

A User Study on MR Remote Collaboration using Live 360 Video

Gun A. Lee* Theophilus Teo Seungwon Kim Mark Billinghurst

School of Information Technology and Mathematical Sciences, University of South Australia

ABSTRACT

Sharing and watching live 360 panorama video is available on modern social networking platforms, yet the communication is often a passive one-directional experience. This research investigates how to further improve live 360 panorama based remote collaborative experiences by adding Mixed Reality (MR) cues. SharedSphere is a wearable MR remote collaboration system that enriches a live captured immersive panorama based collaboration through MR visualisation of non-verbal communication cues (e.g., view awareness and gestures cues). We describe the design and implementation details of the prototype system, and report on a user study investigating how MR live panorama sharing affects the user's collaborative experience. The results showed that providing view independence through sharing live panorama enhances co-presence in collaboration, and the MR cues help users understanding each other. Based on the study results we discuss design implications and future research direction.

Keywords: Mixed Reality, Augmented Reality, remote collaboration, live panorama sharing, view independence.

Index Terms: H.4.3 Computer conferencing, teleconferencing, and videoconferencing; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities

1 INTRODUCTION

Over the last generation there has been active development of new technologies for richer telecommunication and remote collaboration. The main modality of telecommunication has been shifting from audio to video, and now video conferencing is becoming a regular part of people's social life. Furthermore, new technologies, such as 360 video, have emerged to allow people to have even more immersive shared experiences.

Recently, major social networking platforms (e.g. Facebook Live, Periscope Live 360, and YouTube) have enabled people to share live 360 panorama video. This changes the collaborative experience from 2D videos to a more immersive experience. While panorama images and videos are still mostly watched on a 2D screen (e.g. desktop monitor or a mobile device), with Virtual Reality (VR) Head Mounted Displays (HMDs) becoming more affordable (e.g. Google Cardboard VR) more users can now watch 360 panorama in an immersive manner.

While sharing and watching immersive live 360 panorama video is becoming possible, interaction between the remote users mostly uses traditional communication methods, such as verbal communication in text or audio, or sharing emoji icons. With such

limitations, the communication is often largely a passive one-directional experience.

Our research explores enhancing the shared live 360 panorama experience by adding non-verbal communication cues using Mixed Reality (MR) [31]. Compared to prior 360 panorama based remote collaboration systems, the main novelty of our prototype system, SharedSphere (see Figure 1), is that it overlays non-verbal communication cues, such as view awareness and virtual hand gesture cues onto a shared live 360 panorama video, and so enhances the remote collaboration to be an active bi-directional experience. In this paper we report on a user study evaluating our prototype system and investigating how view independence provided by 360 video sharing affects collaboration and the usefulness of the MR communication cues.

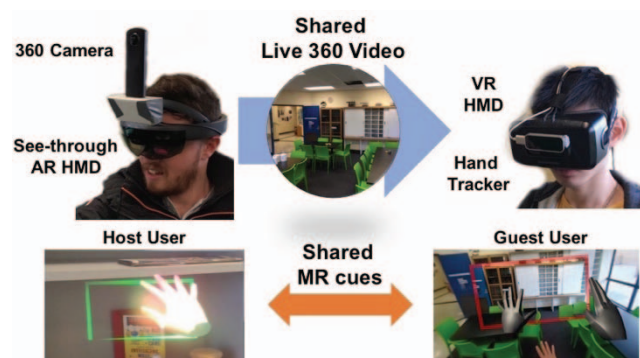


Figure 1: SharedSphere: a MR remote collaboration system.

The main contributions of this paper include:

- Describes design and implementation details of a wearable MR remote collaboration system that is based on sharing live 360 panorama video.
- Reports on a formal user study investigating the benefits of view independency in collaboration using a live 360 panorama video based wearable MR remote collaboration system.
- Reports findings on the benefits of MR communication cues in a shared live 360 panorama video based wearable MR remote collaboration system.
- Provides implications for designing MR remote collaboration interface based on shared live 360 panorama video.

In the rest of the paper we first review related work, and then describe our SharedSphere system, focusing on the MR collaboration cues we have added. Next, we report on a user study with the platform, discuss the results we have found. Finally, we provide some interface guidelines, conclusions and directions for future work.

2 RELATED WORK

In our research we investigate MR remote collaboration that is based on sharing a live 360 panorama video. Here we review related work in MR remote collaboration and live 360 panorama video sharing, and compare our approach to these prior works.

* gun.lee@unisa.edu.au

2.1 Live 360 Panorama Sharing

Since the introduction of QuickTime VR [6], 360-degree panorama pictures and video has been a popular medium for capturing full surroundings of both real-world and virtual scenes. However, due to its requirement for off-line stitching of multiple pictures or video clips, sharing a live stream of 360 panorama video hasn't been actively used for telecommunication. Some of the prior works tried sharing static panorama images [4] or showing a live 2D video insert into a shared static panorama image [32][38] for remote collaboration using handheld mobile devices. With recent advances in mobile hardware technology, affordable 360 panorama cameras (e.g., Ricoh Theta S, Insta360, Samsung Gear 360, etc.) have become available in the consumer market. Some of these even support live streaming of 360 panorama video into social networking platforms. Facebook, Periscope, and YouTube, now support sharing of both pre-recorded and live streaming 360 panorama video, and watching them on a mobile device.

There are a few examples of research projects that explore how live 360 panorama video can be used for remote collaboration. For example, Kasahara *et al.* [22] developed the Jackin Head, headgear with nine cameras around the user's head to capture and share live 360 panorama at the wearer's eye-level, with its orientation stabilized with computer vision techniques. The remote user watched the shared panorama in an HMD, giving them the impression that they were sharing the local user's viewpoint. Their research mostly focused on capturing and sharing the experience with live 360 panorama video with stabilisation, but not on active interaction between the users using MR cues. The person sharing the panorama video did not wear any display, hence voice communication was the main modality for collaboration. In comparison, our research focuses on MR visual cues shared in both directions between the two users for improved non-verbal communication.

Tang *et al.* [42] attached a 360 panorama camera on a monopod mounted on a backpack to share the wearer's surroundings from an exocentric viewpoint. The remote viewer watched the shared panorama video on a tablet with an orientation sensor. They investigated its use in a remote guided touring scenario, and found challenges in communicating location and orientation information because of lack of support for sharing gesture and other non-verbal communication cues. Our research tries to overcome these problems by applying MR techniques to provide spatial and non-verbal communication cues in a shared live 360 panorama.

