Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction

Theophilus Teo

University of South Australia Adelaide, Australia theophilus_hua_lid.teo@mymail. unisa.edu.au Louise Lawrence University of South Australia Adelaide, Australia louise.lawrence@mymail.unisa.edu. au **Gun A. Lee** University of South Australia Adelaide, Australia gun.lee@unisa.edu.au

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

ABSTRACT

Remote Collaboration using Virtual Reality (VR) and Augmented Reality (AR) has recently become a popular way for people from different places to work together. Local workers can collaborate with remote helpers by sharing 360-degree live video or 3D virtual reconstruction of their surroundings. However, each of these techniques has benefits and drawbacks. In this paper we explore mixing 360 video and 3D reconstruction together for remote collaboration, by preserving benefits of both systems while reducing drawbacks of each. We developed a hybrid prototype and conducted user study to compare benefits and problems of using 360 or 3D alone to clarify the needs for mixing the two, and also to evaluate the prototype system. We found participants performed significantly better on collaborative search tasks in 360 and felt higher social presence, yet 3D also showed potential to complement. Participant feedback collected after trying our hybrid system provided directions for improvement.

CCS CONCEPTS

• Human-centered computing → Mixed / augmented reality; Collaborative interaction; Computer supported cooperative work.

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK © 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-5970-2/19/05...\$15.00

https://doi.org/10.1145/3290605.3300431

Matt Adcock CSIRO Canberra, Australia matt.adcock@csiro.au

KEYWORDS

Mixed Reality, Virtual Reality, Remote collaboration, 360 panorama, 3D scene reconstruction, interaction methods

ACM Reference Format:

Theophilus Teo, Louise Lawrence, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May* 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300431

1 INTRODUCTION

Remote collaboration technology can enable a user at a local work place to quickly receive help from another expert user in a remote location. For example, a video call allows the remote guest user to quickly understand the situation of the local host user. In remote collaboration it is important to consider how the local host user can capture and broadcast a view of their surroundings. However, video calls have many limitations such as sharing a small field of view (FOV), limited resolution, or fixing the view of the remote guest user to that of the local host user. To overcome such limitations, it could be more efficient if the remote guest user could immerse themselves in a view of the local host user's environment. Virtual Reality (VR) technology enables this by having the user wear a Head Mounted Display (HMD) that provides an immersive viewing experience with a wider FOV compared to a standard phone or monitor.

Using VR technology, 360 views of the surroundings can be shared from a panorama camera. Alternatively, other systems allow sharing a 3D reconstruction of a real world scene using a depth sensor and/or photogrammetry. Both of these techniques allow sharing the local host user's environment to a remote guest user, but each has some limitations. Sharing 360 panorama views can provide a high quality view without consuming a large amount of bandwidth but it is a 2D presentation that provides limited depth perception. In

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.



Figure 1: MR remote collaboration system overview.

contrast, sharing a 3D reconstruction supports depth perception as well as the ability to navigate through the 3D model. However, the quality of 3D reconstruction and the amount of bandwidth required to transfer it are directly proportional to each other. So a high-quality 3D reconstruction of the user's environment would require a significant amount of bandwidth and is difficult to update in real time.

In this paper, we present a novel Mixed Reality (MR) remote collaboration system (see Figure 1) that combines 360 and 3D reconstructions into one. This creates a system that aims to merge the advantages of the individual approaches while minimizing the limitations stated above.

Compared to prior work, this paper makes a number of novel and significant contributions:

- (1) A novel MR Remote Collaboration technique that merges 360-views and 3D Reconstruction.
- (2) The first user study that compares 360 live panorama and 3D reconstruction based MR remote collaboration systems.
- (3) The first user study that explores the benefits and implications of combining 360-view and 3D reconstructed scene into a hybrid MR remote collaboration system.

2 RELATED WORK

360 Video Sharing Remote Collaboration

Recently, researchers started to explore how 360 panorama camera can be used instead of a standard camera for remote collaboration. This allows the local host user to capture and broadcast the 360 surrounding view to the remote guest user who could turn their head while wearing the HMD. For example, JackIn Head [16] was a remote collaboration system using 6 cameras constantly capturing videos from different angles and processing them into a 360 high-quality spherical video image to live stream to another user wearing an HMD to view. The system is constructed as a headband that allows easy wearability on the head for the local host user. Tang et al. [37] created a 360 video chat system with a similar set up on the local host user side. In their system, they used a 360 camera on a monopod fixed to a user's backpack to broadcast the 360 surroundings to the viewer watching it on a tablet device. Most recently, the Shared Sphere system [21] used a 360 panorama camera attached to a Microsoft HoloLens [23] to capture and share the user's surroundings. With this system, both local host and remote guest users could look around independently while sharing visual communication cues through MR visualisation.

These systems provide easy access to the 360 surroundings of the local host user by the remote guest user, who can look around independently. However, the viewing position of the remote guest user is strictly controlled by the local host user. So the remote guest user will not be able to look at a certain corner of a room or behind any occluded objects by walking closer, unless the local host user goes there.

Collaboration using 3D Scene Reconstruction

In order to overcome the limitation of a fixed viewing position, 3D scene reconstruction can be used instead. It allows the surroundings of the local host user to be reconstructed either live [1, 7, 8, 14, 15, 27] or beforehand [25, 30], and send the 3D reconstruction to the remote guest user. In recent years, researchers have been evaluating the capabilities of 3D reconstruction for remote collaboration.

Izadi et al. [14] showed a real-time 3D reconstruction method that enables a user to scan and reconstruct their surroundings using an RGB-D Camera. This allowed users to freely move through the reconstructed scene and interact with the virtual content. However, the texture information was omitted, so the visual details for this system was limited to mesh level.

The Holoportation system [27] used multiple depth cameras to scan and reconstruct a person in a space, including the visual texture. This could be viewed through a HoloLens, enabling a remote person to appear as an Augmented Reality (AR) image in the user's real space. This created interaction between remote people that was close to face to face collaboration. However, this system required a large setup, significant network bandwidth and special hardware such as a set of depth cameras at fixed positions. This limits the portability and the accessibility of the system. To overcome this, Dong and Höllerer [8] developed a system using the built-in RGB-D camera on HoloLens to scan and reconstruct surroundings in real time. This allowed for quick and easy set up for 3D reconstruction, however the quality and resolution for the reconstructed scene are significantly lower than Holoportation.

