Exploring Enhancements for Remote Mixed Reality Collaboration

Thammathip Piumsomboon
University of South Australia
Adelaide, South Australia, Australia
thammathip.piumsomboon@unisa.edu.au

Arindam Day
University of South Australia
Adelaide, South Australia, Australia
arindam.dey@unisa.edu.au

Barrett Ens
University of South Australia
Adelaide, South Australia, Australia
ensbm001@mymail.unisa.edu.au

Youngho Lee
Mokpo National University
Jeonnam, South Korea
youngho@ce.mokpo.ac.kr

Gun Lee
University of South Australia
Adelaide, South Australia, Australia
gun.lee@unisa.edu.au

Mark Billinghurst
University of South Australia
Adelaide, South Australia, Australia
mark.billinghurst@unisa.edu.au

ABSTRACT

In this paper, we explore techniques for enhancing remote Mixed Reality (MR) collaboration in terms of communication and interaction. We created CoVAR, a MR system for remote collaboration between an Augmented Reality (AR) and Augmented Virtuality (AV) users. Awareness cues and AV-Snap-to-AR interface were proposed for enhancing communication. Collaborative natural interaction, and AV-User-Body-Scaling were implemented for enhancing interaction. We conducted an exploratory study examining the awareness cues and the collaborative gaze, and the results showed the benefits of the proposed techniques for enhancing communication and interaction.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**;

KEYWORDS

Mixed reality, remote collaboration

ACM Reference Format:
Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring Enhancements for Remote Mixed Reality Collaboration. In *Proceedings of SA'17 Symposium on Mobile Graphics & Interactive Applications.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3132787.3139200

1 INTRODUCTION

Mixed Reality (MR) as introduced by Milgram and Kishino [Milgram and Kishino 1994], blends real and virtual worlds along the reality-virtuality continuum comprising of Augmented Reality (AR) and Augmented Virtuality (AV) technology, as shown in Figure 1. AR overlays virtual objects into the real world, while AV captures real objects and superimposes them into the virtual environment. In this research, we explore enhancing remote collaboration between AR and AV users.

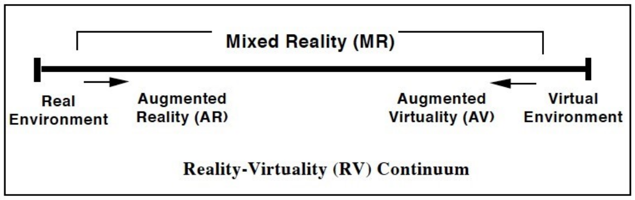

Figure 1: Milgram and Kishino's Mixed Reality on the Reality-Virtuality Continuum .[Milgram and Kishino 1994]

One of the main goals of remote collaboration systems is to enable people who are far apart to feel like they are in the same space. MR technology provides unique capabilities to achieve this goal. Firstly, MR can provide an immersive experience in either real or virtual environments. Secondly, being in a 3D environment enables a person to use their natural ability for spatial interaction. Majority of collaborative AR and Virtual Reality (VR) systems focused on collaboration between users in either only AR or VR situations. There were several prior researches that demonstrated such systems [Billinghurst et al. 2001; Chenechal et al. 2016; Grasset et al. 2005; Kiyokawa et al. 1999; Oda et al. 2015]. These systems used different viewpoints in AR or VR to support different collaborative roles, often a remote expert supervising another user who is performing a real-world or virtual task. Furthermore, they often shared virtual 3D content or captured real environment in a desktop scale workspace. In contrast, our system offers MR technology and supports a room size 3D reconstruction, which an AR user can share with a remote collaborator in AV. In this way, the AR and AV users can experience a shared space and collaborate on real-world tasks. There are many possible applications of this type of system such as emergency response, remote maintenance, education and others.

In this research, we propose techniques for enhancing communication and interaction in remote MR collaboration. In face to face collaboration, head pose and eye gaze are important communication cues for sharing focus of attention. Previous systems showed the importance of awareness cues, such as virtual pointers [Duval et al. 2014; Greenberg et al. 1996; Oda et al. 2015] or embodied

---

SA'17 Symposium on Mobile Graphics & Interactive Applications, November 27-30, 2017, Bangkok, Thailand