2.2 Mixed Reality Remote Collaboration

Researchers have explored enhancing 2D video conferencing on a handheld mobile device with sharing spatial communication cues such as a pointer and drawing annotations [8][12][24][29]. To overlay annotations onto the real world objects researchers have used computer vision techniques, such as simultaneous localization and mapping (SLAM), to track and map the 3D environment [13][14][25]. These works have shown that augmented visual cues are helpful for communicating spatial information and improving the sense of being together in video conferencing using handheld mobile devices. Others have further investigated using wearable interfaces [1][15][17][19], and also sharing other modalities through augmented visual cues, such as sharing eye gaze [2][15][18][19] or hand gestures [1][17] overlaid onto the shared live video stream of the real-world environment.

These prior works focused on sharing egocentric viewpoint through live 2D video stream, hence the remote user's view was dependent on the view of the user sharing the video. To overcome such limitation and let the remote users have control of their view, researchers investigated leveraging SLAM based 3D mapping to provide navigation between key frame viewpoints used for

reconstruction [33][34]. Other researchers have also looked into sharing 3D reconstructed scenes to provide more freedom in navigation. Tait and Billinghurst [41] proposed an MR collaboration system that used offline 3D reconstruction of a physical room shared with a remote person watching and navigating the scene on a desktop interface. Sodhi *et al.* [40] attached a pair of depth cameras on a handheld mobile device to capture and share 3D reconstruction of the physical space. Their system also tracked the remote user's hand using a depth camera and visualised a virtual hand in the shared environment for pointing.

Researchers also experimented using a HMD instead of a handheld device to collaborate in a shared 3D reconstructed environment captured with a depth camera [9][10][27][43]. These systems typically supported sharing gesture cues by visualising remote user's hands, either captured with a depth camera [9][43] or represented by virtual hands controlled with handheld controllers [27]. Most recently, Piumsomboon *et al.* [36] proposed a full-body adaptive avatar for sharing gestures cues in MR remote collaboration systems.

While sharing 3D reconstructed scene for MR remote collaboration enables remote users to navigate in the shared environment with full 6 degrees of freedom, this approach usually has limitations under dynamically changing scenes and the visual quality of the 3D reconstructed model is typically worse than a video image. As an alternative, researchers also investigated sharing 360 panorama to provide the freedom of looking around to the remote user in MR remote collaboration, although it is limited to changing viewing direction but not position. One of the early examples is a work by Billinghurst *et al.* [4] which overlaid a pointer or drawings on a shared static panorama image. Recently, Müller *et al.* [32] and Ryskeldiev *et al.* [38] investigated sharing a live 2D video insert into a static panorama image on handheld mobile devices. Their systems also overlaid visual cues (drawn annotation) onto the shared panorama image. However, to our best knowledge there hasn't been much work on overlaying visual cues on shared live 360 panorama video with both users using wearable displays.

Compared to these previous works, our research focuses on enriching shared live 360 panorama video with MR non-verbal communication cues. The main novel aspect of our system is that it allows sharing of hand gestures and view awareness cues, enabling richer two-way communication between the users both using wearable displays. Similar to our earlier prototype [20], most recently, Cai *et al.* [5] demonstrated a system showing remote user's avatar (virtual head and hands) in a shared live 360 panorama. Compared to our work which shares an egocentric viewpoint, their work visualises the remote user's avatar from third person's point of view as the 360 camera is mounted on a backpack of the local user.

To our best knowledge, no formal user study was conducted with a shared live 360 panorama based MR remote collaboration system that shares visual non-verbal communication cues. Tang *et al.* [42] ran an observational user study to investigate the benefits and problems of live 360 panorama based video conferencing, yet no MR visual cues were used. View independence has been actively investigated in MR remote collaboration systems sharing live 2D video [7][26] or sharing 3D reconstructed scenes [9][13][41], yet not much with sharing live 360 panorama as in our study. The closest prior work is by Kasahara *et al.* [22] in which they focused on how does stabilizing the 360 view reduces motion sickness of the remote user. Their work also included an observational study on communication behavior but without direct comparison between dependent and independent views, nor the person sharing the live 360 video received back any visual communication cue.

3 SHAREDSPHERE: A MR REMOTE COLLABORATION SYSTEM

Figure 1 shows an overview of the SharedSphere system. On the left, the host user is sharing his/her experience while wearing a 360 panorama camera mounted on a see-through Augmented Reality (AR) HMD. The 360 panorama camera captures the surroundings and shares it with the guest user (on the right of the figure). The guest user wears a VR HMD to watch the shared live panorama scene. The system also captures and shares the guest user's hand gestures as a non-verbal communication cue, and displays it on the see-through HMD the host user is wearing.

In this section we explain the main features of the system: view independence and MR visual communication cues. The overall goal of the system design is to help the remote guest user feel that they are in the same space as the local host user through sharing live 360 panorama video, and to support natural non-verbal communication through visualising MR visual cues. The current system significantly expands on an earlier prototype [20], such as adding a halo to the virtual hand and porting to the HoloLens which provides improved head pose tracking with less drift and slightly wider field of view (35 degrees compared to 23 degrees in the previous work). In the rest of the section we give a detailed overview of the current system for understating the user study setup.

3.1 View Independence

One of the main benefits of sharing 360 video is that the viewer can have an independent view from the person streaming the video. They can freely look around at different parts of the video streamed from the camera on the host user's head. However, as the 360 camera is physically mounted on the host user's head, the orientation of the video depends on the host user's head motion. When the host user looks down (as shown in Figure 2 (b)), the guest user's view also changes according to the host user's head motion. This also occurs when the host user is looking left or right which results in the shared scene turning around in the guest user's view, causing disorientation and restricting the guest user's viewing freedom.