Similarly, BundleFusion [7], created a real-time and highquality 3D scanning technique for large-scale scenes that allows a user to view and access the scene with texture using information obtained from an RGB-D stream. This allows a user's surroundings to be scanned and shared with another user in real time. However, this system mainly focused on one-way communication, lacking support for interaction between users. The local host user was not able to see or communicate with the remote guest, compared to an older work by Adcock et al. [1] used a Spatial AR light annotation mechanism on top of a fused RGB-D scene to achieve remote collaboration between the two sides. In another case, Joachimczak et al. [15] presented a 3D reconstruction system that reconstructed people and objects using Kinect [24] and share with another user using HoloLens, however it did not provide the ability to reconstruct the surroundings as a whole.

Nuernberger et al. [25] developed a technique that reconstructs a scene through a set of photographs. This allowed a user to snap their viewpoint to a particular photo in the scene so that it creates the illusion of being immersed into the scene. Farin et al. [10] and Rhee et al. [32] also presented a similar technique of using either 360 panorama images or video as a background scene or light source in the virtual world. All of these systems used image and photo illusion techniques, to create an immersive 3D experience for the user, but did not have live image updates and so had limited support for collaboration. We can therefore say that for a quick set up, using 360-video in VR is a good option while for a more realistic user experience, 3D reconstruction could be better [4].

Visual Cue in Remote Collaboration

In a remote collaboration system, nonverbal communication through user interaction should be considered as it could affect the user performance even greater than verbal communication. For example, in AR and VR collaboration the use of virtual replicas [9, 26] could enable a user to better understand the relationship between objects and carry out a task quickly. This also applies when AR cues such as hand gestures and gaze indicators are used to provide virtual communication aids [30]. In addition, visual cues can also be used as a communicator for workspace awareness [6] as sometimes a local worker needs to understand the view of a remote expert [2]. For example, it can be important for both users to be aware of each other's viewing direction [13, 21, 28], and this can help both local host and remote guest users to mutually understand the context in a remote collaborative task.

Sharing a life-size avatar could also be useful especially for MR remote collaboration. Researchers [15, 27, 29] have demonstrated the use of AR to visualize a life-size human avatar through sensors scanning. Further enriching the idea, Piumsomboon et al. [31] introduced Mini-Me that combines visual cues with an adaptive human avatar to give more variation on interactive visual cues in a MR remote collaboration.



Figure 2: Prototype system features.

Compared to prior work in MR remote collaboration systems, our work propose combining 360 video and 3D reconstruction to complement each other. In this paper, we describe the system design and implementation of our novel MR remote collaboration system that merges live 360 video and 3D reconstruction, and report on a user study that investigates how merging the two techniques would be useful for enhancing remote collaboration.

3 SYSTEM OVERVIEW

Our prototype system design focuses on binding two elements, 360 Remote Collaboration and 3D Remote Collaboration, into a single piece of prototype system. In this way, the prototype system allows a remote guest user to collaborate with a local host user through both live 360 video and reconstructed 3D scene from local host user's side (see Figure 2).

360 Remote Collaboration (360 mode)

For 360 video based remote collaboration, our prototype system is built based on our prior work [21] which implements MR Remote Collaboration by sharing live 360 panorama video from a local host user's perspective to a remote guest user. A local host is equipped with a Microsoft HoloLens [23] AR HMD mounted with a Theta V [33] 360 panorama camera to share surroundings and view visual cues made by the remote guest users. The remote guest user wears a Samsung Odyssey [34] VR HMD mounted with a Leap Motion [19] hand tracking sensor to view the shared surroundings and share visual cues.

While the collaboration takes place, a coloured rectangle view frame is shown to each user which represents the viewing direction of the other user. On top of that, the remote guest user can also use their hand to point or make gestures in the shared surroundings with the help from the hand tracking sensor. This allows nonverbal communication as an interaction method which enhances the collaborative experience. To provide better user experience, the prototype system also supports 360 video stabilization that automatically stabilizes the video using orientation tracking sensor by counter-rotating the 360 video sphere when the local user turns his or her head.

3D Remote Collaboration (3D mode)

We introduce a 3D mode that shares a 3D scene by using the static 3D model reconstructed from the local host user's surrounding. To achieve this, we use Agisoft PhotoScan [3] software to perform photogrammetric processing of 2D digital camera images and generate 3D spatial data. The use of a static 3D scene reduces the data bandwidth requirement.

To aid 3D perception as well as to increase the connection between the virtual and physical worlds, we use an HMD supporting positional tracking to allow natural walking as the navigation method in the 3D scene. Unlike the 360 mode where both the visual cues and the interaction method are held in a first person view shared between the two user, in 3D mode, both users watch each other from a second person view in the 3D scene.

Visual cues, Annotations and Avatar

Similar to our prior system [21], the remote guest user can use natural hand gestures and visual cues as nonverbal communication methods. This includes a view frame that indicates the viewing direction of a user, animated 3D hand model for making gestures, and a ray pointer for pointing. There are also additional features such as visualisation of an arrow or halo to indicate or highlight the view frame and virtual hands when they are out of the user's FOV.

With the system in two modes, visual cues and annotations appear from different viewing perspective. In 360 mode, both local host and remote guest users shares the same first person view. However in 3D mode, both users have control over their own view position allowing them to move independently in the scene. As this also changes the view mode into a second person view, it can be difficult to understand the other user's view direction by just looking at a view frame floating in the 3D environment.

To give both users a clearer indication of the other user's perspective in 3D mode, we introduce a simple avatar made out of a cube and a few cylinder-shaped objects representing the head (including eyes and nose) and body of the user. The avatar is positioned based on the user's location in the shared 3D space. In addition to the view frame shown in front of the avatar, the avatar's head is also rotated according to the user's head motion to allow better indication of user's viewing direction.

In addition to the view frame, a live video feed cut out from the 360 video is shown as an inset to the view frame in 3D mode. This allows the remote guest user to better understand and be aware of the changes made in the 360 reconstructed scene despite using a static 3D model.

Integration of 360 and 3D Modes

By combining the 3D mode with the 360 mode, we present a prototype system with the capabilities for a user to switch between the two modes (see Figure 2). Figure 1 outlines the system setup that is similar to our previous work [21] while supporting a new feature to switch between 360 and 3D modes. As part of the system setup procedure, the local host needs to construct the 3D model of his or her surroundings and send it to the remote guest in advance. Once the remote guest loads the 3D scene, the system will be ready for connection.