ACM ISBN 978-1-4503-5410-3/17/11.
https://doi.org/10.1145/3132787.3139200

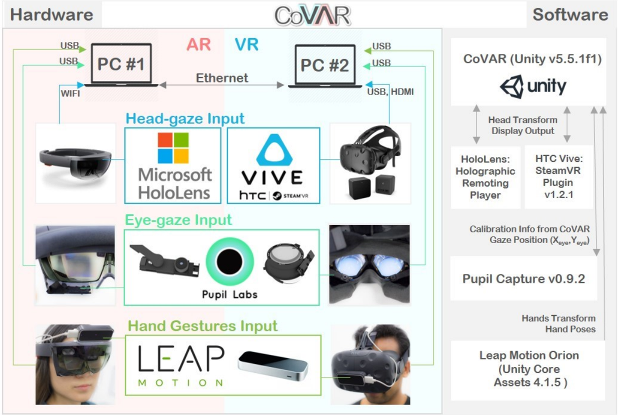

Figure 2: CoVAR system overview.

hand gestures[Sodhi et al. 2013], to support effective communication. We provide remote embodiment using an avatar's head and hands to represent users and further enhance the collaboration with awareness cues and AV-Snap-to-AR interface. For interaction enhancement, we demonstrate collaborative interaction and AV-User-Body-Scaling. We also report on an exploratory study to learn the effects of awareness cues and collaborative natural interaction in remote collaboration.

In this paper, we present CoVAR (Collaborative Virtual and Augmented Reality) system, a novel room-scale mixed-platform remote collaboration system. Our work combines and extends earlier research in MR collaboration, remote embodiment in collaborative systems, and collaboration enhancement.

2 SYSTEM OVERVIEW

CoVAR is a multi-user system with server-client architecture. Either the AR or AV user can be the host server to support different use case scenarios. Both AR and AV sides use Pupil Labs system for eye tracking and Leap Motion system for hand pose recognition. For the AR side, we use Microsoft HoloLens, which is connected to a computer through Holographic Remoting Player via WIFI. On the AV side, HTC Vive with SteamVR platform runs on a second computer. The two computers are connected through Ethernet with TCP/IP connection. CoVAR was developed using Unity 5.5.1f1. The system overview is illustrated in Figure 2.

2.1 System Setup

Both AR and AV user's spaces had the same dimension of 3.5 x 3.5 meters. Only the AR user's space had furniture while the AV user's space was kept empty. The equipment used in each side of the setup is shown in Table 1. Procedures to setup and calibrate each space to align them are described in Table 2. The alignment error depends on the reconstruction quality. When the environment was well reconstructed and aligned, we had the alignment error at around 1-4 centimeters.

Table 1: Hardware used in AR and AV setup

AR side	AV Side
(1) a Windows 10 laptop - Intel Core i7 at 2.7 GHz, 32 GB RAM, and NVIDIA GeForce GTX 780M,	(1) a Windows 10 laptop -Intel Core i7-6700HQ at 2.6 GHz, 16 GB RAM, and NVIDIA GeForce GTX 1070,
(2) Microsoft HoloLens,	(2) HTC Vive Kit,
(3) a Pupil Labs eye tracker,	(3) a pair of Pupil Labs eye trackers with a binocular mount for the HTC Vive
(4) a Leap Motion sensor and a custom-made mount unit	(4) a Leap Motion sensor and VR mount unit

Table 2: AR/AV spaces alignment procedures

Step	User	Procedure
1	AV	Setup and calibrate the HTC Vive Lighthouse system for a room-scale tracking (in our case the interaction space is 3.5x3.5 meters).
2	AR	Place the HoloLens at a position that is to be the origin of the AR space (Figure 3b), aligning the front camera to the origin for all axes. This is the default calibration position.
3	AR	Launch and reconstruct the surrounding using the HoloLens Image-based Texturing software [hol [n. d.]].
4	AR	Share the reconstructed mesh to the VR user. The current implementation requires the model to be downloaded from the HoloLens and shared onto the VR server machine manually but we are developing live reconstruction and sharing.
5	AV	Load the mesh into the scene for VR user (Figure 3c).
6	AR	When the AR user begins CoVAR application, the HoloLens must be placed at the original calibration position.

2.2 System Inputs

To create a seamless collaborative experience, CoVAR provides three common input methods across our MR platform: head-gaze, eye-gaze, and hand gestures.

2.2.1 *Head-gaze Input.* Head-gaze is an input method based on user's head movement. This data is provided by the Head Mounted Display (HMD)'s tracking data. On the HoloLens, the localization is provided by the spatial mapping technology, and on the Vive, by the Lighthouse system through SteamVR platform. The head-gaze location is calculated as the point of intersection between a ray cast from the center of user's Field-of-View (FoV) perpendicular to the view plane and the first object it hits.

