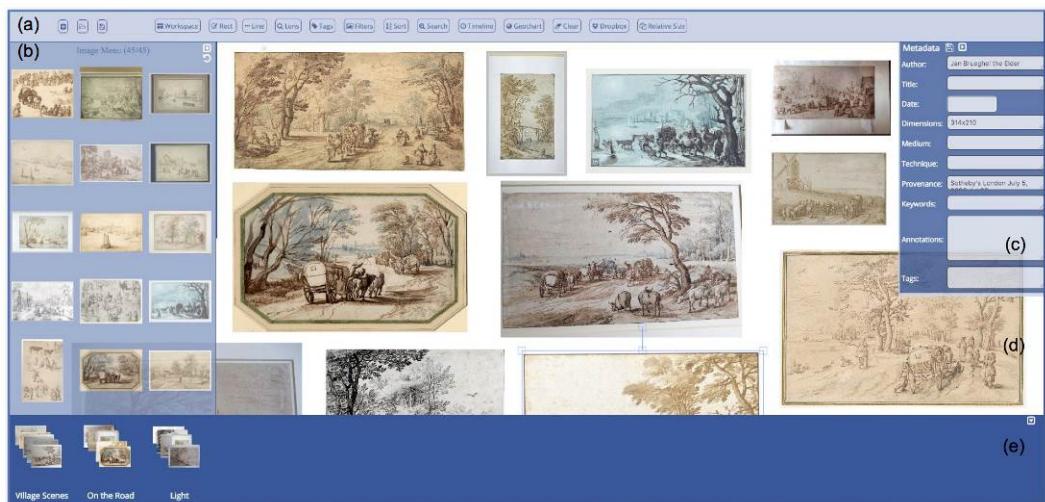


Figure 2. The ARIES interface includes (a) a toolbar and four views: (b) image menu, (c) metadata, (d) lightbox canvas and (e) group menu. Image menu, metadata and group menu are retractable, enlarging the lightbox canvas. Works of art on the lightbox canvas are displayed in relative size.

The interface (Fig. 2) consists of four permanent views that serve different purposes, a toolbar, and keyboard shortcuts that facilitate some operations. The image menu view (Fig. 2b) shows a thumbnail for every image that has been uploaded by the user. These are presented in a three-column grid and can be reorganized by dragging the images around within the view. The purpose of this menu is to make a large number of images easily accessible. When dragged, images can also be dropped on the lightbox canvas (Fig. 2d) or group menu views (Fig. 2e).

The lightbox canvas view is where image exploration and interactive visual comparison take place. Images within this view can be rotated, scaled and translated.

The group menu view organizes groups of images. In this context, a group is a set of images that users put together in order to systemize their findings as needed for a given project. Once a group is created and named in the top menu of the application, it is added to the group menu view. Images from the image menu view can be added to a group by drag and drop. A group is displayed as a stack of thumbnails showing the last five images added to it. The right mouse button, when hovering over a group, enables other operations, such as visualizing every image in the group, throwing all images within the group to the lightbox canvas, removing a group, and annotating a group with metadata.

The metadata view (Fig. 2c) displays and allows users to edit information regarding the works of art represented in the images. It includes a set of specific elements such as author, title, year of creation, provenance, and medium to identify and describe the work of art. Once an image is selected in the lightbox canvas, image menu or group menu, the elements within the metadata view are available to receive the appropriate data by the user on the top-right corner of the application. Batch annotation of multiple images with the same metadata is also supported.

Along with the different views, ARIES has a toolbar (Fig. 2a) for the user to select the tools they wish to use via mouse or keyboard shortcuts. The shortcuts enable a user to remove images from all views, enlarge images to fit the entire viewport occluding other images on the lightbox canvas view, and hide/show the three views (image menu, metadata, group menu) to allow for more space to work on the lightbox canvas view.

## Workspace Management

The different views in ARIES allow users to organize and compare images. Any image on the user's hard drive can be uploaded, making it available as a thumbnail in the image menu to be explored immediately.

The current state of the application can be saved as a project as soon as the user begins to interact with the system. Saved projects are kept in individual folders, and the user has the option of returning the system state to what it was at the moment the project was saved by using the Open Project option. A JSON file is used to store the state of the four views of a project as well as copies of all images uploaded to ARIES.

Once a project is saved, it is also ready to be shared with other users. To share, the system uploads a saved project to an account on the Dropbox file hosting service. After being prompted to log in with a Dropbox account, a directory is created in the user's Dropbox file system and all images present in their project are uploaded to this directory. Images are located at the root along with the JSON file defining the ARIES project. This folder can be shared with other Dropbox users, and the system can download a project folder from Dropbox and load the project back into ARIES.

## Image Exploration

An important goal for ARIES is to allow users to organize, compare and annotate art images. Unlike physical light boxes, whose main function is to bring together images for comparison by the human eye, ARIES can leverage digital image processing to manipulate images by size, color, and background. Image comparison consists of two key steps: detecting the similarity or difference among images and interpreting what was detected. In the first step, fine features are automatically or interactively detected. However, interpreting the complex visual details must be driven by a human taking the requirements of the task and context into account. Thus, in ARIES we integrate computing capabilities to detect important details so that the system can guide experts in their analyses.

Interpretation is central to art historians' core task of analyzing works of art images. An art historian analyzing an image is able to make subjective assumptions that would otherwise be difficult to make solely by purely computational techniques. For instance, a computational tool can detect that two paintings captured at different occasions by any device (camera or scanner) have differences, but it cannot identify if these differences are due to deterioration of materials used in the

original painting over the years or if images are from different paintings. Art experts can use not only their experience, but also their understanding of artistic practice to understand image relationships.

Typical comparative strategies include juxtaposition and superimposition.<sup>14</sup> ARIES implements these strategies to simplify the identification of similarities of different images and interpret complex visual details according a particular task. Images can then be annotated to record new discoveries.

### Dynamic overlays

With the dynamic overlay function supported by the lightbox canvas view, images can be placed on top of each other in a single environment. Image opacity is dynamically altered when layered, a function that is akin to working with physical image transparencies. This function allows the user to see through multiple layers of images, thus recreating an important feature of real light boxes (Fig. 3). Similar shapes that are shown in many images can be manually aligned on top of one another to allow for a better comparison.



Figure 3. This image exemplifies the use of the dynamic overlay tool. Two different images were added to the lightbox canvas, and one of them was moved on top of the other. The opacity of the top image was modified automatically, allowing the user to analyze both images simultaneously. In this example, the alignment was based on the soldier's head.

The dynamic overlay feature makes an art historian's job much easier when comparing two different visual surrogates for a work of art, that is, a photograph and a print after a painting, or two different photographs that appear to be of the same painting but may not be. When one is placed on top of the other, the similarities and differences of the print and photograph become immediately apparent. While the dynamic overlay of images is a simple feature, it is an indispensable tool for the art historian's work.

### Lens tool

The lens tool reduces the distance between images to be visually compared. By bringing selected areas from two separate images together side by side, this tool simplifies comparisons. Once portions of the images are placed side by side, art historians can address many research concerns including differences in the execution, style, and technique of the works of art; whether they are by the same artist; whether one is a copy of the other; if one or the other shows the work in a conserved state; or if they are two different works of art entirely.

To use this tool, the user selects two images on the lightbox canvas to be displayed in the largest possible sizes, occluding other images. With this new view, the user must manually select two

matching points, one in each image, to pinpoint a precise area. These points are the center of two rectangles that highlight the image details of interest to the user. Using the mouse in a swiping motion and keeping the cursor outside of the rectangles, the user can move both rectangles at same time while the system displays a view of the slices of the images contained in the rectangles side-by-side with the rectangle of the reverse image (Fig. 4). When the mouse cursor is inside the rectangles, the swiping motion alters the sizes of the rectangles to visualize bigger or smaller slices of the original images. The user can also scroll the mouse to zoom in or out on the slices. This tool is a special kind of Magic Lens<sup>15</sup> to assist in detecting similarities and differences between images as well as in improving the accuracy of the decision and the eye-saccade time. Fig. 4 shows how an art historian used the tool to compare a digital image of a painting with a digitized photographic reproduction of a similar painting. The tool allowed the art historian to determine that the images reproduced the same painting and, as a result, she was able to fill in gaps in that painting's provenance.



Figure 4. This figure shows an example of the lens tool applied to two images containing young girls. The matching points are represented by white crosses and centered within the rectangles that surround a precise area for further inspection and analysis. Once the points are marked, the user can move the mouse to update the slices within the rectangles.

### Rectangle tool

This feature-matching function provides a very simple way for art historians to explore portions of multiple images of the same work of art. By examining different types of images of a work of art, such as x-rays, infra-red reflectograms, prints, drawings and historical photographs, one can make judgments about changes in the artist's intent from drawing to finished work, changes to the work over time, and its representation in other media such as prints.

The user must first manually create a stack of overlapped images and then use the tool to draw one rectangle on the top image (Fig. 5a and 5b). After occluding every element on the viewport, the system displays a thumbnail view of the original stacked images and their mapped rectangles in the bottom part of the lightbox canvas view (Fig. 5c). On the top, slices of two of the stacked images contained in the rectangles and a heat map relating the slices are shown. The heat map, ranging from white to red, presents the pixel differences between the slices of the two images. The displayed slices can be modified by selecting different images among the thumbnails. With this structure, users can move the relative rectangles of all images in all directions by using the mouse in a swiping motion (Fig. 5d) and can zoom into rectangles simultaneously to examine specific details in the set of images. During the replacement, one of the slices displayed on the top of the lightbox canvas stays static while another one is updated with the corresponding portion of the image inside the rectangle, and the heat map is automatically updated. The overview+detail interface<sup>16</sup> allows users to go through all the images according to their previously established mapping, providing a better understanding of the behavior of the image feature detection while retaining the context of the task at hand.

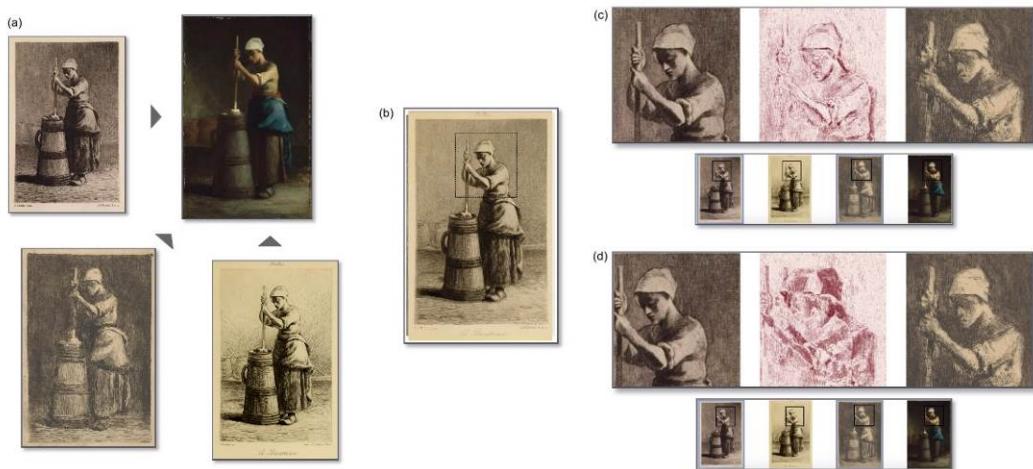


Figure 5. This figure shows (a) the application of the rectangle tool to the four images. (b) The images are stacked based on the woman's head alignment, and one rectangle was drawn around it. (c) One of the possible results provided by the tool: on the bottom, the system displays thumbnails of all stacked images and their rectangles, and on the top, slices of original images contained within rectangles are displayed. To help the user identify differences and similarities, a heat map is displayed that represents the pixel differences between the images. (d) A different result after moving the rectangles according to the relative position previously established. In this case, only the left image was moved while the right one keeps the same position.

## Tags

In order to store discoveries regarding the images, ARIES provides a tag tool. This feature allows art historians to share their observations with other members of a given project or to store them for their own future use. In the lightbox canvas, the user can create a rectangle surrounding a feature in an image and add an annotation related to the feature. A tag, in this context, is the composition of the rectangle and the annotation. The user can create as many tags as desired to record important findings. Once a tag is created, its annotation appears on one of the fields of the metadata view where it can be updated.

## Filters

ARIES provides several filters: grayscale, brightness, contrast, color overlay, and edge detection.<sup>17</sup> These filters can be used to emphasize or remove features from images, providing art historians with a better basis for visual comparison.

Note that filters, dynamic overlays, lens and the rectangle tool were designed to meet the specific requirements discussed in the Requirements section.

## Metadata Exploration

Besides exploring the contents of images visually, users can also add to and explore the metadata associated with the images. Below, we describe the metadata exploration functionality supported by ARIES.