To support the remote guest user to switch between 360 and 3D modes, we use the 3D tracking data from the remote user's HMD to allow real time virtual coordinate mapping between the two modes. This allows the remote user to navigate between 360 and 3D modes, and also physically walk while in 3D mode without roaming out from the designed available walking space. Initially, the remote user position in the virtual space in the two modes are aligned with the local user's position in the physical space. This allows the system to pin for the initial position so it creates an illusion that the remote guest user would feel as if he or she is walking out from the local host user's perspective when switching from 360 mode to 3D mode.

When the system launches, both users' position and rotation values are saved and constantly synchronized over the network as it allows the view frame and other visual cues to reposition and visualize correctly on the scene. However, as the 360 mode requires only a rotation value from both users, the position value is not used for visualisation but only kept internally. When the remote guest user enters the 3D mode, both users' position and rotation values are used to reposition both users at the appropriate coordinate in the shared space. The remote guest user is free to walk and discover around in the 3D mode and also allowed to switch between modes at any time as they want.

Implementation

Our prototype system is developed using Unity (ver 2017.3.1f1) game engine running on a PC (Intel Core i7-6700 3.40GHz CPU, 16GB DDR4 RAM, NVIDIA GeForce 950 GPU) with Windows 10 operating system that drives the VR HMD on the remote guest user's side. The remote guest user wears Samsung Odyssey HMD [34] that supports 1,440 x 1,600 pixels resolution with a refresh rate between 90Hz to 60Hz. On the other end, the local host user wears a Microsoft HoloLens [23] AR HMD running our prototype software built using the same Unity game engine.

The user coordinates and visual cues are synchronised over a Wifi connection. To share the live 360 video feed, we used a Ricoh Theta V [33] that supports 4k resolution live streaming mounted on top of the HoloLens. While the 360 camera we used supports streaming over Wifi in lower resolution and higher delay, we used a 10 meter high speed 3.0 USB connection for the user study. In order to set up the system to support 3D mode, the 3D model of the local host user's physical environment has to be reconstructed using a photogrammetric processing software [3] and shared to the remote guest user's side to be loaded into the system. After that, both local host and remote guest users can start the system and connect over the network.

To switch between 360 and 3D modes during the collaboration, the remote guest user has to perform a specific hand gesture using their hands (thumbs up on both hands). This allows the remote guest user to move away from the local host user and focus on a different object from a different viewpoint, navigating in the 3D scene as explained in the previous section.

4 USER STUDY

We conducted a user study to compare and evaluate our system and identify its benefits and limitations. This allows us to use the data to explore the validity of the idea of combining 360 panorama video and 3D reconstruction techniques into a hybrid remote collaboration system. We used object searching tasks to collect data and feedback for evaluation. This section describes our research questions, study design and set-up including tasks, conditions, measures, and the full procedure of the user study.

We divided our user study into two parts. The first part (Part A) of the study was designed as a user experiment comparing between the 360-only and 3D-only remote collaboration systems to observe the benefits and limitations of each approach, while the second part (Part B) was designed to evaluate the proposed hybrid system that combines the two. As our hybrid system allowed users to freely switch between the two modes, there is a risk that some of the participants might end up using the hybrid system only in one mode which will basically render the comparison of the three options invalid. As an alternative, we decided to compare the two systems (360-only and 3D-only) in the first part of the study, while evaluating the hybrid system in the second part of the study. Here we first describe the user study environment and set-up that is common in both parts, then describe the details of the study design of each part.

Study Environment and Set-up

The user study was conducted in two rooms connected to each other through a door. The local host user was in a room sized approximately $6m \ge 8m$, while the size of the other room where the remote guest user was in was approximately $3m \ge 4m$. Figure 3 shows the two rooms.



Figure 3: Experimental setup: task description (top left) and study environment (top right and lower left)

The local host user's room was used as the main task space for collaboration, and it had furniture including a bookshelf, two desks, a small round table and chairs. All of the furniture except the round table was placed towards the wall or at corners in order to leave free space for the user to walk around in the room. The 3D reconstruction of this first room was prepared in advance as described in section 3. The other room mainly had a free walking space for the remote guest user wearing a VR headset, except a couple of desks at one of the corners with computers for running the VR system and collecting data. We note that we had to scale the 3D reconstructed scene from its actual size to 90% to reduce the space requirement on remote guest user side for movement, however, the user still needed to physically walk or move in order to navigate in the 3D scene.

As apparatus for the user study, we used our prototype system as described in section 3. In Part A of the study, the system worked as a 360-only or 3D-only in the relevant condition and participants were not be able to switch between the modes. In Part B of the study, users were able to freely switch between the two modes.

An actor was employed in both parts of the study to maintain the flow. We noted that we may not be able to capture sufficient amount of data if letting participants work in pairs, since the task could be completed by the local user alone, especially when a proactive local user is paired with a submissive remote guest user. This could also make the experiment finish without using any of the system features. In addition, if participants worked in pairs, the remote guest participant may not be aware of that they could instruct the local host user to move to a certain place. Therefore, we employed an actor to play the role of a partner with whom a participant needs to collaborate.

Part A - 360 vs. 3D

The aim of the first part of the study was to compare the two remote collaboration mediums: 360 live panorama video and 3D reconstructed scene. The main research questions in this part were:

- (1) How does the remote collaboration medium affect the task performance of collaborative object searching task?
- (2) How does the remote collaboration medium affect the user experience including usability, sense of being together, and motion sickness?
- (3) Would a certain type of medium (360 or 3D) be a dominant choice for remote collaboration? Or would there be room for complementing each other?

Experimental Design. This part of the study was designed as a within-subject experiment with two conditions, 360 and 3D, where the participants used either the 360 or the 3D version of the system to collaborate. The order of condition was counterbalanced between participants. As both 360 and 3D representation of the remote environment is only shown to the remote guest user, we had participants play only the role of the remote guest while we had an actor playing the role of the local host. This is because the presentation of the captured task environment (360 or 3D mode) mainly affecting the remote guest user. However, note that we let the participants also experience the local host side later in Part B. As the collaboration task was in a remote expert scenario where the remote guest user gives instruction and the local host user mostly follows, the impact of the actor's behaviour on the task performance and experience was not that concerning.

