Mixed-Reality for Object-Focused Remote Collaboration

Martin Feick
htw saar
66117 Saarbruecken, Germany
martin.feick@ucalgary.ca

Anthony Tang
University of Calgary
Calgary, AB T2N 1N4, Canada
tonyt@ucalgary.ca

Scott Bateman
University of New Brunswick
Fredericton, NB E3B 5A3, Canada
scottb@unb.ca

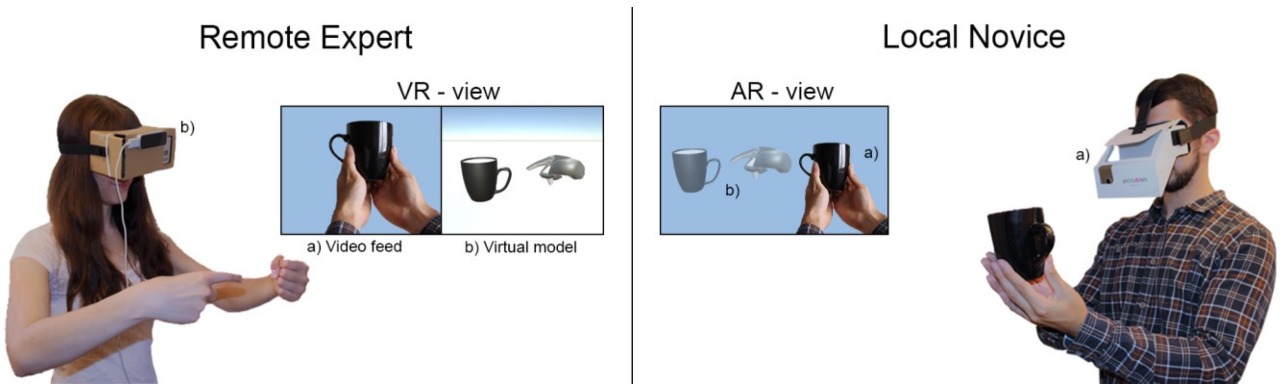

Figure 1: The expert (left) wears a Google Cardboard with a Leap Motion Controller attached, allowing her to interact with a virtual proxy object (b) and to see a shared video-feed (a) from the novice. The novice (right) wears an Aryzon Augmented Reality device. He sees the expert's virtual proxy object and her captured gestures (b).

ABSTRACT
In this paper we outline the design of a mixed-reality system to support object-focused remote collaboration. Here, being able to adjust collaborators' perspectives on the object as well as understand one another's perspective is essential to support effective collaboration over distance. We propose a low-cost mixed-reality system that allows users to: (1) quickly align and understand each other's perspective; (2) explore objects independently from one another, and (3) render gestures in the remote's workspace. In this work, we focus on the expert's role and we introduce an interaction technique allowing users to quickly manipulation 3D virtual objects in space.

Keywords
Mixed Reality; Object-Focused Remote Collaboration; AR; VR

1. INTRODUCTION
Recent work has focused on supporting and addressing the challenges of remote support scenarios (e.g., [7,10,11,14]). In remote support scenarios, a worker (referred to as the local novice) seeks the assistance of an expert located in a different location (the remote expert) [10,11,14,17]. In these scenarios there is often a shared object or workspace that is the focus of the collaboration. A canonical example of this would be a local novice who needs assistance repairing an engine, and who would like to receive guidance from a remote expert. Systems designed to support remote support scenarios often provide video streams of collaborators views and/or shared virtual objects, and has identified three common requirements that arise for remote support systems regardless of the technology that is employed. These three specific requirements have not been well-supported in previous remote support systems:

(1) **Perspective understanding:** the need for collaborators to understand what their collaborators are looking at and the ability to see what their collaborator is seeing [14,18];
(2) **View independence:** the ability to explore objects independently from the collaborator [1,18], and
(3) **Deixis support:** the ability to support some form of deictic reference that allow collaborators to quickly reference particular parts of a shared object or workspace [12].

In this work we propose a low-cost mixed-reality system that meets the three requirements above. Particularly, we focus on the long-studied remote expert scenarios, where a remote expert assists a local novice in performing a physical task focused on a shared object. Below, we summarize how our system better supports the three requirements when compared to video mediated technology and our preliminary work in evaluating our system.

2. Requirements for Remote Support Systems
Even though the three requirements listed above have been investigated in the past, the majority of current systems are video-mediated systems similar to Skype or Google Hangouts. Video mediated systems are insufficient, because they can present the following challenges.