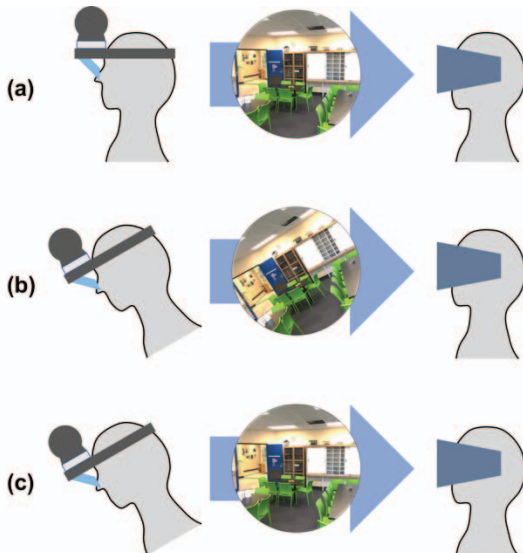


Figure 2: View independence: (a) the shared panorama is properly oriented when the local host (left) user's head is posed upright; (b) dependent view – the shared panorama rotates as the local host turns his head; (c) independent view – the shared panorama is shown at fixed orientation to the remote guest (on the right), independent from the local host's head rotation.

To overcome this problem, the system tracks the host user's head orientation with a tracking sensor and counter-rotates the panorama, so that the world view shown on the guest user's side remains at a fixed orientation independent from the host's head motion (see Figure 2 (c)). In case the guest user wants to follow the host user's view direction, the guest user can choose different viewing modes. For example, fixing their view to match the host user's view, or rotating freely only in the pitch direction so that the panorama scene will remain level, yet the heading will be synchronised with the host user's view direction.

3.2 Mixed Reality Visual Communication Cues

3.2.1 View awareness cues

The guest user can look around the shared scene independently from the host user's view direction, so it becomes difficult for the guest user to tell where the host user is looking (and vice versa for the host user). To help the users understand each other's view direction, the system represents a *View frame* of the other user with a coloured rectangle (see Figure 3 (a) and (c)). This shows the guest user exactly what the host user is seeing and vice versa.

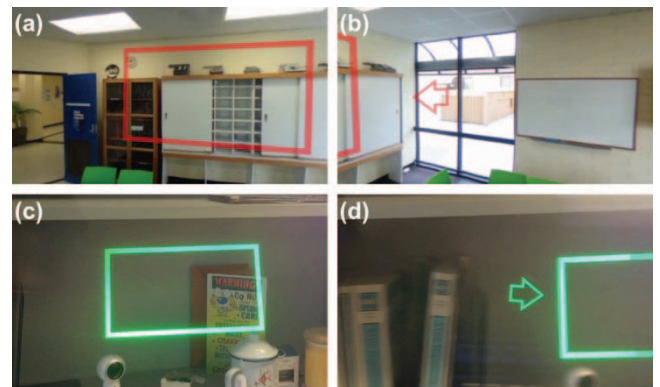


Figure 3: View awareness cues: (a) a *View frame* indicating the host user's view in the guest user's view, (b) a *View arrow* in the guest user's view, (c) & (d) a *View frame* and a *View arrow* in the host user's view.

When the two users are looking in very different directions, the other user's *View frame* will be out of the view and can't be seen. In such cases, the system shows *View arrow*, a virtual arrow that points in the direction where the other user is looking (see Figure 3 (b) and (d)). When the user rotates their head, the size of the arrow gets smaller as the user's view gets closer to the other user's view, and it disappears when the *View frame* is fully visible.

3.2.2 Virtual hand

Our system also supports natural gesture communication. Gestures are one of the main non-verbal communication cues that help facilitate collaboration. In the SharedSphere system the guest user can easily see video of the local host user's hand gestures captured in the shared 360 panorama. On the other hand, to allow the host user to see the guest user's hand gesture, the system shows a virtual representation of the guest user's hand on the host user's see-through HMD (see Figure 4). The virtual hand is animated based on the guest user's hand motion captured using a hand tracking sensor. During an initial demonstration we noticed various hand gestures made by remote guest users. They included deictic gestures, such as pointing at objects or direction, presenting or introducing objects with an open hand, as well as symbolic (e.g., indicating numbers or showing actions), and social gestures (e.g., waving hands or thumbs up for expressing preference).

When the virtual hand is out of the field of view, it becomes hard for the host user to recognize or follow the guest user's virtual hand. To overcome this problem, we took an approach similar to the concept of Halo [3], and added a glowing effect around the virtual hand. The size of the glow increases as the virtual hand moves away from the view frustum so that at least a part of the glow will be still visible to the user. The glow was visible as long as the system can track the guest user's hand, regardless of the two users facing opposite directions. From pilot trials we found it is better to show the glow even when the virtual hand is visible, as turning it off broke its connection with the virtual hand making it less obvious how it is related to the hand position. A preliminary test was held to verify the usefulness of the halo with four people trying to find and tell the pose (paper, scissors, rock) of a virtual hand randomly held around the user's view. Each user having several trials with and without halo, we found users taking significantly ($t(62)=-5.14, p<.001$) less amount of time with halo ($M=3.2$ seconds, $SD=1.1$; without halo $M=6.0$, $SD=2.8$). Without halo, there were also few cases where people took extremely longer than other trials (more than 20 seconds) that we had to exclude them as an outlier. Note that unlike in the test where users purposely looked for the virtual hands, in a real world task, it is very unlikely that a user will notice the virtual hands placed outside the field of view without being told.

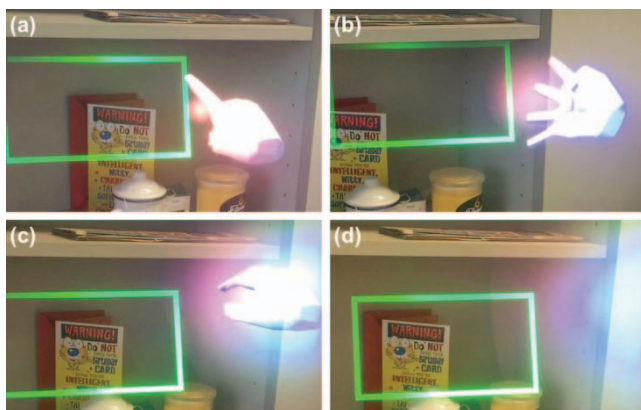


Figure 4: A *Virtual hand* and its halo in the host user's view: (a) deictic gesture, (b) symbolic gesture indicating number four, (c) & (d) halo enlarged as the virtual hand moves out of the view.