Experimental Task. One of the basic tasks in remote collaboration is finding objects of interest and manipulating them collaboratively [18]. In this study we focused on object search task which is usually the first step in remote collaboration that requires the remote user to understand and explore the task space. We used 45 pieces of Duplo Lego blocks in different colours and sizes as objects of interest that the participants need to find among other physical objects (e.g. books and boxes) present in the task space. As a remote guest user, participants were given with a description of target objects to look for, and they were to guide and instruct the actor as a local host user in the physical task environment to find the target objects and mark them by placing a sticky note in front of them. This may include asking the actor to physically move in the task environment or lean towards a reachable spot to find or discover objects, especially in the 360 condition. The task ended once the remote guest user found all of the objects and they were all marked using sticky

notes. Each condition had the same number of target objects to find.

There were two types of tasks depending on how the target object was described. The first type of task instructed users to find blocks based on visual descriptions, e.g. in a specific colour and shape. For example, find nine of blue cube-shaped blocks. The second type of task described spatial configuration and relationship between a set of blocks. Participants were required to find a group of objects in a specific colour and are close by, adjacent, parallel or stacked to each other. For example, these can be finding a pair of red and blue cube-shaped blocks next to each other. In each condition participants performed two tasks with different task descriptions.

The task descriptions included text and images shown in the virtual environment and also printed on a paper in the task environment (see Figure 3). For the 360 condition, participants can see the printed task description paper on the round table in the middle of the room through 360 view. In the 3D condition, the task description is visualised at the same place in the 3D space as 2D text and image.

Four unique task descriptions, two in each type, were prepared so they could be used in different combinations with the conditions. Order and combinations of tasks and conditions were counterbalanced to reduce bias. To make the difficulty level balanced between task descriptions, we first categorised the target objects into three levels of difficulty: Easy, Medium, and Hard. Easy objects were in places that are obvious to the users such as on a table without being occluded by any other objects. Medium level difficulty objects were either partially occluded by non-task related objects or on the ground where the users were required to bend their body to see. Hard type objects were those fully occluded by non-task related objects, hence the users had to move around the occluding object to see them. Based on this categorisation, we tried to make each task include the same amount of objects from each difficulty level.

Measurements. As dependent variables, we collected both objective measures and subjective feedback in each condition. The main objective measure was task completion time as an indication of task performance, and was measured based on video recording. For subjective feedback we collected questionnaire responses. For each task, we measured user experience through a Single Ease Question (SEQ) [35], and a few other custom rating items based on prior work [20]. After each condition, to measure the sense of being together we used the Networked Mind Measure of Social Presence Questionnaire (SoPQ) [12] and the MEC Spatial Presence Questionnaire (SpPQ) [38]. We also measured the usability of the system using the System Usability Scale (SUS) [5], and motion sickness using Simulator Sickness Questionnaire

(SSQ) [17]. At the end of Part A, participants were asked to choose their preferred condition under various categories, and also provide qualitative feedback by answering open questions in the post-experiment questionnaire.

Procedure. The experiment started from the researcher briefing the study information to the participant, then the participant signing a consent form once they agreed to participate. The participant was then asked to complete a pre-experiment questionnaire collecting demographic information and an SSQ to measure a ground level of motion sickness. Then a training session was held for the participant to learn the tasks and build up their understanding of the task environment.

After finishing the training, the participant was asked to wear the HMD on the remote user's side and they were provided with a task description of the first task in the first condition. While performing the task, the participant as a remote guest guided the local host user (an actor) to move around in the room and place sticky notes on the correct target objects they found using a combination of verbal and nonverbal communication. After finishing the first task, the participant had a short break along with answering a per-task questionnaire. The participant then proceeded into the second task. After completing the second task, the participant was asked to complete another per-task questionnaire, followed by a per-condition questionnaire and motion sickness questionnaire. This process was repeated with the second condition. The order of conditions was counterbalanced between participants.

After completing both of the conditions, the participant was asked to complete a post-experiment questionnaire.

Part B - Hybrid (360 + 3D) System

In the second part of the study (Part B), we focused on evaluating the proposed hybrid system that combines 360 and 3D modes. We designed this session as an explorative study where the participants tried using the hybrid system on both the remote and local user's side, and provided subjective feedback through a questionnaire.

First, the participants tried the remote guest user's side as they did in Part A but with being able to freely switch between 360 and 3D modes to perform a search task (similar to the task in Part A). While performing the task, we recorded the number of times the participant switched between the two modes, and the amount of time they spent in each mode. After trying the hybrid system, participants were asked to complete a survey questionnaire that asked if switching between 360 and 3D is useful and easy to use, what they liked and what could be improved in the hybrid system, and to rank the three options (360-only, 3D-only, and 360+3D) based on their preference. Next, the participants tried the prototype system on the local host user side with another task where they received instructions from the actor on the remote guest user side, regularly switching between the 360 and 3D modes. The remote guest actor notified the participants verbally when switching between modes as the visual representation on local user side could have only subtle change, making it hard for the participant to notice. After finishing a task on the local host user side, participants were asked to fill out a questionnaire to collect feedback and recommendations for improvements, as well as to choose the preferred mode (360 or 3D) under various categories, as they did at the end of Part A. This was immediately followed by a short debriefing, and the study concluded with providing a complimentary cafe voucher to the participant.

The experiment including both Part A and B took about 70 minutes on average for each participant.

5 RESULTS

In this section, we report the study results with statistical analyses ($\alpha = .05$, unless noted otherwise) and summarising qualitative feedback collected from the participants. We recruited 20 participants (16 male, 4 female) from the local campus community with their age ranging from 18 to 45 years old. Most of the participants (n = 13) were familiar with VR system stating that they have been using VR system at least a few times a month, while a quarter of participants (n = 5) have never used a VR system before.

Part A - 360 vs. 3D

Task Completion Time. On average, participants took a longer time to complete the tasks in 3D mode compared to 360 mode, and also the tasks given in visual description took longer than those described with object relationship (see Figure 4). A Shapiro-Wilk test found some of the conditions were not following normal distribution, so we applied Align Rank Transform (ART) [39] before using a Repeated-Measures ANOVA for factorial analysis. The results showed that there was a significant main effect of both collaboration medium (360 vs. 3D, F(1, 19) = 4.682, p = .043) and task (task1 vs. task 2, F(1, 19) = 45.533, p < .001). No significant interaction between the two was found (F(1, 19) = 0.688, p = .417).