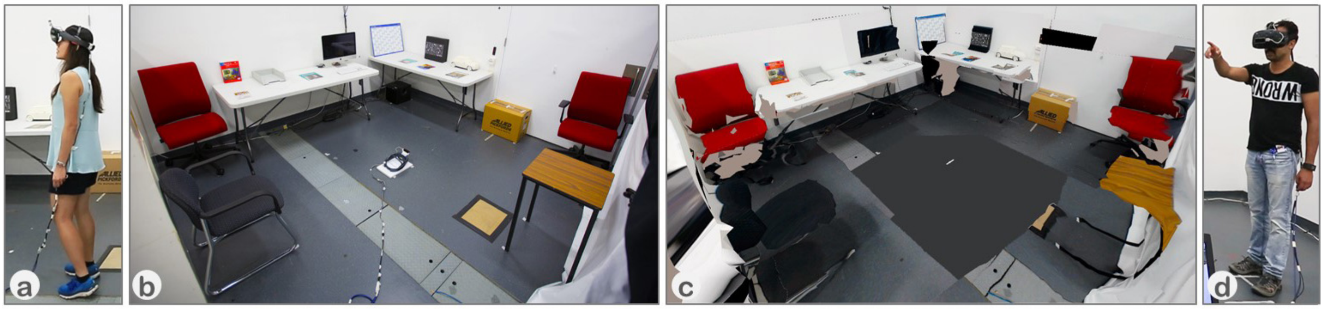

Figure 3: CoVAR system setup: a) AR user, b) AR user's real environment, c) AV user's reconstructed environment, d) AV user.

2.2.2 Eye-gaze Input. To track user's eye-gaze, we used the Pupil Labs eye tracking camera and Pupil Labs Capture software for calibration and tracking. The eye-gaze location is calculated as the point of intersection between a ray cast from the head's center position in the direction of the projected eye-gaze point and the first object it hits.

2.2.3 Hand Gestures Input. HoloLens supports only a small set of free-hand gestures as inputs, while Vive relies on the controllers. To provide a common input method that encourages natural interaction across different platforms, we used Leap Motion for hand tracking and gesture recognition. Leap Motion sensors were mounted on top of the HoloLens and in front of the Vive as shown in Figure 2. CoVAR supports various gesture interaction, e.g. pointing with ray, object grasping, sweeping here/away.

3 COLLABORATION ENHANCEMENTS

3.1 Enhancing Communication

For remote MR collaborations, it is necessary to provide basic visual cues for shared understanding and communication. Our system provides remote embodiment cues that comprise of the avatar's head and hands, representing the remote user's head position, face direction, and hand poses. Although, the local users can see their own hands animated with full degree-of-freedom (dof) of control, in order to save the amount of data exchanged between the users, the remote collaborator's hands are represented with one of the four possible pre-defined hand poses. When one of the predefined poses is recognized, the hand is highlighted in different colors to indicate that the pose is visible to the remote user. The colors for the poses are neutral pose in grey, pointing in blue (See Figure 4a), grasping in red, and thumbs up in green.

3.1.1 Awareness Cues. We use two types of virtual cues to support collaboration awareness. The first is a virtual Field of View cue, FoV, showing the boundary of what the users can see through their display (See Figure 4b). This helps informing a user of what their partner can see. The second cue is the Gaze cue, represented by a ray showing the user's head-gaze or eye-gaze direction (See Figure 4b). This is helpful for pin-pointing the exact gaze location. The FoV cue is represented by a frustum displaying the view volume of the remote collaborator. Different display technologies support different sizes of FoV. Figure 4b illustrates the smaller AR frustum in pink, which matches the size of the HoloLens's FoV. The blue AV frustum has been reduced to half the size of the actual FoV of the HTC Vive's FoV, as we found from pilot tests that it is difficult to recognize when it is in the actual size. As for the gaze cue, it is shown as a ray representing the user's gaze direction. CoVAR supports two types of gaze, head-gaze, in which the ray is perpendicular to the view plane, and eye-gaze, in which the ray is in the direction of the projected eye-gaze point on the view plane.

3.1.2 AV-Snap-to-AR. To assist AV user in better understanding the AR user who is working in the real-world task, CoVAR provides a AV-snap-to-AR point of view interface. When this interface is active, the AV user's head position moves together with the AR user, yet the AV user still has an independent head orientation and minor control of his head position offset from the AR user to avoid simulation sickness. There are two modes in this interface: 1) snapping to AR user's position only and 2) re-orient AV user's orientation to face AR user's facing direction as well. This interface allows AV user to instantly acquire AR user's perspective without having to move or manually teleport to the AR user's location. AV user can return to his/her original position by disabling this interface.