### Relative Size

This tool allows art historians to see the relative size of one image in comparison with all others on the lightbox canvas. This enables them to understand the relationships between different artworks by answering questions like: Is a given work smaller or larger than another? By how much? This tool is useful for planning the hanging of exhibitions as well as to understand how or

where works of art may have been displayed in the past. Displaying works on the screen as being of the same size can lead to misconceptions. Using the dimensions inserted in the metadata view, the images are rescaled on the lightbox canvas, allowing for the real-life proportion of artworks to be evaluated within the application.

## Search

This function allows images to be filtered by the different facets present in metadata, including tags. Once the user clicks on search button in the toolbar, the system displays the metadata elements in a dropdown menu and a box where the user must enter the word to search for. The system finds all images stored in the current project that match the search and displays those images in the search menu view that temporarily substitutes the image menu view. The drag-and-drop functionality is still available for use with images in search menu view as well as an option to return to previous view.

## Sort

The user can sort images from the image menu view or a selected group in the group menu by all categories presented in the metadata view. Once the sort button in the toolbar is clicked, the system displays a dropdown menu with the metadata categories and radio buttons, allowing the user to select between the image menu and a selected group. Images are reorganized in their respective view in ascending order of date or alphanumerical order depending on the chosen metadata element.

## Timeline

For art historians, it is important to study the development of an artist's work over time, categorize changes that have occurred to a single work of art over time, and analyze general trends of artistic techniques and styles over time. Understanding the chronological progression of art and artists throughout the ages helps to determine the "history" of art which can, in turn, shed light on the societies where the art was created.

The timeline tool supports the exploration of artworks over time. ARIES displays a timeline of images from a selected group once the creation dates of the works of art have been entered in the metadata view. By quickly and easily displaying the creation chronology of the given group, stylistic trends or deviations are illuminated. This chronological visualization is designed to display images and their metadata in an interactive way. After selecting a group and clicking on the timeline button in the toolbar, the timeline fits the entire viewport, occluding all other views. On the bottom, a chronological line is shown with markers indicating the images on that date. Once the user clicks on the rectangle representing an image (in the timeline), the image and its metadata appear on the top. By clicking on the arrows placed to the left or right of the screen, the user can browse all images of the group. Fig. 6 illustrates a timeline created in ARIES by an art historian that shows Jean-Francois Millet's "Woman Churning Butter" and several copies and versions of that work by other artists displayed in chronological order. The timeline quickly enabled new insights concerning the differing thematic interpretations and ways of representing the work of art over time.

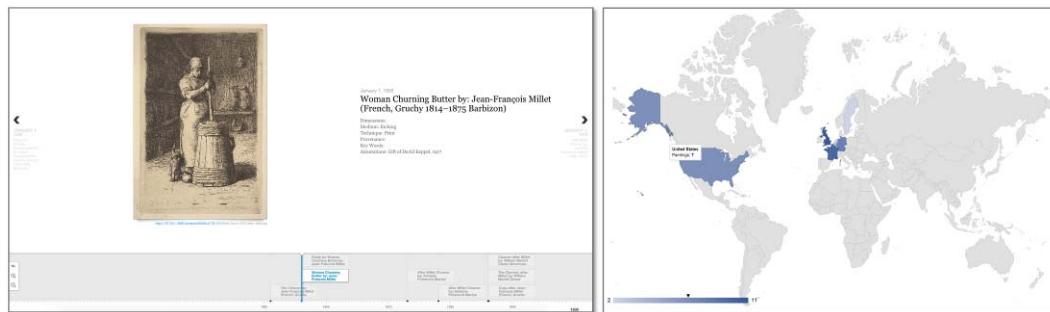


Figure 6 (left). ARIES's timeline. Figure 7 (right). Geochart of a group created in ARIES.

## Geochart

This function allows for a visualization of the geographical location of works of art over time. It facilitates investigations into the history of artistic styles and the popularity of various artists over time in different countries, as well as provides insight into the geopolitical history of the exchange of the works of art.

In ARIES, the user can create geocharts in two different ways: with a group of images or with an image. A geochart of one image displays the map of the world with color markers overlapping countries presented in the provenance metadata of the image. Hovering over the markers, the system provides a visualization of markers in more detail. A geochart for a group of images consists of a heat map that represents the number of artworks in each country (Fig. 7). In both cases, a color legend is shown in the bottom part of the canvas. Geocharts are displayed on lightbox canvas, while all other elements are occluded.

The need for search, sort, timeline and geocharts was not identified during the requirements gathering phase. Instead, it was elicited as art historians interacted with the ARIES prototype, and tried to explore and visualize a large number of images.

## Basic Workflow: Overview

Tasks carried out in ARIES follow a simple workflow consisting of a series of user actions. First, users create a project and upload the images that they will work with; the images' thumbnails are then shown in the image menu view. Users can drag the images they wish to compare over to the lightbox canvas view. New images can be uploaded and dragged from the image menu at any point. By using the image exploration tools, images inside the lightbox canvas are compared, tags are created to record discoveries, and groups are built to organize the images following user-defined criteria. Metadata is added to images in lightbox canvas, image menu or group menu. The user can browse through the images using their metadata to create timelines and geocharts. The project is then saved and shared. Note that these steps can be rearranged or skipped (it is possible that a user may want to simply compare two images, or organize image groups by visual recognition alone).

## IMPLEMENTATION

ARIES was implemented as a web application. We opted for a web-based application for portability and usability—users need not go through the complexities of installing a system. This is especially important for our target users, who have little or no expertise in computing.

The client side was written in JavaScript and HTML5. The jQuery library was used to simplify the JavaScript coding. With the addition of the canvas element with the release of HTML5, JavaScript became a much more viable platform to handle the interactions we needed for the application. The main user interactions are handled through multiple HTML5 canvas elements; the

lightbox canvas view uses the Fabric.js canvas library. Fabric provides an image-handling structure that simplified our implementation. Communication with the application server is handled through AJAX, by making GET and POST HTTP requests. The data transferred between the client and the server is encoded in JSON. The dropbox.js library was used to access the Dropbox Core API and establish communications with the Dropbox servers. The use of Dropbox allows ARIES to support seamless sharing among users through a widely used platform. Other sharing platforms can be easily supported. We implemented the timeline and geochart functionality using Timeline.js and Google GeoChart libraries, respectively.

Since the goal of this work was to design a prototype system to help art historians to visually manipulate and organize art images, we focused our efforts on the client development. To achieve a functional but simple implementation, the server side of ARIES was written in Python and used the framework CherryPy to handle the server deployment and communication with the client. For simplicity, we did not use a database to store the images. Instead, they are maintained in the file system; for each image loaded in a project, ARIES automatically generates and assigns a unique ID. This information is maintained in a JSON record stored on the server side. Every image request issued by the client is based on the currently available IDs. To achieve interactive performance for the client–server communication, we decided to keep two versions of each loaded image: the original resolution and a thumbnail version with a maximum dimension of 128 pixels. The thumbnail version is used to compose the interface menus while the original ones are used when the user loads an image to the lightbox canvas.

## CASE STUDIES

In this section we present two case studies performed by professional art historians, co-authors of this article. They demonstrate how ARIES has helped them in their daily workflow.

### Designing an Exhibition

The following case study shows how the ARIES system helped curators working in two different countries to collaboratively plan, organize and design an exhibition.

#### Background

In order to celebrate the 450th anniversary of the birth of the 17th century Flemish artist Jan Brueghel the Elder (1568–1625), one of the most prominent and successful Flemish artists of his time, an exhibition is planned in the city of his birth, Antwerp, Belgium, in 2019. This exhibition will be the first devoted to his drawings. In the past, his works on paper have been presented rather as supplements to his paintings, and yet it was in drawings that he introduced some of his most revolutionary artistic solutions to landscape art, solutions that influenced an entire generation of artists in both the Southern and Northern Netherlands.

The exhibition will also correspond with the launch of an online catalog of all the artist's known drawings. The catalog and the exhibition are being created jointly by a curator in New York and a curator in Budapest. As a result, many issues arose around the successful sharing of information, ideas, and especially, images. With the help of ARIES, many of these issues have been solved.

As we discussed previously, art historians need to compare, analyze and organize images of works of art. For the Brueghel exhibition, we first needed to decide on a general theme or sequence for the works. Then, we had to gather all the images together in one screen in order to get a sense of what themes the artist was interested in. While currently available tools allow us to do this in a linear fashion, once the images are on one screen, we needed to be able to move them around and create different relevant groupings. With Adobe Bridge, probably the best currently available tool for this type of work, we were able to move images around, but Adobe snapped the images to a grid, making them hard to work with. Further, Bridge did not allow us to separate groupings of images; this had to be done by moving groups of images out of the main screen and into separate folders. This meant that all the images were no longer visible at once on the screen

in their new groupings, so that individual folders had to be opened to see them, and the context of the entire oeuvre of the artist was lost.

Once curators had created their sections of the exhibition, the next step was to share their ideas. Without ARIES, this would have been done by creating a text document, or using Bridge to create PDF files created for each folder. The problem with this process was that the receiving curator had to manually recreate the groups in her environment. Since this back and forth continues for quite some time, the process is quite laborious.

### Designing the exhibition with ARIES

With ARIES, collaboration is much easier and more intuitive. Curator A in Budapest uploaded all the images by the artist into the system and, once there, could move them freely around on the lightbox canvas and group them in a thematic fashion (Fig. 2). After saving the project, curator A shared it with curator B in New York. In this way, curator B was able to open the project and see exactly the same screen as A. Rather than having to recreate the subject groups from a typed list or a PDF file, curator B was immediately able to see precisely which images curator A had placed in which subject groups. Curator B in turn opened each grouping individually on the lightbox canvas, and began working with the groupings herself, adding or subtracting works as desired. The back and forth process continued seamlessly until the curators had agreed on the works for each section of the exhibition.

Once the sections were decided, it was necessary to determine the order the works would be hung in. In this case, the curators wanted to hang them in chronological order, in order to follow the way the artist depicted the various themes throughout his lifetime. To assist in this task, ARIES created a timeline of the works of art and displayed them in chronological order.

Using ARIES, the curators were also able to automatically display the images in relative size to each other on the lightbox canvas (Fig. 2). A relative size display such as this allowed the curators to have a fairly accurate visual idea of what the actual hanging might look like. Once this display was made, again, the two curators were able to save it and send it back and forth to each other for further discussion. The importance of this aspect of ARIES cannot be emphasized enough. Currently there are no tools that do this. To get a sense of the sizes of paintings and how they will look next to each other in an exhibition, curators now make printouts of images of the works of art in relative size and hang them on physical mock-ups of the exhibition walls.

For designing the exhibition, the metadata view of ARIES was also a very important component. This view let the curators add titles, dimensions, medium, techniques, provenance or location, keywords, and any important annotations regarding the images. This allows curators to share the cataloging of the works, as well as communicate thoughts and ideas about a particular work with each other through the sharing aspect of ARIES. One last way ARIES was helpful in the creation of the Brueghel exhibition was in the process of asking for loans from other institutions. ARIES helped map the locations of the drawings with its geochart tool (Fig. 7) and a list of possible loans was more easily drawn up from the resulting map.

## Analyzing Reproductions of Works of Art

The following case study demonstrates how art historians on our team used ARIES to identify works of art from Mary Jane Morgan's collection, find new information on the history of their ownership (provenance), visualize current locations of the paintings, and make a stylistic analysis of various etched copies by other artists published in promotional materials for Morgan's 1886 auction.

### Background

Morgan's collection included contemporary paintings from Europe and the United States with an emphasis on the French Academic School. Academic artists often copied themselves or made close variants of a given composition: altering the size of the work of art, medium employed (pastel, watercolor, oil on panel, and so forth), or making slight adjustments in the composition

such as changing the direction of a person or an animal depicted. Moreover, many of the paintings were then reproduced as limited edition or unlimited edition prints, and inspired copies by other artists. Trying to identify which version of a work of art was included in Morgan's collection was thus very difficult.

Adding to this confusion, titles of paintings were often translated from French, German, Spanish, Dutch, and so on, to English and would even change over time. Further challenges exist because provenance information for the works of art held in both private and public collections are not always complete, measurements not always precise, and attributions change over time. All of these factors made tracing the collection an even more difficult task.

A *New York Times* article, published on 1 March 1886, critiqued the etching reproductions of the paintings included in the auction catalog as to whether or not they were close representations of the original works of art. They had judged that some of the etchings captured the spirit of the original paintings more than others. By utilizing the two surviving visual documentations of the paintings from Morgan's sale, the 1886 auction catalog "The Art Collection" formed by the late Mrs. Mary J. Morgan, and the art critic Charles de Kay's *The Magazine of Art* article "An American Gallery," published in the same year, we were able to make our own comparisons and judgments. While the representations of the paintings found in the auction catalog and the magazine were often close in appearance to the original works of art, they did not match exactly—they were interpretations by other artists working in different mediums.