Subjective Ratings on User Experience. The results from SEQ [35] and custom rating items are shown in Figure 5. All of the rating items were answered on a 7-point Likert scale (SEQ - 1: very difficult – 7: very easy; All other items - 1: strongly disagree – 7: strongly agree). Using a Repeated-Measures ANOVA with ART [39], we found that using the 360 mode was significantly easier than the 3D mode (F(1, 19) = 5.652, p = .028), regardless of the task (task: F(1, 19) = 3.030, p = .098; interaction: F(1, 19) = .085, p = .774). We also



Figure 4: Average task completion time (*: significant effect)



Figure 5: Subjective ratings on user experience (*: significant effect)

found that the 360 mode provides a significantly better understanding of partner's focus (F(1, 19) = 7.797, p = .012), again regardless of the task (task: F(1, 19) = 1.431, p = .246, interaction: F(1, 19) = 1.316, p = .266). No significant difference was found regarding how much participants enjoyed the experience, nor on being able to focus on the task activities.

Social Presence. The social presence questionnaire (SoPQ) [12] included three sub-scales – Co-presence (CP), Attentional Allocation (AA) and Perceived Message Understanding (PMU) – and consisted of eighteen rating items on a 7-point Likert scale (1: strongly disagree – 7: strongly agree). Overall, we found significant difference in favor of the 360 mode (Z = -2.931, p = .003) as shown in Figure 6. We also analyzed each sub-scale and found significant differences in AA (Z = -3.103, p = .002), and PMU (Z = -2.000, p = .045), but not in CP.

Spatial Presence. From SpPQ [38], we used two subscales – Spatial Presence Self Location (SL) and Spatial Situation model (SSM) – that consisted of eight rating items on a 5-point Likert scale (1: Fully Disagree – 5: Fully Agree). Similar to the SoPQ, we analyzed the results as a whole, as well as in each sub-scale. However, the results reported no significant difference.



Figure 6: Results of Social Presence questionnaire (SoP: Overall Social Presence, CP: Co-presence, AA: Attentional Allocation, PMU: Perceived Message Understanding; *: statistically significant).

Simulator Sickness. Figure 7 shows the average score of Simulator Sickness Questionnaire [17] which comprises of fourteen symptoms rated on a scale of (0: none - 3: severe). Overall, participants experienced mild (mostly lower than 1) symptoms of motion sickness in either condition. The results were analyzed with a Friedman test comparing: before the experiment (Initial), after trying 360 mode (360), and after trying 3D mode (3D). For post hoc pairwise comparisons, we used Wilcoxon Signed Rank tests with Bonferroni correction (α = .0167). We found significant differences between Initial and 360 (Z = -3.297, p = .001), and also between Initial and 3D (Z = -3.184, p = .001). However, no significant difference was reported between 360 and 3D modes (Z = -3.297, p = .0297). This indicates that participants were induced by a mild level of motion sickness after attempting the system in either mode. We also conducted a similar set of analysis based on the order of condition experienced, and obtained similar results, i.e. significant increases from the initial measurement (first condition: Z = -3.03, p = .002, second condition: Z = -3.435, p = .001) while no significant difference was seen between conditions (Z = -.0694, p = .0488).

System Usability. In terms of SUS [5], both systems were rated around the average level of usability (360: M = 66.75, Md = 70, IQR = [56.5 - 75.5]; 3D: M = 61, Md = 58.5, IQR = [47.5 - 73.5]). While the 3D mode received lower ratings, the difference between the conditions was not statistically significant based on a Wilcoxon Signed Rank test (Z = -1.942, p = .052).

User Preference. In terms of preference, the participants were mostly split between the conditions across the categories (see Figure 8). Based on Chi-square goodness-of-fit tests comparing against random choices, there was only one category



Figure 7: Results of average simulator sickness score for different conditions. (*: statistically significant).



Figure 8: User preference between 360 and 3D modes as a remote guest (*: statistically significant).

where participants showed strong preference towards 360 over 3D: understanding the partner's focus ($\chi^2(1) = 8.45$, p = .0037).

Even though more than half of the participants preferred the 360 mode in the majority of the categories, more than half of them still preferred 3D mode as their overall preference. Based on qualitative feedback, this was mainly because being inside a 3D environment provided them with a more realistic interaction as well as better understanding of the surroundings. Participants mentioned (P9, Male, Mid 30s, Experienced) *"I felt more feeling of being there"*, (P10, Female, Early 20s, Inexperienced) *"More interaction with the body.* It gives more realistic and sense feeling. It is connecting the vision and body movement together. It simulates the real life experience.", (P16, Male, Mid 20s, Experienced) *"It felt more realistic and interactive than the other condition"*.

A few of the other participants also highlighted their favor of being able to move independently and improving their task performance, mentioning (P5, Male, Early 20s, Experienced) *"It was nice to be able to explore the area without being tied to the other user..."*, (P3, Female, Early 30s, Experienced) *"I was able to walk around, which make me more immersed into the environment." However, there were also participants preferring not to move around by themselves as a benefit,*



Figure 9: User preference based ranking results (*: statistically significant).

mentioning (P11, Male, Late 10s, Experienced) "I found I could focus more on the task when I didn't have to move around...", (P9) "I don't need to worry about moving around while I cannot see what is in front of me in the real world", (P8, Female, Early 30s, Inexperienced) "Walking around a bit more, not sure if I was going to walk into something...". Those who preferred 360 mode in most of the categories mostly valued the ease of understanding their partner's focus and location, i.e. in 3D mode the participant needed to search for the other user's location in the scene whereas in 360 mode it was always in the first person perspective.

Part B - Hybrid (360 + 3D) System

Remote Guest User. Usage data was collected from 14 participants (due to video recording failure) which showed that participants spent, on average, 157.9 seconds (SD = 47.6) using the hybrid system on the remote guest user's side, and switched between the 360 and 3D modes about 7.8 times (SD = 4.6) on overage. Looking at the proportion of time spent in each mode, participants were using the 360 mode slightly more (56.6%) than the 3D mode, yet the difference was not statistically significant (t(13) = 0.88, p = .395). There were six participants who spent more than 70% of their time in 360 mode, while three participants spent over 70% of their time in 3D mode.

After trying the hybrid system, participants (n = 20) were asked to rank (1: best – 3: worst) all the three options they have experienced based on their preference, and the majority (65%) of the participants ranked the hybrid system as the best (see Figure 9). A Friedman test indicated that there was a significant difference in ranking between the three systems ($\chi^2(19) = 8.10$, p = .017). Post hoc tests using Wilcoxon Signed Rank tests with Bonferroni correction ($\alpha = .0167$) showed that participants preferred the hybrid system (360+3D) significantly more than the other systems: 360-only (Z = -2.938, p = .003) and 3D-only (Z = -3.257, p = .001). There was no significant difference between the 360-only and the 3D-only.