Perspective Understanding: Video-mediated systems only present the object from one perspective in a flattened image [8]. Thus, the depth information is missing, and it is challen-

UIST '18 Adjunct, October 14–17, 2018, Berlin, Germany

ACM ISBN 978-1-4503-5949-8/18/10. https://doi.org/10.1145/3266037.3266102

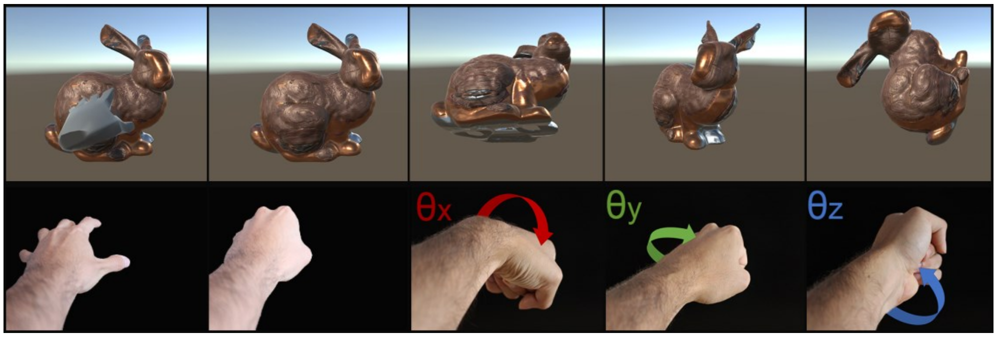

Figure 2: Proprioceptive interaction technique for manipulating virtual proxies (left to right): The user's hand enters the bunny's mesh. By creating a fist, the bunny is selected. Rotations of the fist immediately rotate the bunny.

ging to tell the object's orientation because adjusting the perspective on the object is impossible, which can make the requirement for perspective understanding more challenging. In collaborative mixed reality, we can overcome this by navigating to a collaborator's view in the virtual space. In object-focused remote collaboration this means the user needs to adjust their own view of the object.

View Independence: In video-mediated systems, the expert relies on what the novice shows in the video feed. Generally, there are many issues regarding the camera quality as well as how a person presents the object to the camera (e.g., too close to the camera to see, or out of frame) [8,16,18]. When using a 2D video, the expert can neither explore the object for him/herself nor can they use the object's orientation to clarify and demonstrate instructions. However, these are essential cues in collocated collaboration [15].

Deixis Support: In video-mediated systems, deixis is typically represented in two dimensions through a hand embodiment or telepointer. While 2D gestures have been shown to be helpful for collaboration [13,19], they lack the expressiveness of gestures that are correctly placed with depth [12]. Gestures correctly positioned in 3D space assist in understanding a collaborator's perspective and focus.

3. SYSTEM DESIGN

In Figure 1 we demonstrate our proposed system design. Essentially, it includes two different sites; the Expert Site (ExS) and the Novice Site (NoS).

At the ExS, the expert wears a low-cost Google Cardboard [6] with a Leap Motion Controller [17] attached to it being able to orient the virtual proxy and to track gestures (e.g., pointing); see Fig. 1, left. Experts see a split screen composed of an actual video feed of the novice's workspace, and a virtual environment showing a virtual proxy object (a scanned version of the novice's physical object). The novice can easily create the virtual proxy model by using an application to capture a 3D scan of the physical object (e.g., Qlone [2]).

At the NoS, the novice has the physical object and wears a low-cost Aryzon augmented reality cardboard [5] powered by a smartphone. The smartphone camera also streams 2D video to the ExS providing a shared video feed of the NoS's workspace as proposed in previous work [8,9].

In the end, our system addresses all 3 requirements:

(1) **Perspective understanding:** both expert and novice can see live feeds of the others view of the object;
(2) **View independence:** because a virtual proxy of the physical object is shared, the expert can independently control the virtual object; and
(3) **Deixis support:** the live video allows the novice to gesture at the physical object, the leap motion allows gestures to be made at the virtual object.

In the first iteration, we focus on the ExS. To interact with the virtual proxy objects; i.e., to quickly align perspectives and explore objects, we designed the following free-hand interaction technique.

4. INTERACTION WITH THE PROXY

Our interaction technique is inspired by previous work in the field [3,4,20], and the scene from the science fiction movie Iron Man 2, where Tony Stark's arm represents the orientation of his augmented robot arm. We believe that leveraging proprioception is key to improving interactions with virtual objects.

Proprioceptive: To select the object, experts simply enter the mesh of the object, similar to an interaction shell [20], and create a fist (Figure 2). Thus, the user's hand becomes the virtual object, and changes in hand orientation get immediately displayed on the virtual object (Figure 2). While the fist remains closed, the object stays selected. Hence, the user can manipulate the object around its three axes by bending and/or rotating his/her fist (Figure 2). Rapid clutching (deselecting, readjusting the hand and reselection) allow objects to be naturally rotated around all 3 axes. When an object is selected the virtual hand becomes invisible enabling the user to perform deictic and/or metaphoric gestures.

5. FUTURE WORK

In the future, we want to understand how people would make use of our system and its interaction techniques in simple 3D object docking tasks and collaborations. For example, we believe that our approach of sharing live models of hand gestures will facilitate communications as compared to current systems, which most often share telepointers or annotations (e.g., [7,13]). We are also interested in what the trade-offs are between the currently used interaction techniques for 3D object manipulations (e.g., CAD) and our proprioceptive technique. We believe proprioceptive interactions will be faster to learn and faster when manipulating virtual objects than the interactions used by other 3D object manipulation systems.