4 IMPLEMENTATION

The prototype system is built with AR and VR HMDs for the local host and remote guest users, correspondingly. The local host user wears a Microsoft HoloLens [30] optical see-through AR HMD, while the remote guest user uses an Oculus Rift [35] DK2 VR HMD driven by a PC (Intel Core i7 6700HQ 2.6GHz CPU, 16GB DDR4 RAM, NVIDIA GeForce GTX 1070 GPU) running the Microsoft Windows 10 operating system. The HoloLens and the VR PC communicated over a Wi-Fi connection. To capture the surroundings of the host user, a Ricoh Theta S [37] 360 panorama camera is mounted on the front slide of the HoloLens. The Ricoh Theta S is capable of live streaming stitched 360 panorama video over a USB connection with 1920 x 1080 resolution at 15 frames per second, or 640 x 360 at 10 fps over Wifi (we used a USB connection in the user study). A Leap Motion [28] hand gesture sensor is mounted in front of the VR HMD for capturing the guest user's hand motions which is shared to the host user.

The software was developed using the Unity 3D [44] game engine (v5.5.1f1). To visualise the shared environment on the guest user's side, live 360 panorama video was texture mapped

onto a sphere centered at the guest user's virtual camera. The orientation of the sphere was decided based on the view mode and the head orientation of the users shared with each other over the network. When using the dependent view, the sphere was oriented fixed to the guest user's head, while in the independent view it was following the local host user's head orientation. The system also supported combinations, such as only the heading (yaw) fixed to the guest user's head orientation but the pitch and roll updated based on the host user's head to level the panorama video.

The MR visual cues were overlaid on the panorama sphere for the guest user, and visualised on the optical see-through display for the host user. The view awareness cues were updated based on the shared head orientations of the two users, while the *Virtual hand* was animated based on the hand tracking data. The visualisation of the MR cues was calibrated to compensate for the offset between the panorama camera lens and the see-through display on the HoloLens. In our current implementation, we performed the calibration assuming the focus of the users was at about 1.2 metre of distance which matched well with the experimental setup for the user study. To improve the accuracy of the MR visualisation in future revisions, we plan to use the depth sensors on the HoloLens to update the calibration in real-time.

Visualisation of the MR cues also took the field of view of the displays into account. While the size of the *View frame* shown to the remote guest was matched to the field of view of the HoloLens display, matching the size of the remote guest's *View frame* shown on the HoloLens to the field of view of the VR HMD makes hard to view it on the HoloLens with narrow field of view (approximately 35 degrees) rendering the cue unusable. To overcome this problem, we traded the correct size of the *View frame* for its function of indicating the view direction. In our implementation, the size of the *View frame* representing the guest user was set to approximately 50% of field of view of the HoloLens display. While the *View frame* was not accurately representing the field of view of the remote guest, it did indicate the viewing direction of the remote guest. The Virtual hand also had a similar problem. While it was visualised at full size in the VR HMD, the Virtual hand was scaled down to be more easily viewed on the HoloLens, yet its position and orientation was preserved to make sure the deictic gestures would be kept valid.

5 USER STUDY

5.1 Experimental Design

We conducted a user study with the main goal of investigating the benefit of sharing live 360 panorama video in a MR remote collaboration system. The main research question of the user study was "*How would providing view independence through sharing live 360 panorama video affect the remote collaboration?*" As the main independent variable we chose view independence with two levels: (1) dependent and (2) independent views. With the independent view, the remote guest user can freely control his view direction by turning his head, while in the dependent view, his view direction is fixed to the local host user's view direction. We note that while our prototype system provided more options of viewing modes (e.g., synchronising only the heading) as described in 3.1, we limited it to the two modes (dependent and independent views) for the experiment. Also note that the physical field of view was identical between the two conditions. We postulated that the view independence will affect various aspects of the collaborative experience, including task performance, sense of being together (a.k.a. co-presence), mental task load, and discomfort.

The experiment used a within-subject design with two conditions: dependent view and independent view. Due to the nature of the collaborative task, another factor was the two roles

that participants could take: local host or remote guest. We let the participants experience both roles, hence they had four trials (two conditions x two roles) in total. The order of the conditions and roles were counter balanced between participants.

As an objective measure, we measured task performance with task completion time, recorded in a system log file. We used questionnaires as subjective measures, the Networked Mind Measure of Social Presence Questionnaire [16] and MEC Spatial Presence Questionnaire [45] for measuring sense of being together, and for measuring mental load we used Subjective Mental Effort Question [46]. We also used Simulator Sickness Questionnaire [23] to measure discomfort. After experiencing the two view conditions under each role, we asked users to choose which condition they preferred, and also give feedback on what they liked or disliked about each condition.

The secondary research question of the user study was “How would different MR visual cues overlaid on the shared live panorama help users collaborating in different roles?” We note that our interest was more on exploring and investigating how different visual cues are accepted by the users and what to improve, rather than strictly comparing each of the cues to prove one is better than another which would require each visual cue to be introduced as an independent variable, making the experiment overly complicated. Under this notion, we took an approach of mainly collecting feedback from the participants through rating and open questions.

5.2 Experimental Task and Setup

Similar to prior work [15][25], the experiment took place in a room divided into two sections, one side for the local user (Figure 5 (a) and (b)) and the other for the remote user (Figure 5 (c)). The users were not able to see each other but were able to talk to each other in a natural way. The local host user’s side had a desk and bookshelves where various objects were placed. The local host user wearing a HoloLens stood at the middle of his section being able to freely move around to perform the task. The remote user wearing an Oculus Rift DK2 HMD was seated on a turning chair. For the experiment, the 360 panorama video was shared over USB connection for better resolution (see section 4 for details).



Figure 5: Experimental setup: (a) 360 panorama picture of the local host user’s space, (b) setup on the local host user’s side, (c) setup on the remote guest user’s side.

The experimental task was a search task where both users should collaborate with each other to find a set of target objects in the task space and place them on the desk. The target objects were 54 sticky notes with different symbols printed on them (see Figure

6). Half of the symbols were a combination of common geometric shapes (e.g. a circle in a square, a plus sign in a circle, etc.), while the other half were more complicated shapes with various curves and lines. The target objects were randomly placed around the local host user’s task space. This could be on a wall, bookshelf, or a computer monitor. The target objects were relocated in each condition to prevent participants memorizing their locations.

The set of symbols to look for in each condition was randomised. The target symbol was revealed only to the remote guest user, appearing on the HMD screen one by one, mimicking a situation where a remote expert instructed a local worker to find something in the task space. The participants were free to talk to each other to discuss the target object to find, including describing its shape verbally or with hand gestures. They were allowed to develop their own strategies for completing the task.



Figure 6: Targets symbol (in the bright white square) shown in the guest user’s HMD.