### Using ARIES to illuminate a private collection

ARIES allowed us to bring together both images of the 1886 etched reproductions of the works of art and their photographic surrogates, and superimpose them for a close analysis. We were able to take modern photographs of paintings we believed had been in Morgan's collection and compare them with etchings made at the time of the auction. In the past we would have performed this task in PowerPoint, where each image has to be adjusted individually. While working in ARIES, this function was performed automatically, thus saving us a considerable amount of time. Further, ARIES provided other tools not available in PowerPoint, such as the relative size, lens, and matching point tools, which helped with other aspects of the comparison. In the past we would have also utilized Pinterest to display images from the collection; however, Pinterest would not have allowed us to reorganize images at will—images are displayed only in fixed groups according to the order they are uploaded. To alter the order of display, we had to create entirely new collections, which was a time-consuming process. With a click of a button, in addition to allowing us to bring together images in one space and annotate them, ARIES allowed for seamless visualizations demonstrating the global redistribution of the paintings that once formed part of Mary Morgan's collection.

### Comparing reproductions to images of actual works of art

In the aforementioned 1886 *New York Times* article, W.H. Shelton's (1840–1932) renderings of Jean L.E. Meissonier's (1815–1891) painting "The Vidette," 1812 (lot #227 in the sales catalog) and J.B.E. Detaille's (1848–1912) "A Flag-Officer" were reported as favorably representing the paintings. When a photograph of the Meissonier painting and an image of Shelton's etching were analyzed in ARIES, displayed side by side using the lens tool, a different story was revealed. Further, when the dynamic overlay was employed, using the horse and vidette as reference points (Fig. 3), it was obvious that Shelton had all but forgotten the landscape in his composition. In his rendition, he prioritized the horse and soldier. On the other hand, when Shelton's etching of Detaille's "A Flag-Officer" is compared in ARIES using the lens tool, we observed strong similarities and successful representation of the original work of art. This left us with new art historical questions to ponder, such as why the *New York Times* felt that Shelton's etching was a faithful representation of Meissonier's painting.

In another example we viewed a digital image of a painting from the sale, lot #170, Jean-François Millet's "The Churner," beside etchings created by other artists printed in auction catalogs for the purpose of selling the painting. The etchings examined were from sales catalogs dated

from 1873, 1879, and 1886. The etching included in the 1886 catalog was by the American artist William Merritt Chase (1849–1916), and the etchings included in both the 1873 and 1879 catalogs were by the French artist Adolphe Potemont Martial (1828–1883). We were able to make close observations and track stylistic changes of the works of art by utilizing the lens tool in ARIES—matching corresponding points and then navigating to different corresponding areas of the images simultaneously. The rectangle tool allowed us to get a closer look at the works of art and facilitated a comparison of the 1886 Chase etching, the Millet painting, and the earlier French etchings (Fig. 5). The heat map demonstrated where the works of art matched up and where they deviated from one another. In ARIES we were able to make a quick observation that the French etching of Millet's painting tended to retain the softness and overall tone of the original work of art, whereas the American rendition appeared edgier, more confined; the churner seemed oppressed, as if working in a dark cellar. After adding the dates of creation and creating a group of images to visualize, the art historian easily created a timeline (Fig. 6). This timeline accentuated the stylistic differences and changes of one theme over time both by the original artist and those creating copies after his work in different mediums.

### Discovering missing provenance information

With the help of ARIES, we were able to identify and fill gaps in the provenance for William-Adolphe Bouguereau's "Nut Gatherers," lot #154 from Morgan's sale, which now resides at the Detroit Institute of Arts. The provenance listed on the museum's website does not list ownership before Morgan's 1886 sale. Recently we discovered a photograph of a similar work of art in a 19th century scrapbook of photographic reproductions of paintings that were once offered for sale through Samuel Putnam Avery at his New York gallery at 368 Fifth Avenue. The scrapbook, located in the Rare Books Department at the Metropolitan Museum of Art, has been digitized, allowing us to compare a digital image of that painting with one from the painting at the Detroit museum to see if they represented the same work of art. By utilizing the lens and match-point tools in ARIES, we were able to compare the images closely (Fig. 4). In ARIES, the 19th century photograph and the digital image of the painting corresponded to one another perfectly, which suggests that Samuel Avery, a person known to have sold pictures to Morgan, once owned the picture, a fact which we can now confidently add to the painting's provenance narrative.

## DISCUSSION

### System design

The system we designed and implemented combines several interaction mechanisms that were adapted to meet the requirements we gathered from the art historians. For example, the lens tool is an adaptation of magic lens that makes it easier for users to compare pieces of different images; and the image menu was inspired by Microsoft PowerPoint in making available many images on the left side of the screen easily selectable for analysis and further comparison. An important contribution of this work comes from combining and adapting these mechanisms into a novel unified system that fills an important gap created by the wide adoption of digital images in art research.

### Expert feedback

For art historians, a unique benefit of ARIES is that it provides a dedicated out-of-the-box image workspace with an intuitive interface that supports exploration, manipulation, annotation, grouping and sharing of art images in a single environment that can be shared with multiple users. Those to whom it has been demonstrated have wondered why no one had thought of it before, wish it had been available for past projects, and desire a copy immediately. Using existing technology, art historians working with images on computers have had to go through an extremely clunky and laborious process.

## Lessons for the community

In moving to a new medium, art historians lost access to a useful “tool”: the physical light box. By mimicking the light box and adding features that are enabled by digital media (for example, file organization, image comparison and manipulation), we were able to provide a much improved user experience. Not only has ARIES made it easier for art historians to perform common tasks, but it has also enabled new unexpected uses. The collaborative design of an exhibition across different continents is an example of a novel use art historians discovered, which clearly was not possible with physical light boxes—as discussed earlier, two art historians used ARIES to work together and interactively on planning and organizing an exhibition.

While ARIES was specifically designed to support art history research, the system is general and can potentially be useful to organize and compare digital images by professionals (and researchers) in different areas, including designers, photographers, artists, scientists, and digital humanists. For example, one of our collaborators used ARIES to check printers’ proofs for an event flyer. Photographers and artists might use the tool to compare, contrast and manipulate images as part of their creative process. Historians comparing scans of multiple versions of both printed texts and incunabula might also find the system useful. In the DARPA Memex program,<sup>18</sup> large volumes of images need to be analyzed as experts look for signals of human trafficking in pictures present in online escort ads. Such signals can be found in body features that may indicate an individual is underage, particular markings (for instance, tattoos), or the presence of photos of different people that have the same background (for example, this may suggest multiple individuals in the same organization). ARIES can help experts organize and compare images, streamlining the analysis and the creation of learning classifiers that automatically identify the signals.

As in any interdisciplinary collaboration, we have also faced challenges, including unfamiliarity with each other’s work and the need to learn each other’s “language.” Once we were able to overcome some of these barriers and applied a computer science perspective to the problem, together we were able to develop and refine new ideas that led to an effective solution.

## CONCLUSION

In this article, we introduced ARIES, an innovative system that helps art historians explore and organize collections of images of artworks. Besides discussing our design decisions and describing the system, together with its implementation, we presented case studies carried out by art historians that demonstrate the effectiveness of ARIES and how it greatly simplifies many of the tasks required for art history research and practice.

While the initial prototype provides a significant step toward our original goal, for a wide deployment, several improvements are needed. For example, currently, the user has to manually input metadata for the images; we plan to support additional mechanisms that allow metadata to be imported from different data sources. To improve the collaboration experience, we would like to support real-time interactions, where multiple users can simultaneously work on a shared project and see each other’s modifications. Another direction we would like to explore is the use of ARIES in different domains.

## ACKNOWLEDGMENTS

We thank Professor Dana Karwas and her students, who used our initial requirements for ARIES in their user experience course, and in the process helped develop the initial user experience mock-ups of the system, which served as foundation for our work. We also thank Lucas Brilhante, Felipe Soares Leite, Fernando Silva, and Autumn Wu who contributed to the implementation at different phases of the project. Maria Beatriz Silva narrated the ARIES video. This work was supported in part by the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Google Faculty Award, an IBM Faculty Award, the Moore-Sloan Data Science Environment at NYU, AT&T, DARPA award

FA8750-14-2-023, NSF awards CNS-1229185, CNS-1405927, and CCF-1533564. Opinions, findings and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or DARPA.

## REFERENCES

1. B. Fuchs, L. Isaksen, and A. Smith, “The virtual light box for museums and archives: a portlet solution for structured data reuse across distributed visual resources,” *Museums and the Web*, 2005.
2. A.C. Smith, B. Fuchs, and L. Isaksen, “Vlma: A tool for creating, annotating and sharing virtual museum collections,” *Digital Medievalist*, vol. 4, 2008.
3. “Cultural institute,” [www.google.com/culturalinstitute/home](http://www.google.com/culturalinstitute/home).
4. J. Saklofske, “Newradial: Revisualizing the Blake archive,” *The Journal of the Initiative for Digital Humanities, Media, and Culture*, vol. 2, no. 1, 2010.
5. G. Ciocca, P. Olivo, and R. Schettini, “Browsing museum image collections on a multi-touch table,” *Information systems*, vol. 37, no. 2, 2012, pp. 169–182.
6. U. Hinrichs, H. Schmidt, and S. Carpendale, “Emdialog: Bringing information visualization into the museum,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, 2008, pp. 1181–1188.
7. E. Hornecker, “I don’t understand it either, but it is cool–visitor interactions with a multi-touch table in a museum,” *3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems*, 2008, pp. 113–120.
8. P. Davis et al., “Going deep: Supporting collaborative exploration of evolution in natural history museums,” *Proceedings of 10th International Conference on Computer Supported Collaborative Learning*, 2013.
9. T. Geller, “Interactive tabletop exhibits in museums and galleries,” *IEEE Computer Graphics and Applications*, vol. 26, no. 5, 2006, pp. 6–11.
10. H. Bay, T.uytelaars, and L. Van Gool, “Surf: Speeded up robust features,” *Computer vision–ECCV 2006*, Springer, 2006.
11. D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, 2004, pp. 91–110.
12. M.-J. Lobo, E. Pietriga, and C. Appert, “An evaluation of interactive map comparison techniques,” *Proceedings of the 33rd annual ACM symposium on User interface software and technology (CHI)*, 2015, pp. 3573–3582.
13. M. Elias et al., “Do I live in a flood basin?: Synthesizing ten thousand maps,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2008, pp. 255–264; <http://doi.acm.org/10.1145/1357054.1357100>.
14. M. Gleicher et al., “Visual comparison for information visualization,” *Information Visualization*, vol. 10, no. 5, 2011, pp. 289–309.
15. T. Tominski et al., *A survey on interactive lenses in visualization*, report, EuroVis State-of-the-Art Reports, 2014.
16. A. Cockburn, A. Karlson, and B.B. Bederson, “A review of overview+ detail, zooming, and focus+ context interfaces,” *ACM Computing Surveys*, vol. 41, no. 1, 2008, p. 2.
17. R.C. Gonzalez and R.E. Woods, *Digital Image Processing (3rd Edition)*, Prentice-Hall, 2006.
18. “Darpa memex program,” [www.darpa.mil/program/memex](http://www.darpa.mil/program/memex).

## ABOUT THE AUTHORS

**Lhaylla Crissaff** is a professor in the Department of Geometry at Universidade Federal Fluminense, Brazil. Her research interests cover most aspects of visual computing, especially scientific and information visualization. Lhaylla is the co-author of the book *Geometria Analítica* (in Portuguese), published by Brazilian Mathematical Society in 2013 and she is currently working on a second volume of the book. Lhaylla received her PhD in applied mathematics from Pontifícia Universidade Católica do Rio de Janeiro, Brazil, in 2009. Contact at [lhayllacrissaff@id.uff.br](mailto:lhayllacrissaff@id.uff.br).

**Louisa Wood Ruby** is head of photoarchive research at The Frick Art Reference Library where in addition to being the onsite specialist for Dutch and Flemish Art, she is charged with increasing public awareness of methodological and scholarly trends in digital humanities and art history. She received her doctorate from New York University's Institute of Fine Arts, and has published widely in her field. Recently she has been working on an exhibition of drawings by Jan Brueghel the Elder, and was elected vice president of the Historians of Netherlandish Art. Contact at [WoodRuby@frick.org](mailto:WoodRuby@frick.org).

**Samantha Deutch** is an assistant director of the Center for the History of Collecting at The Frick Collection, The Frick Art Reference Library and is the author and editor of the Center's award winning *Archives Directory for the History of Collecting* (Art Libraries of North America Worldwide Books, 2011). She coordinates the center's research fellowships and-fellows' forums, is responsible for developing research tools, and is a member of the Library's Digital Art History Lab. Currently she is serving as Secretary on the Art Libraries of North America's (ARLISNA) executive board and the sub-committee chair for ARLISNA's Sotheby's Institute of Art Research Award. Additionally, she serves on the steering committee of the New York Digital Art History Group (NYC-DAH), and has been working collaboratively on developing tools for image analysis. Contact at [deutch@frick.org](mailto:deutch@frick.org).