3.2 Enhancing Interaction

3.2.1 Enhanced Interaction. Our system supports three primary input methods; Eye-Gaze, Head-Gaze, and Hand Gestures. These also serve as collaboration cues in our system. Distinct from other collaborative AR/VR systems, CoVAR supports eye-gazed-based interaction. We have developed two eye-gaze interaction techniques. The first is a combination of gaze and gesture manipulation. Users can use an eye-gaze to select an object of interest and then perform a hand gesture to execute an action on the selected object. For example, while looking at a distant object, the user can perform a "Come-Here" gesture, beckoning with their hand, to bring it closer. Users can also use direct free hand manipulation to move objects within their arm's reach. The second technique is a collaborative gaze which requires both the AR and AV users to gaze at the same target object to trigger an action such as revealing hidden information (Figure 4c).

3.2.2 Virtual Body Scaling. Sharing the 3D reconstruction of the AR user's space with the AV user provides a virtual collocation experience. Following this setup, we explore different techniques for

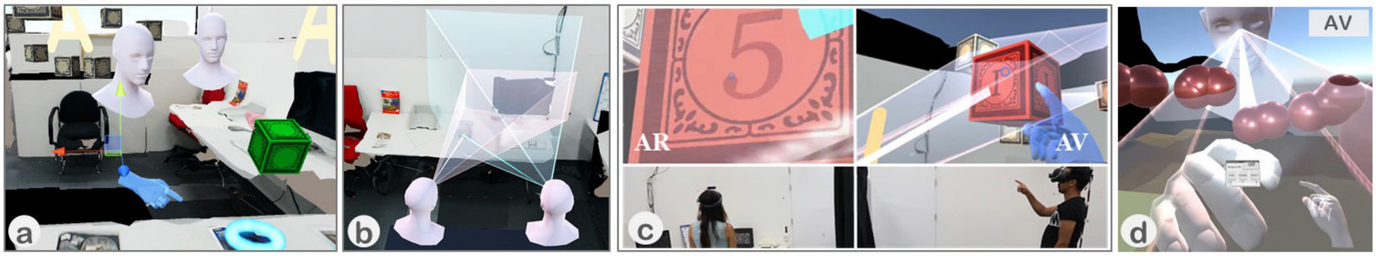

Figure 4: Collaboration enhancements: a) 3rd person's view in virtual collaboration, b) awareness cues (FoV + eye gaze), c) collaborative gaze, d) AV user's point of view in Miniature mode standing on a coffee table (smaller than the AR user).

changing the user's view to enhance the collaborative experience. Our technique enables the AV user to scale bigger or smaller relative to the real world. In the Normal scale mode, we have a 1 to 1 scale between AR and AV users. In the Miniature mode, the AV user is scaled down to a fraction in size compared to the AR user (see Figure 4d). In the God mode, the VR user becomes bigger than the AR user and their room. Other different types of view manipulation could be implemented in the future.

4 EXPLORATORY STUDY

We conducted an exploratory study with 32 participants in 16 pairs (9 were females). We investigated two collaboration enhancements in awareness cues for communication enhancement and collaborative gaze for interaction enhancement using the head-gaze. The two conditions were remote embodiment only (**RO**) and remote embodiment with awareness cues (**RA**). In the first condition, users could see only avatar's head and hands during their collaboration. In the second condition, in addition to avatar's head and hands, users could also see their collaborator's FoV and head-gaze as shown in Figure 4c. For collaborative interaction, collaborative head gaze was used in both conditions to reveal hidden information. This is a within-subject study and the conditions were counterbalanced for every pairs to prevent the learning effect.

4.1 Tasks

The study used a collaborative object finding and placing task, called "Gaze and Place", where users had to search for tagged virtual blocks in an AR or AV interface. They had to find a block tagged with a certain combination of a number and a letter by looking at the same block to reveal the hidden tag information. When two users looked at the same block, it showed a number to the AR user, and a letter to the AV user. Once they found the correct block, one user had to move it to a target place which is only shown to the other user.

4.2 Results

We found significant differences between **RO** and **RA** in mutual gaze rate and distance traveled by users as objective data, and in usability as subjective data.