5.3 Experimental Procedure

Participants participated in pairs. The experiment started with the participants signing a consent form, and answering demographic questions. Then the participants were shown the SharedSphere system and the study task. A training session followed where participants tried the interfaces provided in each role in the collaboration. During this, participants were given an explanation about view independence, the view indication cues, and gesture communication, as they tried out each feature. They also tried performing sample tasks with a couple of target objects. Once familiarised with one side of the system, they swapped roles to try the other side. At the end of the training participants were asked to answer SSQ to ensure they were not experiencing any serious motion sickness before performing the main experimental trials.

The experimental trials were divided into two sessions, one for each role in the collaboration. Participants were assigned to one of the roles randomly, then they swapped their roles after finishing the first session. Each session included two experimental trials in a different condition: dependent or independent view. The order of the condition was counter balanced between participants. After finishing each session, participants were asked to answer a session questionnaire which included ranking the conditions based on their preference, choosing what system feature they liked, and qualitative feedback on how to improve the system.

In each trial, participants performed the experimental task under a given condition. They were asked to find 8 target objects, one by one. The target object location was shuffled in each condition to prevent memorizing their position. After each trial, participants answered a questionnaire with subjective ratings items on the given condition. The experiment took an hour on average.

6 RESULTS

In this section, we report on the statistical analysis ($\alpha=.05$, unless noted otherwise) of the user study results, and summarise the qualitative feedback collected from the participants.

6.1 Participants

We recruited 12 participants (9 male, 3 female) with their ages ranging from 22 to 45 years old ($Mean=31.7$, $Standard\ Deviation=7.3$). Most pairs of the participants knew each other well, with half of them stating they knew each other more than a year, only one pair stated they met about a week ago. They described their relationship with their pair as colleagues (50%), friends (33.3%), or family (16.7%).

Except for one person, participants had been using video conferencing at least a few times a year. Two thirds of them were using it at least a few times a month. Participants were mostly familiar with AR or VR interfaces, with rating 4 or higher on a 7-point Likert item (1: novice ~ 7: expert) asking how much familiar they are with AR/VR interfaces.

6.2 Task Completion Time

Participants took slightly longer time to complete the task under the independent view condition ($M=234.0$ seconds, $SD=73.6$) compared to the dependent view condition ($M=215.8$, $SD=49.1$). A Shapiro-Wilk test indicated that the independent view condition was not normally distributed ($p=.023$), so a Wilcoxon Signed Rank test was used and found no significant difference between the two conditions ($Z=1.16$, $p=.246$).

6.3 Subjective Rating Questionnaires

6.3.1 Co-Presence

We used the co-presence subscale of the Networked Mind Measure of Social Presence Questionnaire [16] which consisted of six rating items on a 7-point Likert scale (1: strongly disagree ~ 7: strongly agree). Figure 7 shows the average rating results of each condition. Participants in the remote guest user role gave a higher rating to the independent view condition ($M=5.514$, $SD=0.830$) compared to the dependent view condition ($M=4.972$, $SD=1.197$). A similar trend was found with the participants in the local host user role (dependent view: $M=5.125$, $SD=1.236$; independent view: $M=5.625$, $SD=0.669$). Factorial analysis of the results using a two-way repeated measures ANOVA revealed there was a significant effect of view independence ($F(1,11)=5.202$, $p=.044$), while no significant difference was found between the roles ($F(1,11)=0.173$, $p=.686$) and the interaction effect was not statistically significant ($F(1,11)=0.018$, $p=.895$).

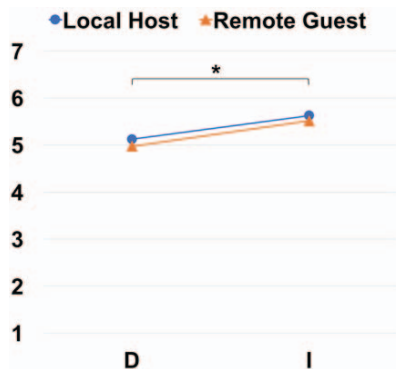


Figure 7: Results of co-presence questionnaire (7-point Likert scale from 1 to 7, the higher the better; D: dependent View, I: independent View; *: statistically significant).

6.3.2 Spatial Presence

To investigate if the conditions affected the participants' sense of being at a remote location, we used questions from the Self Location subscale of the MEC Spatial Presence Questionnaire [45] which consisted of four 5-point Likert scale rating items (1: fully disagree ~ 5: fully agree). As the local host user does not experience being in a remote place, the participants were asked to answer these questions only when they were in the role of the remote guest user. Results showed a trend of participants giving a slightly higher rating in the independent view condition ($M=3.73$, $SD=0.71$) than in the dependent view condition ($M=3.23$, $SD=0.80$), yet the difference was not statistically significant ($t(11)=-2.15$, $p=.0547$).

6.3.3 Mental Effort

To compare the participants' mental effort in each condition, we used Subjective Mental Effort Question [46] which is answered on a rating scale ranging from 0 (not at all hard to do) to 150 (tremendously hard to do). As shown in Figure 8, participants in the role of a local host user gave an average rating of "A bit hard to do" level to both of the view conditions (dependent view: $M=26.4$, $SD=14.4$; independent view: $M=22.7$, $SD=18.5$). As a remote guest user, participants gave average rating slightly above "Fairly hard to do" level (dependent view: $M=44.3$, $SD=30.6$; independent view: $M=40.0$, $SD=28.8$). A two-way repeated measure ANOVA ($\alpha=.05$) indicated that the remote guest users felt significantly more demand in mental effort to complete the experimental task ($F(1,11)=7.867$, $p=0.017$), while the independent view only marginally reduced the ratings ($F(1,11)=0.819$, $p=0.385$). There was no significant interaction effect between the two factors ($F(1,11)=0.014$, $p=0.907$).

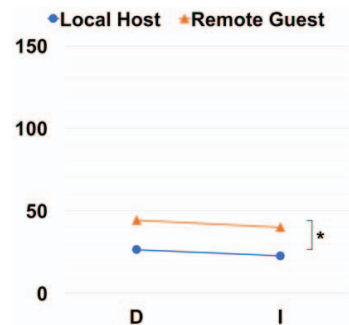


Figure 8: Results of subjective mental effort question (rating scale from 0 to 150; the higher the more mental effort demanded; D: dependent View, I: independent View; *: statistically significant).