Participants who rated the hybrid system as the best mostly enjoyed the ability to switch between the two modes at any time, complementing each other, mentioning (P3) "When combined, I was able to switch and choose the preference I want to use to help aid me in finding that suits me in that situation", (P8) "switching between views was good, made it



Figure 10: Rating on Switching between 360 and 3D modes.



Figure 11: User preference between 360 and 3D modes as a local host (*: statistically significant).

faster and could check objects for clarity", (P10) "Some aspect which is not working well in one (mode), worked fine in the other.", (P12, Male, Early 20s, Experienced) "I was able to use 360 to view my partners when needed, and use 3d to move around myself when I want to explore."

We also asked the participants (n = 20) to rate on a on a 7-point scale (1: strongly disagree – 7: strongly agree) on how much they agree with the statement 'Switching between 360 and 3D is useful' and 'Switching between 360 and 3D is easy to use'. The results (see Figure 10) showed most of the participants rated positively on the switching feature being useful (n = 16, 80%), and easy to use (n = 17, 85%). Onesample Wilcoxon Signed Rank tests showed these ratings were significantly different from neutral level rating (useful: Z = -2.082, p = .005, easy: Z = 3.355, p = .001).

Local Host User. After participants (n = 16 due to set up issues) tried the system at the local host user side, more than half of the participants preferred using 360 mode over 3D mode in all categories (see Figure 11). However, the result from Chi-square goodness-of-fit tests against random choice showed that the preferences were not strong enough to be statistically significant, except for one category: understanding the partner's focus ($\chi^2 = 5.06$, p = .0245). This is therefore suggested as the biggest benefit of the 360 mode.

Qualitative feedback from the participants indicated that the main reason behind their favour of the 360 mode was because of the limited FOV (field of view) on the AR headset that restricted the range of visual cues participants can see at a time. Participants indicated that spotting visual cues from the first person view in 360 mode was easier than from the second person view in 3D mode as they mentioned (P7, Male, Mid 20s, Experienced) "Because FOV was small, locating remote guest user was little bit hard for 3D reconstruction condition", (P9) "It was more obvious to find where he was pointing at.", (P13, Male, Early 20s, Experienced) "The attention grabbing arrow was slightly confusing at times, mainly due to the HoloLens' field of view", (P10) "... Most of the time, I am spending time searching the arrow to check out where the guest is looking at and pointing at. When I found the arrow, it moved again.". However, one of the participants valued the benefit of using 3D mode especially being able to see the 3D avatar of the remote guest without concerning the limited FOV, mentioning (P4, Male, Late 20s, Experienced) "having another character walking around makes me feel like two of us working together."

Suggested System Improvements

Participants were asked about what to improve in the current system in an open-ended question at the end of both Part A and B and also during debriefing. A majority of the participants mentioned needing improvements in visual quality as the color offset, a lighting issue from the camera, and image resolution were obvious problems. This was due to the experimental task which heavily relied on searching for objects by their colors and shapes. From the hardware aspects, a participant mentioned (P8) *"More comfortable headset if you have to wear it all the time!"* due to the prolonged task duration.

There were also useful suggestions made on the system features. With the current implementation of the system, users switching from one mode to another would result a sudden change of viewing perspective and location in the shared space. This potentially caused a mild level of disorientation and motion sickness to the user. Some participants mentioned of adding a transition effect of teleporting from one spot to another upon switching between modes to reduce motion sickness. One of them mentioned (P13, Male, Early 20s, Experienced) *"Perhaps a fast jump to the other users position or a blackout effect when changing to the 360 view to reduce disorientation"*.

Participants also suggested updating the remote guest user's position in the virtual space according to the local host user's position in the physical space when switching between modes. Participants mentioned (P5) "(would be nice) if you could switch back to 360 and see the same perspective as was in 3D", (P2, Male, Late 30s, Experienced) "... the alignment of AR/VR spaces", (P7) "...when user switches from 360 view to 3d reconstruction, I wish camera position and view was the same as 360 view, so that I can start navigating space from AR user's last view."

Other suggestions included improving existing features, such as (P12) "A small miniature (3D model) view that's fixed

to the user view for the 3D vision would help show where the other person is looking more often...", (P15, Male, Mid 20s, Inexperienced) "... an indicator on my partner's display giving the mode the I am currently in" to help better understanding of each other during the collaboration. There were also suggestions about ways to combine the two modes other than simply switching between them, such as (P16) "If the panorama had the freedom to move like the reconstruction experiment then I would gladly use that. Panorama adds more realism to it so (would be nice) if that can be improved on."

6 **DISCUSSION**

360 vs. 3D

Overall, the results indicated that participants performed better with the 360 mode compared to the 3D mode, showing significantly lower task completion time, feeling the task was easier, understanding the partner's focus better, and perceiving improved social presence. However, the 3D mode also provided comparable user experience and usability, and with its benefits including the remote guest being able to navigate the scene independently, lead to participants' preference being split between the two options. Slightly over half of the participants expressed their preference towards the 360 mode, while nearly half of the participants preferred the 3D mode. Having no statistically significant difference in preference indicated participants value both 360 and 3D and potentially need an option to combine the two.

These findings could be influenced by the experiment where colours and shapes were critical. In the experiment, participants were accessing the shared space from the remote guest side through the live video from 360 panorama camera or 3D reconstructed scene using the HMD. This creates an implication of heavily relying on the image quality of the 360 panorama camera, HMD or the reconstructed scene to aid on recognising the correct colours and shapes that did not work well in certain cases. For example, there was a lightning issue that made colours appear different in brighter or darker area. This caused confusion making participants to rely more on verbal communication, spending more time to confirm object attributes especially in 3D mode due to the insufficient quality of 3D scene. We note that the issue of colour presentation in 3D is comparable to camera resolution limitation in 360 camera. Hence, it does not make any particular mode being superior for the given task.

Regarding the effect of simulator sickness in different modes, it was indicated that 3D mode gave a lower average simulator sickness score compared to 360 mode even though participants spent longer time inside the 3D mode compared to 360 mode. This was confirmed as the 3D mode took longer task completion time on average than the 360 mode. While there is a need for further investigation with longer term usage, the qualitative feedback given by the participants suggested the full control of the movements reduced the simulator sickness perceived.