4.2.1 *Objective data:*

For mutual gaze rate, we counted how many times collaborators looked at the same block during the identification task, which enabled them to identify whether it was the correct block. We found the mutual gaze rate of **RA** being significant higher than that of **RO**, t(15) = -3.9, p = .001. For distance traveled by users, we calculated the total movement (in meters) of participants in the environment as an indication of the physical load. A sum of distance traveled by both collaborators together yielded a significant difference, t(15) = -3.2, p = .006, where **RA** had lower total movement than **RO**.

4.2.2 *Subjective data:*

We asked four questions related to usability. (1) How easy was it to use the cue? (2) How useful was the cue for collaboration? (3) How stressful was it to use the cue? and (4) How confusing was the cue to understand? Participants rated on a 5-point Likert-scale. We found a significant difference for ease of use, t(31) = -3.8, p = .001, usefulness, t(31) = -4.0, p < .001, and confusion, t(31) = 3.5, p = .001, all in favor of **RA** over **RO**. We did not find significant difference between conditions in term of stressfulness.

4.3 Discussion

4.3.1 *Awareness cues:*

We observed that awareness cues were useful for both AR and AV users, allowing them to find their collaborator's gaze area quicker with the FoV cue, and an exact gaze target with the gaze-ray. The AV users were also able to empathize the AR user's limited vision by looking at his/her FoV cue. The awareness cues were crucial in improving performance and usability as indicated by a significantly higher mutual gaze rate, reduction of distant traveled by users, and higher rating in subjective feedback.

4.3.2 *Collaborative interaction:*

Head-gaze input provided both awareness cues in FoV and gaze cue, and was also used as the default interaction method for revealing the character on the block. The design coupling of this communication and interaction cues utilized the implicit nature of natural interaction and collaboration cue, resulting in higher efficiency in collaboration indicated by the significantly better performance and user experience. From this finding, we encourage interaction designers to better align communication and interaction cues for a specific task, and to improve the user experience, when designing collaborative interaction.

5 CONCLUSIONS

In this paper, we presented CoVAR, a novel MR system for remote collaboration. CoVAR enables an AR user to capture and share the local environment with a remote user in AV, and collaborate on spatial tasks in the room-size shared space. We implemented techniques to enhance communication including awareness cues and AV-snap-to-AR interface, and also techniques to enhance interaction with collaborative natural interaction and AV-user-body-scaling. We conducted an exploratory study and found evidences supporting benefits of some of our enhancements. In our future research, we will investigate other cues to enhance the collaboration and test the system with a different real-world task.

REFERENCES

[n. d.]. Hololens_Image_Based_Texture. https://github.com/ywj7931/Hololens_Image_Based_Texture. ([n. d.]). Accessed: 2017-03-15.

Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook: a transitional AR interface. *Computers & Graphics* 25 (2001), 745–753.

Morgan L. Chenechal, Thierry Duval, ValÃĺrie Gouranton, JÃĺrÃťme Royan, and Bruno Arnaldi. 2016. Vishnu: virtual immersive support for HelpiNg users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE).* 9–12. https://doi.org/10.1109/3DCVE.2016.7563559

Thierry Duval, Thi Thuong Huyen Nguyen, Cédric Fleury, Alain Chauffaut, Georges Dumont, and Valérie Gouranton. 2014. Improving Awareness for 3D Virtual Collaboration by Embedding the Features of Users' Physical Environments and by Augmenting Interaction Tools with Cognitive Feedback Cues. *Journal on Multimodal User Interfaces* 8, 2 (June 2014), 187–197. https://doi.org/10.1007/s12193-013-0134-z

Raphael Grasset, Philip Lamb, and Mark Billinghurst. 2005. Evaluation of Mixed-Space Collaboration. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '05).* IEEE Computer Society, Washington, DC, USA, 90–99. https://doi.org/10.1109/ISMAR.2005.30

Saul Greenberg, Carl Gutwin, and Mark Roseman. 1996. Semantic telepointers for groupware. In *Proceedings Sixth Australian Conference on Computer-Human Interaction.* 54–61. https://doi.org/10.1109/OZCHI.1996.559988

Kiyoshi Kiyokawa, Haruo Takemura, and Naokazu Yokoya. 1999. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. In *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*, Vol. 6. 48–53 vol.6. https://doi.org/10.1109/ICSMC.1999.816444

Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems* E77-D, 12 (Dec. 1994), 1321–1329.

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. https://doi.org/10.1145/2807442.2807497

Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. 2013. BeThere: 3D Mobile Collaboration with Spatial Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13).* ACM, New York, NY, USA, 179–188. https://doi.org/10.1145/2470654.2470679