6.3.4 Simulator Sickness

Figure 9 shows the results of Simulator Sickness Questionnaire [23] which comprises of fourteen symptoms rated on a Likert scale of (0: none ~ 4: severe). The rating items are aggregated into three subscales (nausea, oculomotor, and disorientation) by taking a sum of seven items in each subscale. While the original questionnaire had a weighting factor for each subscale, for simplicity and better understanding of the results, we report on the values without weights which results in each subscale ranging from 0 (none) to 21 (severe). The statistical analyses on the weighted results also lead to similar conclusions, hence here we only report on the unweighted results.

Overall, the participants reported only mild symptoms in both conditions across all subscales with average ratings not being higher than 7 in most cases. As some of the data set was found not normally distributed, we used Wilcoxon Signed Rank tests and found the overall rating was statistically significantly below average (10) in both the dependent ($Z=-2.41$, $p=.016$) and

independent ($Z=-1.98, p=.048$) view conditions. No statistically significant difference was found between the dependent and independent view conditions in the overall value ($Z=-1.33, p=.184$), nor in any of the subscales (nausea: $Z=-0.99, p=.322$; oculomotor: $Z=-1.35, p=.177$; disorientation: $Z=-1.01, p<.05$).

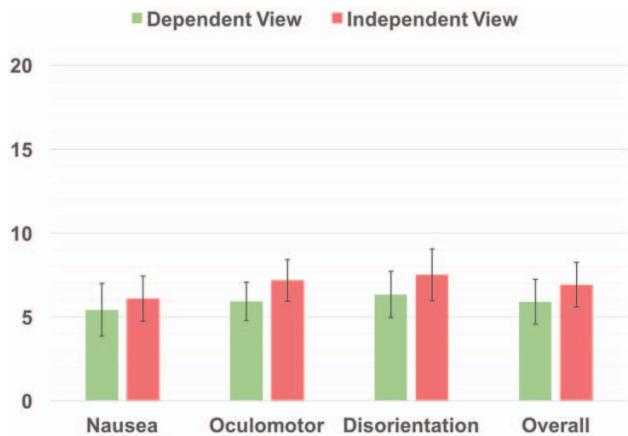


Figure 9: Results of simulator sickness questionnaire (0: none ~ 21: severe; Error bars represent standard errors).

6.4 Preference on Views

6.4.1 Remote guest user

From the remote guest’s point of view, more than half of the participants (7 out of 12) preferred the Independent view, while three participants preferred the Dependent view, and the rest (2 out of 12) showed no preference (see Figure 10). A one-sample Kolmogorov-Smirnov test showed the distribution of the choices was significantly different from evenly distributed random choices ($D_{max}=0.499, p<.01$).

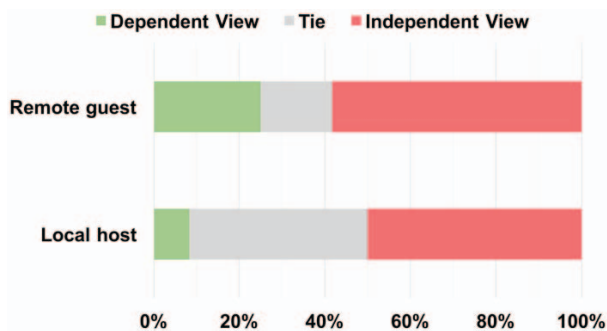


Figure 10: Preference between dependent and independent views.

Those who preferred the independent view explained that one of its benefits was being able to look around freely and work on the task independently. Participants mentioned, “I was able to look to where I wanted to.” “I was able to help my partner look and find things on my side as well.” “I could move my head and see my own view rather than keeping my head in one place and watching my partners view.” One of the participants valued the benefit of having the control of the view even when being aware of the problem of jittery motion in the independent view when the local host move quickly, mentioning, “I like being able to look where I want to, even though it was more choppy.” In contrast, some participants mentioned the independent view was causing less motion sickness than the dependent view, saying “I could freely move to actively help find the object, it also made me feel less sick than the dependent view.” “It caused less dizziness and both users could complement each other in the finding task.”

Those who preferred the dependent view were mostly concerned with the visual discomfort which happened when the local host user moved fast and the view became jittery due to leveling the view got out of sync. They mentioned, “I got very dizzy in the independent view,” “(dependent view) can decrease the discomfort I got.” One of the participant also mentioned directing the local user’s movement being easy, saying, “It was easier to just direct my partner, even if it might have been easier to look for it myself.”

Those who had no preference between the conditions mentioned that they both had pros and cons and so would better to freely switch between them. They explained, “I don’t feel as sick in the dependent view but I also like to be view independent.” “It would be good to switch between the independent and dependent view. When you just want to watch, the dependent view is better but when you want to look around and guide the local user in a more hands-on approach, the independent view is better.”

6.4.2 Local host user

From the local host’s perspective, half of the participants preferred the independent view condition (6 out of 12), while the rest mostly stated no preference between the two conditions (5 out of 12), and only one participant preferred the dependent view condition (see Figure 10). A one-sample Kolmogorov-Smirnov test showed the distribution of the choices was significantly different from random choices ($D_{max}=0.583, p<.01$).

Those who preferred the independent view explained, “I found my partner was able to communicate with me better,” “My partner can look around and help me find the sign as well,” “I can let him find the target instead of me.” “my partner can also look around, which makes it like two people doing the task.”

Among those having no preference, two mentioned both conditions had pros and cons, “I think both views have strengths depending on the task,” “I do enjoy being more involved in the game (collaboration) for a dependent view condition but I feel that it is faster and easier for my partner to do in the independent view.” Others mentioned that they didn’t felt much difference.

One participant preferring the dependent view explained he could be more focused on the task in this condition as the remote partner was behaving passive, saying, “I could focus on finding the symbols (by) myself and it was a lot quicker.”

6.5 Usefulness of Visual Cues

At the end of each session, we asked participants about how much the three visual cues (*View frame*, *View arrow*, and *Virtual hand*) were useful for each role in collaboration. The answers were given with rating on a 7-point Likert scale item (1: very useless ~ 7: very useful). Figure 11 summarises the results.