In terms of user communication and attention awareness, the ratings in Attentional Allocation and Perceived Message Understanding from Social Presence, and ratings from user preference strongly suggested that 360 mode provides an easier understanding of communication and user's focus and attention. This was also consistent with the qualitative feedback given by the participants. We postulate that the first person perspective view from 360 mode plays an important role in supporting the results.

View and Annotation Independence

While our study compared the two modes, we suspect that the implication of user independence could have played a major role. The task performance would have been improved with both modes allowing users to look independently in the scene. This is supported by prior work observing on the effects of view independence in 360 [22] and 3D reconstruction based [36] remote collaboration systems. This also aligns with the result from another similar prior work [11] that allowed independent annotation in a remote collaboration using AR tablet and desktop set up. This suggests need for further investigation to observe the effects of view and annotation independence that can be brought to different configuration modes in a MR remote collaboration.

Implication for merging two modes together

In this paper we proposed a system that allows switching between 360 and 3D modes during a remote collaboration. The purpose of such design was to provide alternative options for a user to perform collaboration through different perspectives and features. The study results indicated that more than half of the participants (13 out of 20) preferred having the ability to switch between 360 and 3D modes as it allows them to perform tasks using the mode they found working the best in the given situation.

However, it is also noticeable that not all of the participants would prefer having a system with both modes (when the two other modes are on offer). The qualitative feedback result suggested that some participants found that the hand tracker was hard to use, yet the main reason lied behind the sudden change of perspective when switching between the modes. Although some participants mentioned the gestures for switching between the modes were easy to use, we observed few of the participants struggling with having their gesture not working properly during the experiment, either the mode not switching at times or switching unintentionally. Hence, few participants preferred having just 360 mode to prevent any occurrence of a similar incident as the sudden change of perspective when switching between modes can potentially lead to disorientation.

Limitations

Although the user study helped evaluating and measuring the system in a controlled environment with an experimental task, there are still many aspects that need to be further investigated. Some of the obvious issues were the limitation of hardware devices and the choice of the task that heavily relied on colour and shape which could have added noises to our results. As suggested by some of the participants, the result could be different if the task did not rely on such aspect. Likewise, if the visual quality of the reconstructed 3D scene was comparable to the live 360 video, the results may have turned out in favour of 3D mode.

Another limitation of the user study was employing an actor instead of letting two participants collaborate together. While we decided to employ an actor to control some of the random variables, it would be also interesting to conduct a future study by recruiting participants in pairs to observe any significant changes upon user preference and other aspects of the user experience. This is especially essential for a remote collaboration system as it involves random pair of users in the real world, and it may also help identifying the issues on the local host side.

The prototype system implementation also had certain limitations that would need improvement in the future studies. One of the obvious issue was the visual cue offset and 3D scene quality that was directly influencing the user experience. More accurate calibration and a better reconstruction of 3D scene could be worth for further development. Another limitation was the height offset in 360 mode as the 360 panorama camera was mounted on top of the MR headset resulting the viewpoint of the remote guest being slightly higher than the local host. While this was not a huge problem in the user study, it would be useful to further investigate how to compensate the height difference between the two users.

7 CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel MR remote collaboration system that integrates 360 panorama video and 3D reconstructed scene. The system allows switching between 360 and 3D modes to introduce variation to solve collaborative tasks. A user study was conducted to evaluate the advantage and limitation of each mode independently, as well as to solicit feedback on merging both modes into a single system. The results showed that participants preferred having both 360 and 3D modes, as it provides variation in controls and features from different perspectives. Participants reported that the main benefit of being able to switch between modes was that they were able to enjoy focusing on performing their own task in 3D mode, while better understanding the partner's focus in 360 mode. They were also be able to take over the movement controls at any time for different tasks in different situations. While comparing between 360 mode and 3D mode, there was some slight evidence supporting the preference of 360 mode due to the easier communication and understanding of the partner in the experiment.

We note that there is a need for further investigation with a larger number of participants and in pairs, as well as revising the task set-up to be less reliant on factors that can be influenced by different lightning. Recruiting participants in pairs will also help further investigating social presence and communication behaviours. Adding other visual cues such as drawing annotation, and full body avatars [2, 15, 27, 29, 31] in the 3D scene would be interesting to explore in the future, as well. We would also like to explore the suggestions from participants such as having a static live video inset representing the local host user's perspective or a transition effect for switch between modes. We expect that conducting further studies with different set-ups and improved system features would build a stronger implication towards a MR remote collaboration system combining 360 videos and 3D reconstruction that will become a new form of approach for telecommunication and remote collaboration in the future.

ACKNOWLEDGMENTS

This research is supported by a CSIRO Data61 PhD scholarship and South Australian Research Fellowship project.

REFERENCES

- [1] Matt Adcock, Stuart Anderson, and Bruce Thomas. 2013. Remote-Fusion: 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. https://doi.org/10.1145/2534329.2534331
- Matt Adcock, David Feng, and Bruce Thomas. 2013. Visualization of Off-surface 3D Viewpoint Locations in Spatial Augmented Reality. In Proceedings of the 1st Symposium on Spatial User Interaction (SUI '13). ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/2491367. 2491378
- [3] Agisoft. [n. d.]. PhotoScan. Retrieved January 1, 2019 from http: //www.agisoft.com
- Mehdi Boukhris, Alexis Paljic, and Dominique Lafon-Pham. 2017. 360° versus 3D Environments in VR Headsets for an Exploration Task. In International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, ICAT-EGVE 2017. 71–78. https://doi.org/10.2312/egve.20171341
- [5] John Brooke. 1996. SUS: a quick and dirty usability scale. In Usability Evaluation in Industry, P.W. Jordan, B. Thomas, B.A. Weerdmeester, and I.L. McClelland (Eds.). Taylor and Francis, London, 189–194.
- [6] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. 2016. Workspace Awareness in Collaborative AR Using HMDs: A User Study Comparing Audio and Visual Notifications. In Proceedings of the 7th Augmented Human International Conference 2016 (AH '16). ACM, New York, NY, USA, Article 3, 8 pages. https://doi.org/10.1145/