**R. Luke DuBois** is associate professor of Integrated Digital Media at NYU's Tandon School of Engineering. He holds a doctorate in music composition from Columbia University and researches topics ranging from user experience to citizen science to assistive technology. A practicing artist and composer, his work uses techniques drawn from data science to create new forms of artistic expression. Contact at [dubois@nyu.edu](mailto:dubois@nyu.edu).

**Jean-Daniel Fekete** is senior research scientist (DR1) at INRIA, scientific leader of the INRIA Project Team AVIZ that he founded in 2007. He received his PhD in computer science in 1996 from Université Paris-Sud. His main research areas are visual analytics, information visualization and human computer interaction. He is currently the chair of the IEEE Information Visualization Conference Steering Committee, member of the IEEE VIS Executive Committee, member of the Eurographics EuroVis Steering Committee, and member of the Eurographics publication board. He is also an ACM Distinguished Speaker. He was the General Chair of the IEEE VIS Conference in 2014, the first time it was held outside of the USA in Paris, associate editor of the *IEEE Transactions on Visualization and Computer Graphics* (TVCG) 2011–2015, the President of the French-Speaking HCI Association (AFIHM) 2009–2013, he was conference chair of IEEE InfoVis 2011, and paper co-chair of IEEE Pacific Visualization 2011. Contact at [Jean-Daniel.Fekete@inria.fr](mailto:Jean-Daniel.Fekete@inria.fr).

**Juliana Freire** is a professor of computer science and engineering and data science at New York University. She holds an appointment at the Courant Institute for Mathematical Science, is a faculty member at the NYU Center for Urban Science and at the NYU Center of Data Science, where she is also the Director of Graduate Studies. She is the executive director of the NYU Moore-Sloan Data Science Environment. Her recent research has focused on big-data analysis and visualization, large-scale information integration, web crawling and domain discovery, provenance management, and computational reproducibility. Freire is an active member of the database and web research communities, with over 170 technical papers, several open-source systems, and 12 US patents. She is an ACM Fellow and a recipient of an NSF CAREER, two IBM Faculty awards, and a Google Faculty Research award. She has chaired or co-chaired workshops and conferences, and participated as a program committee member in over 70 events. Her research grants are from the National Science Foundation, DARPA, Department of Energy, National Institutes of Health, Sloan Foundation, Gordon and Betty Moore Foundation, W.M. Keck Foundation, Google, Amazon, AT&T, the University of Utah, New York University, Microsoft Research, Yahoo! and IBM. Contact at [juliana.freire@nyu.edu](mailto:juliana.freire@nyu.edu).

**Claudio T. Silva** is a professor of computer science and engineering and data science at New York University. His research includes the intersection of visualization, data analysis, and geometric computing, as well as the analysis of urban and sports data. In 2013, he was elected as an IEEE Fellow and in 2014 he won the IEEE Visualization Technical Achievement Award. He helped develop MLB.com's Statcast player tracking system, which won the Alpha Award for Best Analytics Innovation/Technology at the 2015 MIT Sloan Sports Analytics Conference. In 2015, he was elected chair of the IEEE Technical Committee on Visualization and Computer Graphics. Contact at [csilva@nyu.edu](mailto:csilva@nyu.edu).

# 10 Years of MegaMol

## The Pain and Gain of Creating Your Own Visualization Framework

**Michael Krone**  
University of Stuttgart

**Sebastian Grottel**  
TU Dresden

**Guido Reina**  
University of Stuttgart

**Christoph Müller**  
University of Stuttgart

**Thomas Ertl**  
University of Stuttgart

**Editor:**  
Andre Stork  
[andre.stork@igd.fraunhofer.de](mailto:andre.stork@igd.fraunhofer.de)

This article discusses our experience in creating MegaMol, an open-source visualization framework for large particle-based data.

Research in scientific visualization often utilizes small software prototypes for proof-of-concept implementations, performance evaluation, and image generation. In large-scale and particularly in long-running research projects, especially aiming at big data, these prototypes quickly reach their limitations, at least as soon as scientists from application domains start using the software. In this position paper, we discuss our experience in creating MegaMol,<sup>1</sup> an open-source visualization framework for large particle-based data. This article is an extended version of our contribution to the IEEE Vis 2016 workshop “Visualization in

Practice.” MegaMol is developed at the University of Stuttgart, Germany by Ph.D. students and postdoctoral researchers, without the help of dedicated programmers. The development of the MegaMol visualization system started in 2006. It originates from a joint research project between biologists, physicists, material scientists and visualization experts working with large, particle-based data. Within this project, we faced the need for high-performance GPU-accelerated computer graphics to achieve diverse interactive visualizations for our project partners. Having started as research prototype for high-performance computer graphics, MegaMol has grown over more than ten years and has become a software actively used by several Ph.D. students and collaboration partners from physics, biochemistry, thermodynamics and material science. We designed our software to be modular and to support rapid prototype development, that is, to provide the necessary flexibility to visualization researchers. In addition, this approach allows central parts of the software to reach a mature and stable state. Such continuity enables the software to be usable by application domain scientists, which is usually not the case for typical research prototypes illustrating one single concept or algorithm.

This article also highlights aspects of the development process and long-term software architecture decision-making, which we identified to be critical to successfully deliver stable software releases. We discuss the increased workload required for defensive programming, error handling, testing, and software maintenance, but also the benefits of having such a framework.

## DEVELOPMENT & DESIGN PRINCIPLES

MegaMol is written in C++, supporting Windows and Linux platforms, and primarily uses OpenGL and CUDA. The framework is strongly tailored towards fast rendering. Consequently, the software itself only provides a thin layer above OpenGL avoiding overhead during rendering while exposing low-level functionality required for cutting-edge graphics programming. Thus, a central paradigm is that data management and layout on the CPU side has to fit the requirements of the GPU and strictly follows a zero-copy paradigm. Here, MegaMol significantly differs from other visualization frameworks existing when we started the development.



Figure 1. MegaMol is capable of supporting tiled displays, like the VISUS powerwall at the University of Stuttgart. The image shows a prototype of a distributed particle renderer. The data set is a cosmological simulation of dark matter comprising 1 billion particles (a subset of the data provided for the 2015 IEEE SciVis contest, <http://darksky.slac.stanford.edu/scivis2015/>).

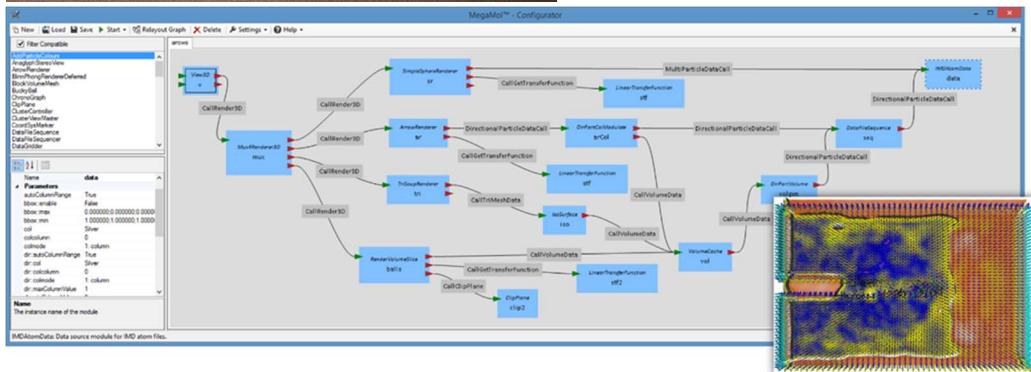


Figure 2. The MegaMol Configurator application for designing module graphs. The depicted module graph generates the visualization shown as inset containing glyphs (spheres and arrows), a scalar field visualization, and isosurfaces. The visualized data set is the simulation of a crack propagation in solid metal. The combination of all three visualizations allows for efficient and effective observation of the data, providing overview (through color and isosurfaces) as well as detail information (through glyphs).<sup>2</sup>

The architecture supports different front ends. For instance, there are GUIs for specific application tasks using the MegaMol back end as engine. The software itself also supports synchronized execution on clusters (e.g., for rendering on large displays as shown in Fig. 1). All functionality in MegaMol is provided as so-called Modules. Modules can be arranged in arbitrary call graphs, where control flow and data flow follow a pull paradigm. An in-depth description of the architecture can be found in the work of Grottel and colleagues.<sup>1</sup>

### Modularization

The Modules, which encapsulate functionality, communicate via strongly typed channels dubbed Calls. Each Module can have an arbitrary number of typed user-adjustable parameters (e.g., numerical values, enumerations, strings, or file names). Developers explicitly specify via which Calls Modules communicate. This approach has several benefits: Task-specific functionality can be composed from an ever-growing set of operations and visualizations provided by existing

Modules (see the example shown in Fig. 2). The separation into Modules establishes strong code ownership, minimizing the potential for conflicts.

A plugin system provides further abstraction, allowing developers to group related Modules and Calls. This allows for removing specialized functionality or specific implementations from MegaMol's core library. This way, the core library retains as few dependencies as possible.

## Defensive Programming & Client-site Debugging

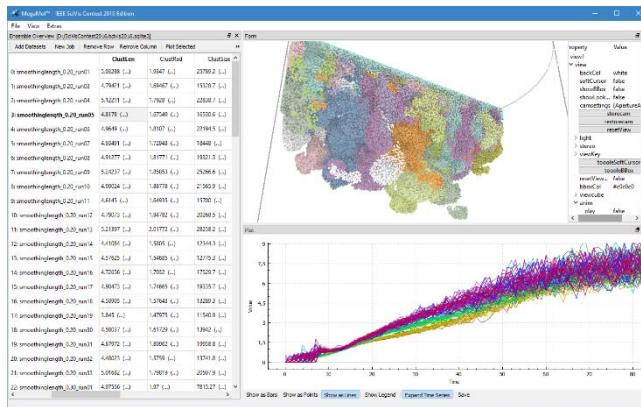
Incomplete error handling might be the most severe issue of research prototypes. Reusable software requires defensive programming, handling errors and exceptions where appropriate to ensure consistent state and spot errors. Besides standard mechanisms like assertions or tracing, MegaMol supports the developer with corresponding utilities to investigate and resolve issues. MegaMol also provides a logging mechanism allowing non-expert users to provide detailed feedback to the developers in an error case.

## Run-time Configuration & Usability

The graph of modules determining the actual functionality can be created using a graphical utility, the MegaMol Configurator shown in Fig. 2. This is a stand-alone C# software, fully supported on Linux using Mono. The Configurator allows users to construct only valid module graphs. All parameter values of the Modules can be edited at runtime within MegaMol, but can also be set to initial values using the Configurator.

The default front end of MegaMol dynamically generates a GUI from the active modules. GUI controls are chosen based on the type of the parameters and are grouped hierarchically. While this gives the user full control over all parameters, no task-based structure exists in this GUI. Therefore, we have created task-specific GUIs for certain applications that present only the relevant parameters to the user and offer higher-level widgets for more complex tasks and interactions. An example is shown in Fig. 3.

Besides a GUI, documentation is an important factor for usability of software. The MegaMol project website provides a wiki and manuals including instructions for building and using MegaMol, as well as tutorials and example files and projects.



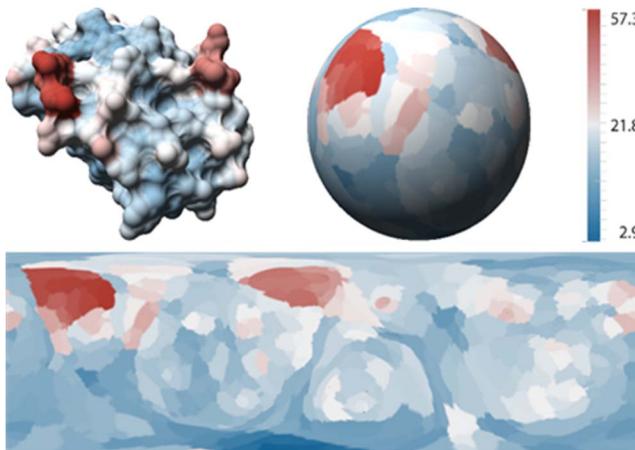
**Figure 3.** The primary MegaMol user interface developed for the 2016 SciVis contest, featuring a tabular database view (left), a 3D rendering of the selected data set (top right), and additional information via 2D plots (bottom right).<sup>3</sup>

## DEPLOYMENT

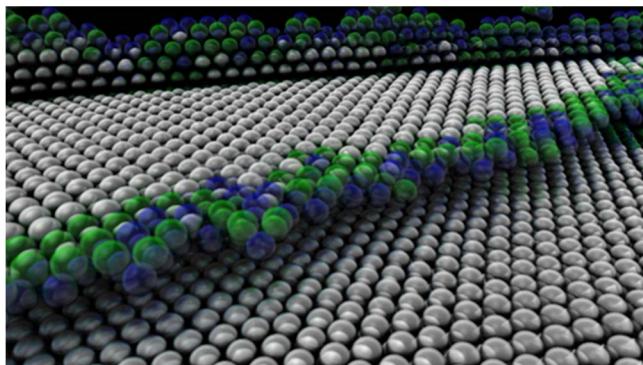
MegaMol is primarily distributed via the project website (<http://www.megamol.org>). On this website, we provide packaged releases and links to the source code repositories, which we have recently consolidated and migrated to GitHub (<https://github.com/UniStuttgart-VISUS/megamol>) for the release of version 1.2. Although the main development takes place on Windows using Microsoft Visual Studio, we opted for the widely used CMake build system (<https://cmake.org/>) due to the necessity of Linux support to users. Of course, since MegaMol is

open source, our code can be reused by third parties in other software projects, which has already happened in the past.