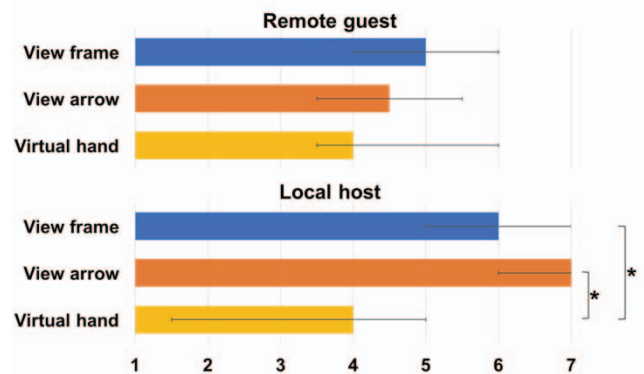


Figure 11: Results of rating on how much each visual cue was useful (1: very useless ~ 7: very useful; Median with Inter-Quartile Range in error bars; *: statistically significant difference).

As a remote guest, participants perceived the cues being marginally useful by rating close to the neutral value of 4 (*View frame*: Median=5, Inter-Quartile Range=[4-6]; *View arrow*: Md=4.5, IQR=[3.5-5.5]; *Virtual hand*: Md=4, IQR=[3.5-6]). Comparing the ratings with the neutral value using one-sample Wilcoxon Signed Rank test showed none of the cues has been rated significantly different from the neutral value. Also, no statistically significant difference was found between the cues with Friedman test ($\chi^2(2)=1.54, p=.463$).

Participants perceived the visual cues being more useful for the local host user. The *View frame* and *View arrow* cues were both rated significantly higher than the neutral value (*View frame*: Md=6, IQR=[5-7], Z=2.91, p=0.004; *View arrow*: Md=7, IQR=[6-7], Z=3.04, p=0.002), while the *Virtual hand* cue was perceived as not very useful (Md=4, IQR=[1.5-5]). A Friedman test indicated there was a significant difference between the cues on how useful each visual cue was for the local host user ($\chi^2(2)=14.63, p=.001$). Post hoc tests using Wilcoxon Signed Rank tests with Bonferroni correction ($\alpha=.0167$) showed that participants felt the *View arrow* cue significantly more useful than the *Virtual hand* (Z=2.91, p=.004), and also the *View frame* cue more useful than the *Virtual hand* (Z=2.88, p=.004), while there was no significant difference between the *View arrow* and the *View frame* cues.

6.6 Preference on Visual Cues

The participants were asked to choose a visual cue they liked most for each role in the collaboration and explain their preference. As a remote guest user, half of the participants (6 out of 12) chose *View frame* as the most preferred visual cue, while three chose *View arrow*, two chose *Virtual hand*, and one participant expressed no preference (see Figure 12). A one-sample Kolmogorov-Smirnov test showed the distribution of the choices was significantly different from evenly distributed random choices ($D_{max}=0.542, p<.01$).

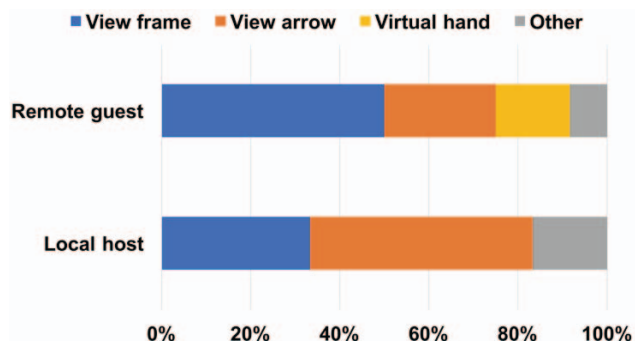


Figure 12: Preference on visual cues.

Those who chose *View frame* explained its usefulness as “I could identify where my partner was looking at,” “The frame rectangle sometimes served as a pointer. We could agree on the right object without pointing with our hands. My partner seemed to find this easier than pointing with his finger,” “I could see where my partner was looking so we could talk about the space we were both focusing on.”

Those who chose *View arrow* mentioned, “It is very useful into pointing where your partner is looking,” “(I) can follow partner,” “Because of the arrow, I can know where he is looking at.”

Two of the participants who found *Virtual hand* useful for the remote guest user explained their choice by mentioning, “I feel like I can catch my partner attention and indicate to her exactly,” “It allows me to provide more feedback to the local user.”

As a local host user, more participants preferred the *View arrow* cue with half of the participants (6 out of 12) choosing it, followed by the *View frame* (4 out of 12). No participant chose

Virtual hand, while one participant liked the “*Viewframe* combined with the *Virtual arrow*,” and another participant mentioned all of the cues were equal. The choices was significantly different from evenly distributed random choices ($D_{max}=0.583, p<.01$).

Those who chose the *View arrow* cue for the local host user explained the reason as being useful for finding where their partner was looking at even when they were looking away from each other. They mentioned, “it helps locating where my partner is looking at,” “I could understand where the partner look at,” “It was nice to direct me towards where my partner was looking,” “I always know where he is looking,” “view arrow can help me to find what my partner was looking at.”

Those who chose *View frame* mentioned “I could see where my partner was looking at as most of the time the communication was related to the objects within his view direction,” “The frame was enough information to identify a symbol, so the hand wasn't really needed. I didn't use the arrow since I usually found the symbol first, so the partner would adjust to see where I was looking.”

6.7 Suggested System Improvements

Participants were asked about what to improve in the current system in an open-ended question. On the local host user's side, the system was relatively well accepted with half of the participants mentioning that there was nothing much, suggesting the interface was working fairly well. Still there were useful suggestions made. Regarding hardware setup, a participant suggested to “make the HoloLens lighter on the front” as the 360 camera mounted on the HoloLens broke the weight balance. Another participant suggested using a better quality camera which is more to benefit on the remote guest side. There were also suggestions regarding improving visual communication cues. Two of the participants mentioned to improve the *Hand gesture* cue to make it more useful. Some participants also suggested adding more cues such as eye gaze, or the partner's face. Another interesting cue suggested was to indicate how the host user's movement is affecting the remote user, for example, indicating to slow down if moving too fast. He mentioned, “I don't have a sense of how my movements are affecting the other person's view, for example, how sick am I making him when I move. Should I move more slowly?”