2875194.2875204

- [7] Angela Dai, Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Christian Theobalt. 2017. BundleFusion: Real-Time Globally Consistent 3D Reconstruction Using On-the-Fly Surface Reintegration. ACM Trans. Graph. 36, 4, Article 76a (May 2017). https://doi.org/10.1145/ 3072959.3054739
- [8] Samuel Dong and Tobias Höllerer. 2018. Real-Time Re-Textured Geometry Modeling Using Microsoft HoloLens. In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 231–237. https: //doi.org/10.1109/VR.2018.8447549
- [9] Carmine Elvezio, Mengu Sukan, Ohan Oda, Steven Feiner, and Barbara Tversky. 2017. Remote Collaboration in AR and VR Using Virtual Replicas. In ACM SIGGRAPH 2017 VR Village (SIGGRAPH '17). ACM, New York, NY, USA, Article 13, 2 pages. https://doi.org/10.1145/ 3089269.3089281
- [10] Dirk Farin, Wolfgang Effelsberg, and Peter H. N. de With. 2007. Floorplan Reconstruction from Panoramic Images. In *Proceedings of the 15th ACM International Conference on Multimedia (MM '07)*. ACM, New York, NY, USA, 823–826. https://doi.org/10.1145/1291233.1291420
- [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. https://doi.org/10.1145/2642918.2647372
- [12] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence. In Seventh Annual International Workshop: Presence 2004. 246–251. http: //cogprints.org/7026/
- [13] Keita Higuch, Ryo Yonetani, and Yoichi Sato. 2016. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 5180–5190. https://doi.org/10.1145/2858036.2858438
- [14] Shahram Izadi, Richard A. Newcombe, David Kim, Otmar Hilliges, David Molyneaux, Steve Hodges, Pushmeet Kohli, Jamie Shotton, Andrew J. Davison, and Andrew Fitzgibbon. 2011. KinectFusion: Realtime Dynamic 3D Surface Reconstruction and Interaction. In ACM SIGGRAPH 2011 Talks (SIGGRAPH '11). ACM, New York, NY, USA, Article 23, 1 pages. https://doi.org/10.1145/2037826.2037857
- [15] Michal Joachimczak, Juan Liu, and Hiroshi Ando. 2017. Real-time Mixed-reality Telepresence via 3D Reconstruction with HoloLens and Commodity Depth Sensors. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017)*. ACM, New York, NY, USA, 514–515. https://doi.org/10.1145/3136755.3143031
- [16] Shunichi Kasahara and Jun Rekimoto. 2015. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera for Remote Collaboration. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST '15)*. ACM, New York, NY, USA, 217–225. https://doi.org/10.1145/2821592.2821608
- [17] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal* of Aviation Psychology 3, 3 (1993), 203–220. https://doi.org/10.1207/ s15327108ijap0303_3
- [18] Seungwon Kim, Gun Lee, Nobuchika Sakata, and Mark Billinghurst. 2014. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 83–92. https://doi.org/10.1109/ISMAR.2014.6948412
- [19] LeapMotion. [n. d.]. LeapMotion. Retrieved January 1, 2019 from https://www.leapmotion.com

- [20] Gun A. Lee, Seungwon Kim, Youngho Lee, Arindam Dey, Thammathip Piumsomboon, Mitchell Norman, and Mark Billinghurst. 2017. Improving Collaboration in Augmented Video Conference using Mutually Shared Gaze. In International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, ICAT-EGVE 2017. 197–204. https://doi.org/10.2312/egve.20171359
- [21] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2017. Mixed Reality Collaboration Through Sharing a Live Panorama. In SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications (SA '17). ACM, New York, NY, USA, Article 14, 4 pages. https://doi.org/10. 1145/3132787.3139203
- [22] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2018. A User Study on MR Remote Collaboration using Live 360 Video. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR '18). Article 51, 153–164 pages. https://doi.org/10.1109/ISMAR.2018.00051
- [23] Microsoft. [n. d.]. HoloLens. Retrieved January 1, 2019 from https: //www.microsoft.com/en-us/hololens
- [24] Microsoft. [n. d.]. Kinect for Windows. Retrieved January 1, 2019 from https://developer.microsoft.com/en-us/windows/kinect
- [25] Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2017. Evaluating Snapping-to-photos Virtual Travel Interfaces for 3D Reconstructed Visual Reality. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17). ACM, New York, NY, USA, Article 22, 11 pages. https://doi.org/10.1145/3139131.3139138
- [26] 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
- [27] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517
- [28] Mai Otsuki, Keita Maruyama, Hideaki Kuzuoka, and Yusuke SUZUKI. 2018. Effects of Enhanced Gaze Presentation on Gaze Leading in Remote Collaborative Physical Tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 368, 11 pages. https://doi.org/10.1145/ 3173574.3173942
- [29] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16). ACM, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2818048.2819965
- [30] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring Enhancements for Remote Mixed Reality Collaboration. In SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications (SA '17). ACM, New York, NY, USA, Article 16, 5 pages. https://doi.org/10.1145/3132787.3139200
- [31] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Article 46, 13 pages. https://doi.org/10.1145/3173574.3173620

- [32] Taehyun Rhee, Lohit Petikam, Benjamin Allen, and Andrew Chalmers. 2017. MR360: Mixed Reality Rendering for 360° Panoramic Videos. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (April 2017), 1379–1388. https://doi.org/10.1109/TVCG.2017.2657178
- [33] Ricoh. [n. d.]. Theta V. Retrieved January 1, 2019 from https://theta360. com/en/about/theta/v.html
- [34] Samsung. [n. d.]. Odyssey. Retrieved January 1, 2019 from https: //www.samsung.com/us/computing/hmd/windows-mixed-reality/ hmd-odyssey-windows-mixed-reality-headset-xe800zba-hc1us/
- [35] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of Three Onequestion, Post-task Usability Questionnaires. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). ACM, New York, NY, USA, 1599–1608. https://doi.org/10.1145/1518701. 1518946
- [36] Matthew Tait and Mark Billinghurst. 2015. The effect of view independence in a collaborative ar system. *Computer Supported Cooperative Work (CSCW)* 24, 6 (2015), 563–589.

- [37] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration with 360° Videochat: Challenges and Opportunities. In Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17). ACM, New York, NY, USA, 1327–1339. https://doi.org/10.1145/3064663.3064707
- [38] Peter Vorderer, Werner Wirth, Feliz R Gouveia, Frank Biocca, Timo Saari, Futz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, Tilo Hartmann, et al. 2004. MEC spatial presence questionnaire (MEC-SPQ): Short documentation and instructions for application - Report to the European community, project presence: MEC (IST-2001-37661). Technical Report.
- [39] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963