A large software framework like MegaMol, aimed both at developers as well as at end users, requires a substantial amount of maintenance. We use bug tracking and feature request systems to organize tasks. The main developers of MegaMol are Ph.D. students and postdoctoral researchers, that is, scientists who are not intended to spend much time on system development and maintenance. However, not fixing bugs or adding new critical features will make a software project lose users rapidly, all of them in the worst case. For many of our visualization research projects, however, having users was hugely beneficial or even crucial in the past. We, therefore, have to balance carefully how much work to put into research and how much into maintenance. This can be seen as a long-term investment, which will be discussed in the next section.



**Figure 4.** Examples for biomolecular visualization in MegaMol: Solvent Excluded Surface, Molecular Surface Globe, and Molecular Surface Map of a protein colored by temperature factor. The legend shows the cool-warm color gradient mapped to the temperature factor values.<sup>4</sup>

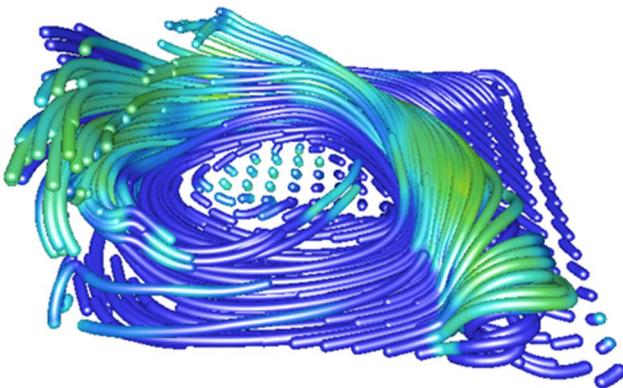


**Figure 5.** Interactive visualization of spherical particles rendered with transparency and ambient occlusion within a consistent lighting model.<sup>5</sup>

## DISCUSSION & CONCLUSION

When we started developing MegaMol more than a decade ago, we clearly underestimated the huge workload this would cause. Repeatedly, the question was raised whether a Ph.D. student could afford this non-research-related workload. In hindsight, we believe it is worth to develop and maintain your own code base for problems not yet solved satisfactorily. Given the range and complexity of visualizations we can now create with MegaMol, we find our choice well justified. All participating institutions and people are today satisfied with the result, concerning both software and scientific publications. MegaMol has been used successfully in many research projects that involved particle visualization in the past. Examples include the visualization of particle simulation data from cosmology (see Fig. 1), material sciences (see, e.g., Fig. 2 & 3), and structural biology (see Fig. 4). We also used MegaMol two times for the annual IEEE SciVis contest (2012 and 2016) and were able to win the contest in both occasions.<sup>6,3</sup> MegaMol was instrumental in creating our contest entries, as we were able to reuse existing modules for particle visualizations and the underlying rapid prototyping idea enabled us to develop new visualizations

efficiently. However, the functionality of MegaMol has evolved beyond simple particle visualization as shown in Fig. 5. Nowadays, MegaMol also includes for example volume rendering, flow visualization (see Fig. 6), or information visualization and visual analytics. Especially the latter has become increasingly important for the visual analysis of large, multivariate simulation data sets. Furthermore, our experience is that having a stable framework has a positive effect on the quality of the research prototypes, making them more usable, more stable, and better testable. As mentioned above, another positive effect is that the development of complex prototypes is less effort once the framework is mature, since a lot of functionality is readily available. One prominent example in our case is the built-in support for tiled displays like the VISUS Powerwall shown in Fig. 1. Adding such functionality to standalone research prototypes is usually accomplished via copying and pasting of respective code snippets, which is more work and typically more error-prone.



**Figure 6.** Flow field stream tubes generated by point-based ray casting of implicit surface descriptions on the GPU.

What remains difficult is the acceptance and appreciation of the scientific community for a framework like MegaMol. While we are still convinced that it helps to be more productive once the framework is mature, the cost of the development work invested in a software system cannot be mitigated through publications since these would be engineering-centered. We believe the most important thing for a scientific software project is the continuous commitment of the individual developers against all odds. This includes sacrificing time for the development and maintenance of the framework, which could otherwise be invested in a research project.

Another issue that we want to raise here is that the development of a framework like MegaMol within a university research project is always people-centric. For us, staff equals researchers, mostly Ph.D. students. This leads to several problems, in addition to the abovementioned conflict of investing time in the framework instead of research. First, there is usually a limited persistence in staff, since Ph.D. students usually leave the university after graduation. Secondly, Ph.D. students typically have limited experience with large-scale software, since they are often fresh graduates. However, such a project requires a main architect at all times to stay on track. In our case, this post was filled by Sebastian Grottel, which led to a difficult situation after he left the University of Stuttgart after graduating in 2012. Sebastian went to TU Dresden, but the MegaMol project remained in Stuttgart. MegaMol was kept alive through the continued dedication of Sebastian and other people involved in the project. In hindsight, we should have planned ahead for that situation, however, the actual alternative remains unclear since it requires another person who invests time in the project. One solution could be the recently announced funding for software maintenance by the German Research Foundation or the involvement of other sponsors. We also briefly pondered the commercialization of the project, but discarded the idea due to the high risk. An important factor for the successful commercialization of MegaMol is a sufficiently large user base. This, however, leads to the conflict that a large user base requires dedicated people that can invest substantial time into user support, and without sufficient funds, it is impossible to hire these people speculatively. Conversely, if the user base grows suddenly, sufficient support has to be provided (and hired) at once. Otherwise, those new users would have to wait for the fulfillment of their requests, which makes it likely they will turn away from MegaMol as suddenly again.

## ACKNOWLEDGMENT

This work was partially funded by Deutsche Forschungsgemeinschaft (DFG) as part of SFB 716.

## REFERENCES

1. S. Grottel, M. Krone, C. Müller, G. Reina, and T. Ertl. MegaMol – A prototyping framework for particle-based visualization. *IEEE TVCG*, 21(2):201–214, 2015.
2. S. Grottel, P. Beck, C. Müller, G. Reina, J. Roth, H.-R. Trebin, and T. Ertl. Visualization of Electrostatic Di-poles in Molecular Dynamics of Metal Oxides. *IEEE TVCG*, 18(12):2061–2068, 2012.
3. P. Gralka, S. Grottel, J. Staib, K. Schatz, G. K. Karch, M. Hirschler, M. Krone, G. Reina, S. Gumhold, and T. Ertl. 2016 IEEE Scientific Visualization Contest Winner: Visual and Structural Analysis of Point-based Simulation Ensembles. *IEEE CG&A*, 2017. (accepted).
4. M. Krone, F. Frieb, K. Scharnowski, G. Reina, S. Fademrecht, T. Kulszewski, J. Pleiss, and T. Ertl. Molecular surface maps. *IEEE TVCG*, 23(1), 2017.
5. J. Staib, S. Grottel, and S. Gumhold. Visualization of particle-based data with transparency and ambient occlusion. *CGF*, 34(3):151–160, 2015.
6. K. Scharnowski, M. Krone, F. Sadlo, P. Beck, J. Roth, H.-R. Trebin, and T. Ertl. 2012 IEEE visualization contest winner: Visualizing polarization domains in barium titanate. *IEEE CG&A*, 33(5):9–17, 2013.

## ABOUT THE AUTHORS

**Michael Krone** is a postdoctoral researcher at the University of Stuttgart’s Visualization Research Center. Contact him at michael.krone@visus.uni-stuttgart.de.

**Sebastian Grottel** was a postdoctoral researcher at the TU Dresden’s Computer Graphics and Visualization lab. He was lead architect and main developer of the MegaMol framework.

**Guido Reina** is a postdoctoral researcher at the University of Stuttgart’s Visualization Research Center. Contact him at guido.reina@visus.uni-stuttgart.de.

**Christoph Müller** is a Ph.D. student at the University of Stuttgart’s Visualization Research Center. Contact him at christoph.mueller@visus.uni-stuttgart.de.

**Thomas Ertl** is a full professor at and the head of the University of Stuttgart’s Visualization Research Center and Institute of Visualization and Interactive Systems. Contact him at thom-as.ertl@vis.uni-stuttgart.de.

Contact department editor Andre Stork at andre.stork@igd.fraunhofer.de.

# Putting VR/AR to Work

**Amit Agrawal**

Kleene Closure Consulting

**Editor: Amit Agrawal**

amit.agrawal@me.com

Virtual Reality (VR) and Augmented Reality (AR) have received a lot of attention in the last few years. This article studies the current state of the VR/AR penetration in the enterprise and highlights a few examples where these technologies have been successful.

As virtual reality (VR) and augmented reality (AR) technologies endeavor to find their place in the consumer space, their contribution to the enterprise is becoming increasingly clear. The benefits accrued are (a) cost reduction and faster times to market by reducing the need to develop physical prototypes and finding defects early, (b) the ability to inexpensively study and address human-machine interface (HMI) issues, and (c) easy enterprise-wide collaboration and training.

Upskill Technologies, working with GE, used Google Glass to demonstrate performance improvement of a user undertaking equipment maintenance tasks by 34%, in a first-time use of the technology.<sup>1</sup> Other case studies from Boeing, GE Healthcare, and other firms have shown productivity improvements on an average of 32%. To put this in perspective, productivity growth in the US has averaged 0.5% from 2011 to 2016, compared to 3% from 1996 to 2005. Other notable examples include the use of Hololens at Japan Airlines to provide supplemental training for engine mechanics and flight crew trainees as well as Bechtel and Industrial Training International (ITI) partnering to train workers on operating cranes in VR.

We detail the expected future advancements below.

## REDUCED COST AND FASTER TIME TO MARKET

Building physical prototypes is expensive and time-consuming; replacing them with virtual prototypes allows for several iterations for the cost of a single prototype. This not only reduces cost but also makes iterations more efficient, thereby providing a faster time to market. As Jacques Delacour, CEO of OPTIS, a virtual prototyping company, puts it, “The goal is to have zero physical prototypes.”

Autodesk (Inventor, VRED), ESI Group (IC.IDO), and OPTIS (OMS2, SPEOS, Thea RT) all provide functionalities for VR in the enterprise, especially in the automotive and aerospace industries. Among these, OPTIS stands out for its focus on physically accurate simulations. For example, it treats light as multi-spectral energy rather than three-color channels, does a physically accurate simulation of light transport, and then maps the resulting energy into the device range by accurately simulating the human perceptual system. Among other things, it models

- The polarization of light required for simulation of glare.

- An average of about a 100 channels for light modeling, which can be decreased or increased depending on requirements.
- Mapping of high dynamic range energy information into the device range by simulating human perception. For example, OPTIS models glare depending on the person's age and can predict who will be able to read the displays.

This year, OPTIS launched light painting with Theia-RT 2017 (see Figure 1).



Figure 1. Theia-RT speeds up lighting design: designers paint the desired lighting, and the system automatically figures out the parameters of the lighting system. (Photo credit: OPTIS.)

## ENABLING NEW APPLICATIONS

Another area of advancement in AR/VR includes novel applications where none existed before or were even possible.

An example in this category is Aeroglass (<http://glass.aero>; see Figure 2), which provides a solution to pilots' unique need to visualize terrain, navigation, traffic, instruments, weather, and air-space information, with access to vital safety procedures and protocols—all within the confines of a cockpit.



Figure 2. AR navigation in a cockpit. Pilots can see aerial navigation data overlaid on top of their HMDs.

This solution can be used with various head-mounted displays (HMDs) including Osterhout Design Group's smart glasses (<http://www.osterhoutgroup.com/products>) and Epson Moverio, among others. The app won the best app of the Auggie Awards (<http://events.bizzabo.com/AWE2017/page/1007706/2017-augie-awards>) at the Augmented World Expo this year.

Similarly, 8ninth developed a Holographic Workstation for stock trading<sup>2</sup> (see Figure 3) for Citi Traders on the Hololens platform, which clearly demonstrates viable visualization and collaboration possibilities as the technology becomes more widely available.



**Figure 3.** Stock trading using AR. The stock trader (top) looks at the visualization of stocks to find a desirable trade. He then discusses it with his client (bottom) who views the close-up of the visualization and gives a go-ahead to proceed with the trade.

## IMPROVED IMMERSION

A third area of advancement in AR/VR is the ability to provide better immersive experience by having a larger field of view, an untethered experience, or a multi-sensory experience, especially haptics. Even though the advancements in this area are primarily being driven by consumer VR/AR, the results will benefit the enterprise, especially when dealing with HMI.