However, there are many more questions we need to study in the future. How can people dis/engage with the system? Does the continuous link between the two sites provide enough fidelity? How do we present the VR/AR – views; side-by-side split or a picture-in-picture approach? Eventually, we aim to provide a system that allows richer object-focused remote collaboration which is accessible and affordable for everyone.

6. REFERENCES

1. Matt Adcock, Stuart Anderson, and Bruce Thomas. 2013. RemoteFusion: real time depth camera fusion for remote collaboration on physical tasks. In Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry (VRCAI '13). ACM, New York, NY, USA, 235-242. DOI: http://dx.doi.org/10.1145/2534329.2534331
2. Qlone Application. 2018. Qlone, the all in one tool for 3D Scanning. *URl: https://www.qlone.pro/*
3. Mark Billinghurst, Piumsomboon Tham, and Bai Huidong. (2014). Hands in space: Gesture interaction with augmented-reality interfaces. *IEEE computer graphics and applications*, *34*(1), 77-80.
4. Joan De Boeck, Erwin Cuppens, Tom De Weyer, Chris Raymaekers, and Karin Coninx. 2004. Multisensory interaction metaphors with haptics and proprioception in virtual environments. In Proceedings of the third Nordic conference on Human-computer interaction (NordiCHI '04). ACM, New York, NY, USA, 189-197. DOI: https://doi.org/10.1145/1028014.1028043
5. Aryzon Cardboard. 2018. Aryzon: Augumented Reality powered by your smartphone. *URl: https://www.aryzon.com/*
6. Google Cardboard. 2018. Google Cardboard – Google VR. *URl: https://vr.google.com/cardboard/*
7. Omid Fakourfar, Kevin Ta, Richard Tang, Scott Bateman, and Anthony Tang. 2016. Stabilized Annotations for Mobile Remote Assistance. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 1548-1560. DOI: https://doi.org/10.1145/2858036.2858171
8. Martin Feick, Terrance Mok, Anthony Tang, Lora Oehlberg, and Ehud Sharlin. 2018. Perspective on and Re-orientation of Physical Proxies in Object-Focused Remote Collaboration. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Paper 281, 13 pages. DOI: https://doi.org/10.1145/3173574.3173855
9. Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Hum.-Comput. Interact.* 19, 3 (September 2004), 273-309. DOI= http://dx.doi.org/10.1207/s15327051hci1903_3
10. Lei Gao, Bai, H., Piumsomboon, T., Lee, G., Lindeman, R. W., & Billinghurst, M. (2017). Real-time Visual Representations for Mixed Reality Remote Collaboration.
11. Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14). ACM, New York, NY, USA, 449-459. DOI: https://doi.org/10.1145/2642918.2647372
12. Aaron M. Genest, Carl Gutwin, Anthony Tang, Michael Kalyn, and Zenja Ivkovic. 2013. KinectArms: a toolkit for capturing and displaying arm embodiments in distributed tabletop groupware. In Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13). ACM, New York, NY, USA, 157-166. DOI: https://doi.org/10.1145/2441776.2441796
13. Saul Greenberg, Carl Gutwin, & Mark Roseman. (1996, November). Semantic telepointers for groupware. In *Computer-Human Interaction, 1996. Proceedings., Sixth Australian Conference on* (pp. 54-61). IEEE.
14. Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 957-966. DOI: https://doi.org/10.1145/2702123.2702345
15. Russell Kruger, Sheelagh Carpendale, Stacey D. Scott, and Saul Greenberg. 2003. How people use orientation on tables: comprehension, coordination and communication. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work* (GROUP '03). ACM, New York, NY, USA, 369-378. DOI= http://dx.doi.org/10.1145/958160.958219
16. Terrance Mok and Lora Oehlberg. 2017. Critiquing Physical Prototypes for a Remote Audience. In Proceedings of the 2017. Conference on Designing Interactive Systems (DIS' 17) ACM, New York, NY, USA, 1295-1307. DOI: https://doi.org/10.1145/3064663.3064722
17. Leap Motion. 2018. Leap motion controller. *URl: https://www.leapmotion.com*
   Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15). ACM, New York, NY, USA, 405-415. DOI: https://doi.org/10.1145/2807442.2807497
18. Anthony Tang, Fakourfar, O., Neustaedter, C., & Bateman, S. (2017). *Collaboration in 360 Videochat: Challenges and Opportunities*. University of Calgary.
19. Anthony Tang, Carman Neustaedter, and Saul Greenberg. "Videoarms: embodiments for mixed presence groupware." *People and Computers XX—Engage*. Springer, London, 2007. 85-102.
20. Piumsomboon Tham, Altimira, D., Kim, H., Clark, A., Lee, G., & Billinghurst, M. (2014, September). Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on* (pp. 73-82)