In VR, a large field of view is extremely important to provide immersion. Among the current players, HTC Vive provides the best field of view with a horizontal field of view at 100°. Two new players are VR Union (<http://vrunion.com>; headset: Claire VR) with a field of view of 170°, and Star VR (<http://www.starvr.com>) with a 210° horizontal and 130° vertical field of view.

See the sidebar, "Further Resources," at the end of the article for links to more tools and products in this area.

## CONCLUSION

Whether VR/AR succeeds in the consumer space—and we certainly hope that it does—it is clear that it is here to stay in the enterprise.

## REFERENCES

1. M. Abraham and M. Annunziata, "Augmented Reality Is Already Improving Worker Performance," *Harvard Business Review*, 2017; <https://hbr.org/2017/03/augmented-reality-is-already-improving-worker-performance>.
2. I. Kar, "Citigroup Wants to Bring Microsoft's Hololens and Augmented Reality to Stock Trading," *Quartz*, 2016; <https://qz.com/650872/citigroup-wants-to-bring-microsofts-hololens-and-augmented-reality-to-stock-trading>.

## ABOUT THE AUTHOR

**Amit Agrawal** is the principal at Kleene Closure Consulting. His research interests include machine learning and its applications, serious games, and non-photoreal rendering. Agrawal received a PhD in Geometric Modeling from the Institute of Robotics and Intelligent Systems at the University of Southern California. Contact him at amit.agrawal@me.com.

## SIDE BAR: FURTHER RESOURCES

A few companies provide add-ons to existing HMDs to make an untethered experience possible. Notable among these at the time of writing this article are TPCast (<http://uploadvr.com/tpcast-wireless-vive-impressions>), DisplayLink (<http://www.displaylink.com/vr>), and Sixa (<http://www.tomshardware.com/news/sixa-rivvr-wireless-vr-tested,34064.html>; Rivvr). The challenge is compressing the data streams, maintaining a low latency, while preserving a good experience. HP addresses the problem by putting the workstation in a backpack (HP Z VR Backpack G1 Workstation; <http://www8.hp.com/h20195/v2/GetDocument.aspx?docname=4AA7-0460ENUC>).

Even though a few companies are exploring full body suits for haptics (for example, AxonVR [<http://axonvr.com>], Synesthesia suit [<http://www.wired.com/2015/12/rez-infinite-vr-suit/>], Rapture Vest [<http://www.vrdb.com/hardware/rapture-vest>], and Teslasuit [<https://teslasuit.io/>]), for the enterprise Go Touch VR (<http://www.wareable.com/vr/go-touch-vr-haptic-finger-accessory-7765>), Tactical Haptic (<http://www.roadtovr.com/tactical-haptics-2-2-million-seed-investment-grant-haptic-vr-controller>), and Omnipulse (<http://venturebeat.com/2017/05/14/cornells-haptic-skin-gives-vr-a-more-human-touch/>) technologies may be the ones to watch because they may provide a better problem–solution fit.

# Optimizing Scientist Time through In Situ Visualization and Analysis

**John Patchett and  
James Ahrens**

Los Alamos National  
Laboratory

**Editor:**

Theresa-Marie Rhyne  
[theresamarierhyne@gmail.com](mailto:theresamarierhyne@gmail.com)

What would happen if an asteroid impacted the earth in the middle of the ocean? Would a giant tsunami engulf the distant coastlines? Galen Gisler, a Los Alamos scientist, designed a series of simulations using a Los Alamos National Laboratory (LANL) multiphysics simulation code that was run for thousands of hours on thousands of processing cores of LANL supercomputers. A critical part of answering these scientific questions was the process of gathering and analyzing simulation outputs. Our role as visualization and analysis developers at Los Alamos allowed us to work with Gisler and suggest the use of in situ processing to optimize his analysis time.

In situ processing produces reduced-size persistent representations of a simulation's state while the simulation is running. The need for in situ visualization and data analysis is usually described in terms of supercomputer size and performance in relation to available storage size. In situ processing is considered "critical technology for achieving scientific discovery at exascale."<sup>1</sup>

After in situ outputs have been produced, a domain scientist such as Gisler must study the output to gain scientific insight. Output types, sizes, and quantities affect the scientist's ability to gain understanding of the represented simulation. Different types, sizes, and quantities of data may require bigger, more extensive, and expensive compute resources for the process of understanding. The choices of types, sizes, and quantities affect the time the scientist must spend interacting with the data to gain understanding.

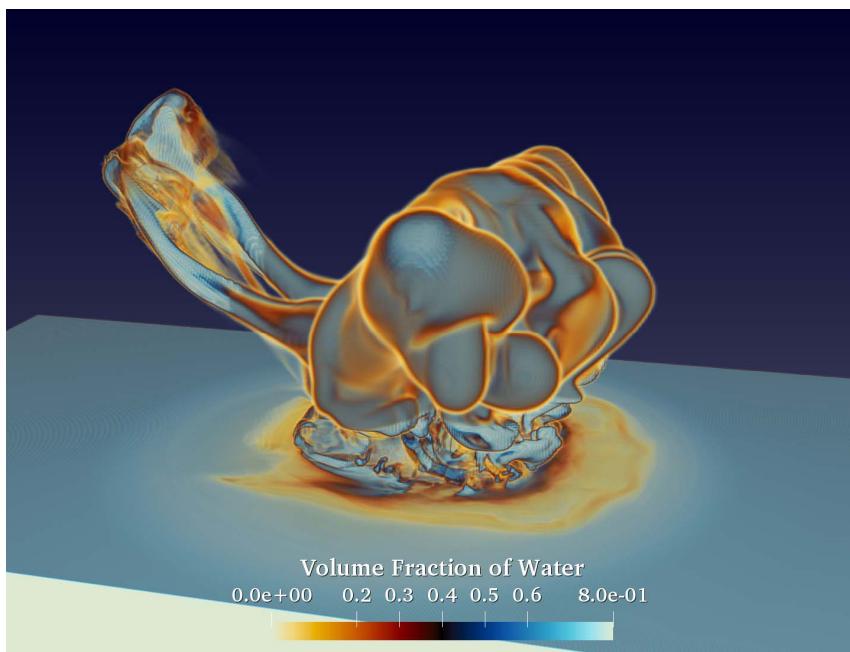
The time spent, the power consumed,<sup>2</sup> and the footprint and general capability of libraries<sup>3</sup> producing in situ visualization and data analysis products are important. Those visualization and data analysis products cannot be evaluated without the corresponding usefulness of the products. That usefulness needs to be measured in the time required for domain scientists to reach insight. Once domain scientists start attempting to understand the data that was stored during the simulation, they should be able to work quickly and efficiently to understand what data they have and what story it exposes. The visualization community must consider the effectiveness of the data visualization and analysis techniques provided on the ability of a domain scientist to quickly and easily assess the products for scientific insight.

End-user time must be a focus area when designing in situ solutions. As a community of visualization and data analysis experts, we should focus on ensuring that domain scientists' time is

spent studying their domain science. The more time they spend managing their data, the less successful we are.

## THE DEEP-WATER IMPACT

Scientific workflows observed while working with Gisler and LANL's Asteroids Team are used as examples for this article. The team's use of simulation to study asteroid-generated tsunamis (AGTs) provided opportunity for the development of these ideas. Visualization and data analysis examples of this AGT work can be found in several LANL publications<sup>4,5</sup> and in Figure 1. Specifically, we will reference the yA31 ensemble member.



**Figure 1.** An image from the Deep Water Impact Ensemble Data Set showing the volume fraction of water with a highly customized color map to show structure as the water moves vertically into the upper levels of the atmosphere. Water is a potent greenhouse gas. Visualizations like this are driving asteroid researchers to be more concerned about the atmospheric repercussions of asteroids impacting deep ocean water.

yA31 is a three-dimensional simulation of a 250-meter-diameter basalt asteroid racing through the atmosphere and impacting five-kilometer-deep ocean water at a 45° angle. This hypothetical phenomenon was mapped onto a computational simulation code called xRage,<sup>6</sup> a parallel multiphysics Eulerian hydrodynamics code that is developed and maintained by the ASC program at LANL. xRage uses a continuous adaptive mesh refinement (AMR) technique that allows smaller computational cells in areas of interest and larger, and thus fewer, cells in other areas, which enables more efficient use of the supercomputer. The simulation contains only three materials: water, air, and asteroid. As the asteroid progresses through the atmosphere and impacts the water, the computational grid refines in areas of entropy, increasing the number of cells in the simulation. The simulation used nearly 50,000 discrete simulation time steps, called cycles, to represent 95 seconds of simulated time. In total, the yA31 simulation consumed 761,000 CPU hours. The simulation included 476 visualization “dumps” that were written to a shared parallel file system. A representative sample of the entire simulation is available at <http://dssdata.org>.

Checkpoint or restart files are written regularly. They allow a simulation job to stop and then restart from that stopping point. This is necessary because supercomputing queues impose a maximum time allowed for any single computational job. In the case of yA31, 51 separate jobs were

run over the course of three weeks, each picking up where the other left off. The activity is performed so that the simulation can be restarted from a specific state should it fail<sup>7</sup> or to support the understanding of the simulation.<sup>8,9</sup>

Gisler frequently checked on the progress of the simulations. He made adjustments where necessary that were applied at restarts. Commonly, the total cell count in the AMR grid would grow beyond the available supercomputing resources. Areas or zones of the simulation would be forced into coarser resolutions to manage the total cell count. At one point during the yA31 simulation run, the number of processing cores doubled from 512 to 1,024. This was done to allow for more total cells and to make the simulation advance more quickly.

## THE COMPUTATIONAL-SCIENCE PROCESS

We want to emphasize the importance of optimizing a scientist's interactive-analysis time. Domain scientists want to spend more time studying their data and less time trying to get access to it. Supercomputers can take hours or days to schedule a job. Long-running simulations are broken into discrete jobs that run as long as the scheduling policy allows, usually 8 to 24 hours each, and might produce a couple of high-resolution data products during each job. Simulations of this nature do not require an *in situ* process to be fast because the process may only get to run a few times per day. The time to produce the data product may not even be evaluated as long as the data product is useful and keeps the domain scientist focused on his or her science rather than data management.

The domain scientist will evaluate time spent in analysis versus simulation time when he or she sits down to look at the results of the simulation days, weeks, or months after it was originally scheduled. It might be to check that the simulation is progressing or to study the outputs to develop conclusions about the simulation. *In situ* activities that stop the simulation from advancing at reasonable rates; that force the scientist to deal with restarts, file systems, and codes rather than his or her domain science; or that provide insufficient or broken data products are a waste of domain scientist time.

## FLEXIBILITY VS. ACCESSIBILITY

Patchett et al.<sup>11</sup> introduced the idea of flexibility and accessibility in evaluating data products produced during *in situ* processing. One end of the spectrum, a checkpoint or restart contains all of the simulation states sufficient to restart the simulation. These types of data products are very flexible and enable new derived analysis products based on them. They are said to be *explorable*.<sup>1</sup> Typically, they are not easily accessible and require specialized tools and substantial computing resources to extract new data products. The other extreme is a single scalar value. For instance, our AGT example contained interesting scalars such as the total water in the stratosphere and total energy transferred from one element to another. When the saved data is only one scalar value representing the whole simulation state, there isn't more to be gleaned by relooking at that scalar value. The simplicity of a single scalar value makes it easily accessible but not very flexible.

The most flexible data products produced during *in situ* processing typically consume more storage space and take longer to access than the less flexible data products. This puts the most flexible output in contention with the desire to produce accessible data products. Imagery is very accessible. It can contain a large amount of information, but, like a single scalar value, an image is difficult to mine for more information than what is readily apparent.

In terms of domain scientist time, single scalar values are extremely easy to view. Standard image formats are also very easy to access and view even from remote supercomputers. In the worst case, a scientist would need to copy small image files to a desktop computer. Large data products will likely stay on the supercomputer. They are too big to easily move and require the domain scientist to spend time processing them in some way to produce more accessible data products.

## IN SITU TECHNIQUES

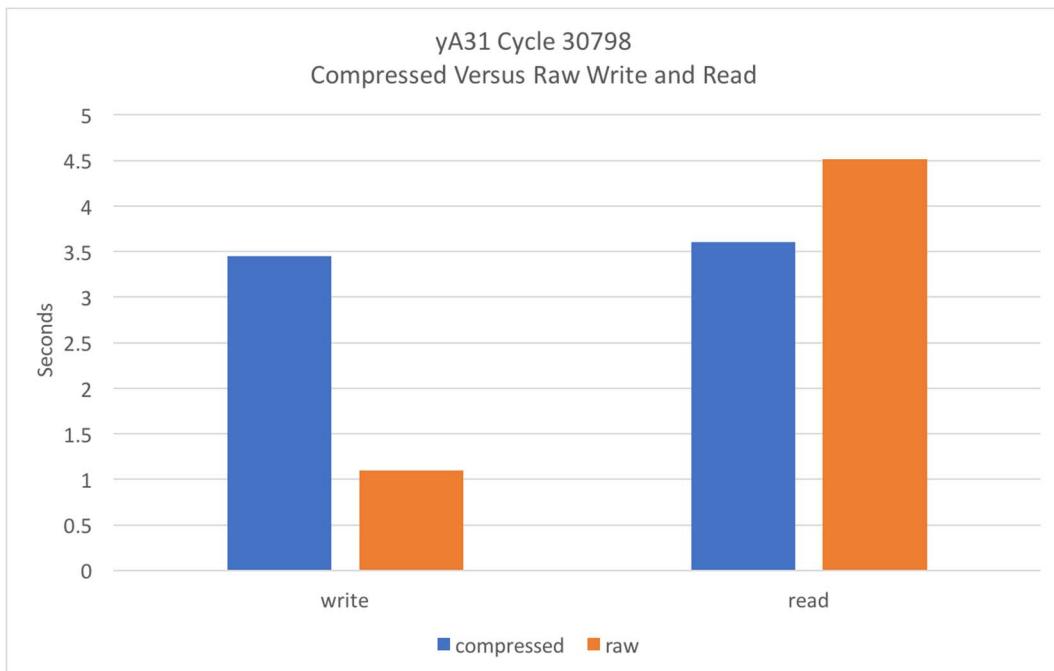
Our work with Gisler shows that several task areas have great promise for in situ processing. These focus around lessening the problems with post hoc data management and analysis. Data size creates issues with resources to process, long-term curation, and exploration. That is accessibility. Lossless compression of data trades compute time for disk space. Lossy compression can not only save disk space but also speed up the entire visualization pipeline. Extraction of salient features is a common visualization activity and is frequently identified as an effective data reduction technique. Using common data formats is an easy and effective way to increase accessibility. Computing and integrating accessible metadata into common file formats makes finding data easier. These are areas that have been suggested previously<sup>12</sup> for further research. Driven by the AGT simulation work, we present four specific examples of our work in support of the production of flexible and accessible in situ data products:

- lossless compression,
- sampling onto a regular grid,
- feature extraction, and
- metadata storage.

### Lossless Compression

Simple compression using known standard compression libraries can potentially save space, and thus time, in the supercomputing environment. Lossless compression is already supported, and transparent to the user, in some tools. The in situ compression of data products has to be balanced with the cost of creating them in supercomputing time,<sup>3</sup> scientist time spent decompressing them for access, the cost of long-term storage, copies and reads to the long-term storage, and the risk of error during the compression operations.

While we would assume using compression would always help, some simple experiments with the native VTK (Visualization Toolkit) support for lossless zlib compression has yielded interesting results. Figure 2 shows parallel read-and-write performance using 512 processes to read and write 512 partitioned files using VTK's XML multiblock native reader and writer, the same writer used by our simulation codes. Compression and write performance is slower than writing raw data to disk. Writing occurs during the simulation runtime. It might appear that compression is a bad idea, because it consumes three times the supercomputing time. Writing compressed data for the 476 saved time steps of the yA31 simulation would have added fewer than 20 minutes during the six weeks of runtime. When reading all of the time steps, nearly 8 minutes of interactive time is saved. This is the time that matters—the time the scientist is performing interactive post hoc analysis.



**Figure 2.** For compressed versus raw data, the time to write to disk *in situ* and then ingest the data into the visualization tool.

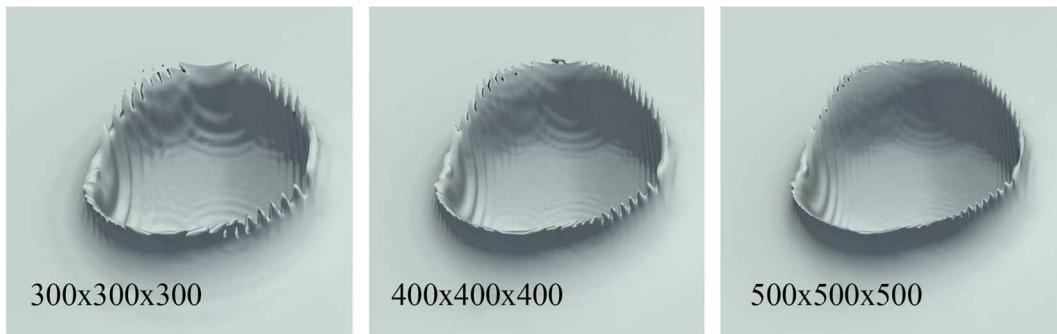
We anecdotally observed that the smaller compressed files had less variability in read time than the larger raw files. We believe this was due to the inherent nature of a shared file system, and a longer read was more likely to get disturbed by other users. More study is needed to validate this.

## Sampling

Sampling is a tradeoff between the quantity of data desired and the quantity of data that is reasonably saved and used. In our AGT simulation work, we note that the decision to store higher-resolution data products consumes more disk space and requires proportionally more time to read it into memory for post hoc analysis. During the yA31 simulation run, 476 total time steps were saved to represent 95 simulated seconds at every 0.2 seconds of simulated time. Higher-resolution sampling must be tempered with the realization that the higher the sampling rate, the longer it will take an end user to load the data and study it.

For the nonsampled raw dataset, the end user waited hours to gain sufficient compute resources, then minutes to load each time step, and then minutes to render each time step. Sampling was the solution to these lengthy interactive wait times.

Figure 3 shows a large-scale feature from a time step in the AGT dataset at different sampled resolutions:  $300^3$ ,  $400^3$ , and  $500^3$ . The  $300^3$  visualized data on the far left is completely sufficient to see the crater feature from the asteroid impact. It is just under 600 Mbytes total, while the  $400^3$  is 1.4 Gbytes and the  $500^3$  is 2.7 Gbytes. The higher-resolution data products are necessary for producing presentation visualizations for marketing and other materials. The feature of scientific interest is visible in the much smaller, more quickly loaded, and more quickly rendered intermediate data product.



**Figure 3. Three sampled resolutions showing the same feature.**

The three different sizes of sampled datasets cover an interesting range of accessibility. A domain scientist can easily and quickly use a commodity tool like ParaView to interactively work through an entire time series of output with the smaller dataset using a desktop computer. A user's desktop computer is always available, unlike a queue-based supercomputer, which might have a lengthy wait time to access the resource. We chose the  $300^3$  sampling size for interactive exploration because the scientist's desktop could accommodate this size and the time series could be copied onto the scientist's desktop over the lunch hour.

We found that by using very low sampling rates, we could quickly and effectively develop visualizations that showcased the large-scale features of interest across many time steps. Those visualization pipelines could then be applied to much larger datasets in noninteractive modes to produce high-quality presentation visualizations.

It is possible that desired features might not be available in the coarsely sampled versions of this dataset. In our AGT example, even the finely sampled versions were insufficient for studying the fine-scale asteroid features. Using domain knowledge and simulation knowledge, those features could be explicitly extracted.

## Feature Extraction

Explicit feature extraction is another mechanism that can be used to lower the size and increase the accessibility of the data. In our AGT example, after studying the tsunami generation potential, the domain scientist was interested in studying the asteroid<sup>14,15</sup> during entry. The spatial-sampling technique was too coarse to capture any details of the asteroid. The sampled data products were not sufficiently flexible to study the asteroid entry. We found that a simple threshold operation could effectively capture the asteroid feature at a fraction of the total dataset size. A regularly sampled version of the whole spatial domain of the dataset could then effectively be used to capture the coarse-level detail, while a full-resolution version of the asteroid itself could also be preserved to provide detail at a completely different spatial scale than the full-dataset sampling described earlier.

This type of reduction improved the accessibility of the dataset such that a modest amount of supercomputing to a large desktop computer could be used to perform interactive post hoc analysis. This is important because the alternative is to use supercomputing resources that require the domain scientist to wait in the queue for access to the compute resources, spend time waiting to load the complete dataset, and then spend time extracting and rendering the feature to produce a visualization like that seen in Figure 4. In this case, the extracted feature was roughly two orders of magnitude smaller than the complete dataset, which requires a similar drop in computing needs and time to visualize the feature of interest.

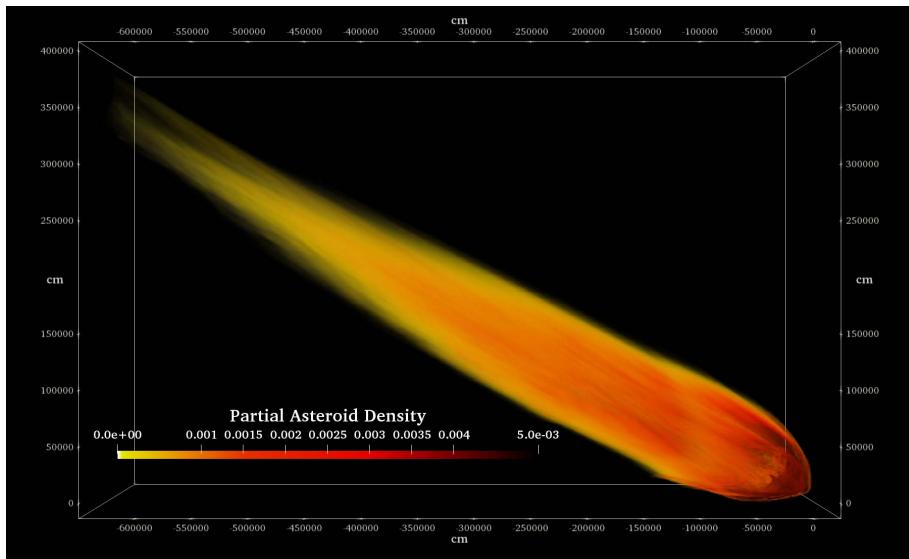


Figure 4. By extracting just the asteroid feature, high-quality imagery of the relatively small asteroid could be produced quickly and with fewer resources than processing the whole dataset. The size of this feature makes it prohibitively costly to get a similar resolution with sampling.

## Metadata

Support for storing and later accessing metadata is essential for a multipurpose in situ visualization and for data analysis tools.

Although much of the metadata could be regenerated based on other more flexible saved data products, metadata can help domain scientists sort through and search for specific data. Recent advances to the VTK XML file type readers and writers have enabled datasets to be augmented with arbitrary metadata such as the version of the simulation code, the name of the input deck, the domain scientist, simulation cycle, and simulated time. In our AGT example this might also include water in the atmosphere, energy transferred from asteroid to air and water, etc. Metadata like this could be leveraged by emerging tools, such as those found in the Cinema project,<sup>17</sup> to allow domain scientists to more quickly and interactively identify visualization and data analysis products of interest to them. Annotations in the visualizations, like those seen in Figure 5, help domain scientists with situational awareness, especially when sorting through quantities of analysis products that are part of ensembles or time series. This metadata will save scientist time, but it must be integrated and accessible.

## CONCLUSION

Gisler's simulation studies support the current consensus that asteroids less than 250 meters in diameter do not pose a substantial tsunami threat.<sup>17</sup> We attempt to increase the accessibility of data analysis to domain scientists such as Gisler by considering their time while managing and studying simulation data. We argue that the time to produce the data products in situ can be offset by the time for the domain scientists to gain access to and preserve those data products. The more time they spend managing their data, the less successful we are.

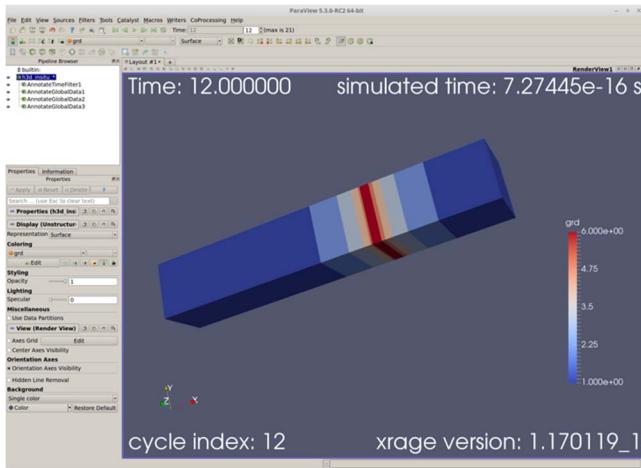


Figure 5. ParaView-based post hoc analysis displaying data produced during in situ processing and showing metadata annotations.

## ACKNOWLEDGMENTS

Many people have contributed to the development of these ideas through discussions or collaborations. We would especially like to thank our LANL team and collaborators, including Boonthanome Nouanesengsy, Pascal Grosset, Quincy Wofford, David Rogers, Jon Woodring, Galen Gisler, the entire Eulerian Application Project, and our ASC and Office of Science program management. We would also like to thank Sandia National Laboratory visualization team members Kenneth Moreland and Patricia Crossno; Eric Brugger of Lawrence Livermore National Laboratory; Hans Hagen from the University of Kaiserslautern; and Berk Geveci, David DeMarle, and Andy Bauer of Kitware Inc.

## REFERENCES

1. A.C. Bauer et al., “In situ methods, infrastructures, and applications on high performance computing platforms,” *Computer Graphics Forum*, vol. 35, no. 3, 2016, pp. 577–597.
2. J. Aherns, “Increasing scientific data insights about exascale class simulations under power and storage constraints,” *IEEE Computer Graphics and Applications*, vol. 35, no. 2, 2015, pp. 8–11.
3. K. Moreland, “The tensions of in situ visualization,” *IEEE Computer Graphics and Applications*, vol. 36, no. 2, 2016, pp. 5–9.
4. J. Patchett et al., “Visualization and Analysis of Threats from Asteroid Ocean Impacts,” *YouTube*, 2017; <https://youtu.be/95z0qRNFFxs>.
5. J. Patchett and G. Gisler, *Deep Water Impact Ensemble Data Set*, technical report LA-UR-17-21595, 2017-05-02 (rev.1), Los Alamos National Laboratory Technical Report, 2017.
6. M. Gittings et al., “The RAGE radiation-hydrodynamic code,” *Computational Science & Discovery*, vol. 1, no. 1, 2018, p. 015005.
7. J. Daly, “A model for predicting the optimum checkpoint interval for restart dumps,” *Int'l Conf. Computational Science (ICCS03)*, 2003, pp. 3–12.
8. B. Whitlock, J. M. Favre, and J. S. Meredith, “Parallel In Situ Coupling of Simulation with a Fully Featured Visualization System,” *Proc. 11th Ann. Eurographics Symp. Parallel Graphics and Visualization (EGPGV 11)*, 2011.
9. U. Ayachit et al., “Paraview catalyst: Enabling in situ data analysis and visualization,” *Proc. First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization*, 2015, pp. 25–29.

10. J. M. Patchett et al., “In Situ and Post Processing Workflows for Asteroid Ablation Studies,” *EuroVis 2017 Eurographics/IEEE VGTC Conf. Visualization*, 2017.
11. K.-L. Ma, “In situ visualization at extreme scale: Challenges and opportunities,” *IEEE Computer Graphics and Applications*, vol. 29, no. 6, 2009, pp. 14–19.
12. P. Lindstrom and M. Isenburg, “Fast and efficient compression of floating-point data,” *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 5, 2006, pp. 1245–1250.
13. F. Samsel et al., “Employing Color Theory to Visualize Volume-Rendered Multivariate Ensembles of Asteroid Impact Simulations,” *Proc. 2017 CHI Conf. Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 1126–1134.
14. J. Patchett et al., *Visualization and Analysis of Threats from Asteroid Ocean Impacts*, technical report LA-UR-16-26258, Los Alamos National Laboratory, 2016.
15. J. Ahrens et al., “An image-based approach to extreme scale in situ visualization and analysis,” *Proc. Int'l Conf. High Performance Computing, Networking, Storage and Analysis*, 2014, pp. 424–434.
16. D. Morrison and E. Venkatapathy, *Asteroid Generated Tsunami Workshop: Summary of NASA/NOAA Workshop*, technical report, 2017; <https://ntrs.nasa.gov/search.jsp?R=20170005214>.

## AUTHOR BIOS

**John Patchett** is the ASC Production Visualization Lead at Los Alamos National Laboratory. Contact him at [patchett@lanl.gov](mailto:patchett@lanl.gov).

**James Ahrens** is a senior scientist at Los Alamos National Laboratory. Contact him at [ahrens@lanl.gov](mailto:ahrens@lanl.gov).

Contact department editor Theresa-Marie Rhyne at [theresamarierhyne@gmail.com](mailto:theresamarierhyne@gmail.com).

**PURPOSE:** The IEEE Computer Society is the world's largest association of computing professionals and is the leading provider of technical information in the field.

**MEMBERSHIP:** Members receive the monthly magazine *Computer*, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.

**COMPUTER SOCIETY WEBSITE:** [www.computer.org](http://www.computer.org)

**OMBUDSMAN:** Direct unresolved complaints to [ombudsman@computer.org](mailto:ombudsman@computer.org).

**CHAPTERS:** Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.

**AVAILABLE INFORMATION:** To check membership status, report an address change, or obtain more information on any of the following, email Customer Service at [help@computer.org](mailto:help@computer.org) or call +1 714 821 8380 (international) or our toll-free number, +1 800 272 6657 (US):

- Membership applications
- Publications catalog
- Draft standards and order forms
- Technical committee list
- Technical committee application
- Chapter start-up procedures
- Student scholarship information
- Volunteer leaders/staff directory
- IEEE senior member grade application (requires 10 years practice and significant performance in five of those 10)

#### PUBLICATIONS AND ACTIVITIES

**Computer:** The flagship publication of the IEEE Computer Society, *Computer*, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.

**Periodicals:** The society publishes 13 magazines, 19 transactions, and one letters. Refer to membership application or request information as noted above.

**Conference Proceedings & Books:** Conference Publishing Services publishes more than 275 titles every year.

**Standards Working Groups:** More than 150 groups produce IEEE standards used throughout the world.

**Technical Committees:** TCs provide professional interaction in more than 30 technical areas and directly influence computer engineering conferences and publications.

**Conferences/Education:** The society holds about 200 conferences each year and sponsors many educational activities, including computing science accreditation.

**Certifications:** The society offers two software developer credentials. For more information, visit [www.computer.org/](http://www.computer.org/) certification.

#### NEXT BOARD MEETING7

7-8 June 2018, Phoenix, AZ, USA

#### EXECUTIVE COMMITTEE

**President:** Hironori Kasahara

**President-Elect:** Cecilia Metra; **Past President:** Jean-Luc Gaudiot; **First VP,**

**Publication:** Gregory T. Byrd; **Second VP, Secretary:** Dennis J. Frailey; **VP,**

**Member & Geographic Activities:** Forrest Shull; **VP, Professional &**

**Educational Activities:** Andy Chen; **VP, Standards Activities:** Jon Rosdahl;

**VP, Technical & Conference Activities:** Hausi Muller; **2018-2019 IEEE**

**Division V Director:** John Walz; **2017-2018 IEEE Division VIII Director:**

Dejan Milojevic; **2018 IEEE Division VIII Director-Elect:** Elizabeth L. Burd

#### BOARD OF GOVERNORS

**Term Expiring 2018:** Ann DeMarle, Sven Dietrich, Fred Douglass, Vladimir Getov, Bruce M. McMillin, Kunio Uchiyama, Stefano Zanero

**Term Expiring 2019:** Saurabh Bagchi, Leila DeFloriani, David S. Ebert, Jill I. Gostin, William Gropp, Sumi Helal, Avi Mendelson

**Term Expiring 2020:** Andy Chen, John D. Johnson, Sy-Yen Kuo, David Lomet, Dimitrios Serpanos, Forrest Shull, Hayato Yamana

#### EXECUTIVE STAFF

**Executive Director:** Angela R. Burgess

**Director, Governance & Associate Executive Director:** Anne Marie Kelly

**Director, Finance & Accounting:** Sunny Hwang

**Director, Information Technology & Services:** Sumit Kacker

**Director, Membership Development:** Eric Berkowitz

**Director, Products & Services:** Evan M. Butterfield

#### COMPUTER SOCIETY OFFICES

**Washington, D.C.:** 2001 L St., Ste. 700, Washington, D.C. 20036-4928

**Phone:** +1 202 371 0101 • **Fax:** +1 202 728 9614

**Email:** [hq.ofc@computer.org](mailto:hq.ofc@computer.org)

**Los Alamitos:** 10662 Los Vaqueros Circle, Los Alamitos, CA 90720 **Phone:** +1 714 821 8380

**Email:** [help@computer.org](mailto:help@computer.org)

#### MEMBERSHIP & PUBLICATION ORDERS

**Phone:** +1 800 272 6657 • **Fax:** +1 714 821 4641 • **Email:** [help@computer.org](mailto:help@computer.org)

**Asia/Pacific:** Watanabe Building, 1-4-2 Minami-Aoyama, Minato-ku, Tokyo 107-0062, Japan

**Phone:** +81 3 3408 3118 • **Fax:** +81 3 3408 3553

**Email:** [tokyo.ofc@computer.org](mailto:tokyo.ofc@computer.org)

#### IEEE BOARD OF DIRECTORS

**President & CEO:** James Jefferies

**President-Elect:** Jose M.F. Moura

**Past President:** Karen Bartleson

**Secretary:** William P. Walsh

**Treasurer:** Joseph V. Lillie

**Director & President, IEEE-USA:** Sandra "Candy" Robinson

**Director & President, Standards Association:** Forrest D. Wright

**Director & VP, Educational Activities:** Witold M. Kinsner

**Director & VP, Membership and Geographic Activities:** Martin Bastiaans

**Director & VP, Publication Services and Products:** Samir M. El-Ghazaly

**Director & VP, Technical Activities:** Susan "Kathy" Land

**Director & Delegate Division V:** John W. Walz

**Director & Delegate Division VIII:** Dejan Milojević

## Call for Papers

COMPSAC is the IEEE Computer Society Signature Conference on Computers, Software and Applications. It is a major international forum for academia, industry, and government to discuss research results and advancements, emerging challenges, and future trends in computer and software technologies and applications. The theme of COMPSAC 2018 is *Staying Smarter in a Smartening World*.

Computer technologies are producing profound changes in society. Emerging developments in areas such as Deep Learning, supported by increasingly powerful and increasingly miniaturized hardware, are beginning to be deployed in architectures, systems, and applications that are redefining the relationships between humans and technology. As this happens, humans are relinquishing their roles as masters of technology to partnerships wherein autonomous, computer-driven devices become our assistants. What are the technologies enabling these changes? How far can these partnerships go? What will be our future as we deploy more and more "things" on the Internet of Things - to create smart cities, smart vehicles, smart hospitals, smart homes, smart clothes, etc.? Will humans simply become IoT devices in these scenarios and if so, what will be the social, cultural, and economic challenges arising from these developments? What are the technical challenges to making this all happen - for example, in terms of technologies such as Big Data, Cloud, Fog, Edge Computing, mobile computing, and pervasive computing in general? What will be the role of the 'user' as the 21st Century moves along?

COMPSAC 2018 is organized as a tightly integrated union of symposia, each of which will focus on technical aspects related to the "smart" theme of the conference. The technical program will include keynote addresses, research papers, industrial case studies, fast abstracts, a doctoral symposium, poster sessions, and workshops and tutorials on emerging and important topics related to the conference theme. A highlight of the conference will be plenary and specialized panels that will address the technical challenges facing technologists who are developing and deploying these smart systems and applications. Panels will also address cultural and societal challenges for a society whose members must continue to learn to live, work, and play in the environments the technologies produce. Authors are invited to submit original, unpublished research work, as well as industrial practice reports. Simultaneous submission to other publication venues is not permitted. All submissions must adhere to IEEE Publishing Policies, and all will be vetted through the IEEE CrossCheck Portal. Papers relevant to COMPSAC topics published in IEEE journals in the last three years or accepted for publication can as well be proposed for live presentation. Further info is available at [www.compsac.org](http://www.compsac.org).

**Standing Committee Chair:** Sorel Reisman, California State University, USA

**Steering Committee Chair:** Sheikh Iqbal Ahamed, Marquette University, USA

**General Chairs:** Shinichi Honiden (NII, Japan)

Roger U. Fujii, Fujii Systems, 2016 IEEE Computer Society President

**Program Chairs in Chief:**

Jiannong Cao (Hong Kong Polytechnic University, Hong Kong)

Stelvio Cimato (University of Milan, Italy)

Yasuo Okabe (Kyoto University, Japan)

Sahra Sedigharvestani (Missouri University of Science & Technology, USA)

**Workshop Chairs:** Kenichi Yoshida (University of Tsukuba, Japan)

Ji-Jiang Yang (Tsinghua University, China)

Hong Va Leong (Hong Kong Polytechnic University, Hong Kong)

Chung Horng Lung (Carleton University, Canada)

**Local Chairs:** Hironori Washizaki (Waseda University, Japan)

Nobukazu Yoshioka, NII, Japan

### Important Dates

#### Main Conference papers

Due date: 31 January 2018 (extended)

Notification: 31 March 2018

#### Workshop papers

Due date: 10 April 2018

Notification: 1 May 2018

#### Camera Ready and Registration

Due date: May 15, 2018



Now there's  
even more to  
love about your  
membership...



Read all your IEEE Computer Society  
magazines and journals **your WAY** on

# myCS



- ON YOUR COMPUTER
- ON YOUR SMARTPHONE

- ON YOUR eREADER
- ON YOUR TABLET

### Introducing myCS, the digital magazine portal from IEEE Computer Society.

Finally...go beyond static, hard-to-read PDFs. Our go-to portal makes it easy to access and customize your favorite technical publications like *Computer*, *IEEE Software*, *IEEE Security & Privacy*, and more. Get started today for state-of-the-art industry news and a fully adaptive experience.



► LEARN MORE AT: [mycs.computer.org](http://mycs.computer.org)