When participants were asked what to improve on remote guest side, the most prominent problem mentioned by seven people was to reduce jitter and make the view smoother. This happened when local host users were moving or turning their head quickly. One of the participants who had troubles with their partner moving fast mentioned, “I think the VR side was very sickening. I don't think I can work with this system for more than 5 minutes in its current stage. need to reduce physical head movement.” The problem of motion sickness was present with both the dependent and independent views, as in the dependent view the visual motion of the view was not matching the remote guest's head motion (usually staying still) at all, while in the independent view, even though the orientation of the panorama video was leveled, if the local host made fast head movement the system still lagged behind with leveling and stabilising the panorama video into correct orientation, causing a jittery effect. Participants mentioned, “higher frame rate would make a big difference,” “smooth independent view,” “make it less blurry when moving,” “Frame rate, smoothing,” “reduce the shake.”

Other suggestions included improving existing features, such as “lower the camera angle, as it is too high,” and making more use of “gesture.” There were also suggestions of new features to add. As it was understood that there were pros and cons between the dependent and independent views, two participants suggested to let the user switch between the two views. One of them suggested

this could work as “snapping to my partner’s view” as needed. Another interesting suggestion was to add visual cues for indicating users to know when “their gaze intersects with their partner” for improving sense of sharing focus with each other.

7 DISCUSSION

7.1 Dependent vs. Independent Views

Overall the results indicated that the independent view positively contributed to the overall collaborative experience which is similar to the findings from prior work with 2D video conferencing [7][26] and 3D reconstruction based remote collaboration [41]. Participants preferred the independent view more as over half of the participants expressed their preference towards the independent view, while less than 25% preferred the dependent view. Qualitative feedback explained the main benefit of the independent view as giving the remote user freedom to control his or her own view direction, which resulted in the local users feeling that their partner was able to communicate better and also work on the task independently and in parallel.

The results from the co-presence questionnaire were in line with the participants’ preference. The independent view was rated as having significantly higher co-presence which comprised of rating items on the participants noticing and sensing each other’s presence, and paying attention to each other [16]. In the dependent view, most of the visual cues (except the *Virtual hand*) are not so useful for noticing each other’s activities. In contrast, with the independent view, the *View frame* and *View arrow* actively indicate changes in the partner’s view direction helping the users to be more aware of each other’s activities and focus, which may have contributed to improving the sense of being together. While not strong enough to reach statistical significance, the results of the spatial presence questionnaire also showed a supportive trend of remote guest users having a greater feeling of being ‘shifted’ to a remote environment (where the host user is located) with the independent view which could have contributed to co-presence.

Regarding the effect of view independency on task performance, no statistical evidence was found which is different from the findings of prior works [7][26][41]. From observations, we noticed that while the participants were able work in parallel and save time to find a target, it also took time for collating views compared to the dependent view in which the two users always have the same view hence no need for asking each other to pay attention to what he or she is looking at. We also noticed that the task performance was strongly influenced by collaboration styles especially with the independent view. In some cases, remote guest users showed passive behaviour of only following the local user’s view and confirming what the local host has found, while on the other extreme remote guest users tried to perform the task on their own. Results of the subjective mental effort questionnaire also indicate that task performance and load might be more dependent on the type of tasks and different roles played in collaboration that would need further investigation in the future.

The results from simulator sickness questionnaire indicated participants using the remote guest interface had mild symptoms of discomfort. While there is a need for further investigation with longer term usage, yet we note that no significant difference was found between the dependent and independent view conditions. In line with this result, qualitative feedback given by the participants regarding their preference also showed mixed opinions. The problem of motion sickness was present in both dependent and independent views, as neither of them give full control of the view to the remote guest user. In the dependent view, the remote guest users head motion mismatches the motions of the video view all the time. On the other hand, while the independent view allows the guest users to control their own view direction, still the

translational motion is dependent on the local host’s movement. This suggests motion sickness would be also largely dependent on the type of tasks and motions involved in the scene, as it was reported in prior work [22] which found image stabilisation significantly reducing motion sickness in general but not always.

7.2 Visual Cues

We note that verbal communication still remained as the main communication method, and the introduction of visual cues is to complement, but not to replace nor compete with talking. As users would not have traded off verbal communication for a visual cue, our study focused on comparing visual cues among themselves.

The two visual cues indicating the partner’s view, *View frame* and *View arrow*, were well received by the participants. They rated the *View frame* and *View arrow* as being more useful than the *Virtual hand*, especially in the local host user’s role. This was also reflected in their choice of preferred visual cue with the majority of the participants choosing either the *View frame* or the *View arrow* over the *Virtual hand*. We note that the experimental task being a collaborative visual search task, understanding each other’s focus is very important, hence the two view related cues would have received more preference over the *Virtual hand*.

While the *Virtual hand* could be useful for pointing to identify targets [40], *View frame* also worked as a pointing method which rendered the use of hand gestures not very necessary for performing the experimental tasks. We observed that participants tended to verbally describe the target symbols rather than using hand gestures, affirming verbal communication being the main modality. Despite half of the symbols were more complex shapes, rather than trying to give detailed descriptions using hand gestures, remote guest participants tended to just briefly describe certain features of the shape and simply look for the shape through trial and error. We note that the *Virtual hand* would not have been enough for describing complex shapes in the air. It would be interesting to add a function to the system which allows users to draw virtual annotations as in previous work based on 2D video conferencing [25][14] or static panorama [4][38], and further investigate how this helps communication between the users sharing live 360 panorama.

While none of the local host users chose the *Virtual hand* as the most preferred cue, some remote guests did. We postulate that the virtual hand was an additional method for the remote guest users to express themselves, while for the local users, knowing the focus or view of the remote guest user would have been enough for performing the task. In addition, the narrow field of view of the host user’s display still limited the visibility of the virtual hand, despite having halo. While the hand gestures were not very useful in the experimental task of visual searching, different types of collaborative tasks could take benefit of richer hand gestures, such as when needing to explain how to manipulate objects, as Higuch *et al.* [17] also suggested. Hand gestures could be also useful as non-verbal communication cue in social interactions. During the user study, we observed some participants moving their hands while talking without necessarily raising them for the system to recognise and show them as a virtual hand. We postulate that capturing and sharing such minor hand gestures could enable the users to better communicate with each other. This would need further investigation in the future.

It is also notable that the *View arrow* cue was more preferred by the local host users than the remote guests. Local users preferred the *View arrow* over the *View frame*, while it was the other way around in the remote guests’ case. Also, the participants gave higher ratings on the usefulness of the *View arrow* when working as a local host than as a remote guest ($Z=2.78$, $p=.0054$). This could have been partly due to the narrow field of view of the AR display used by the local host users. The *View frame* went out of

the local host user's field of view time to time, while the *View arrow* was always present within the view. In such a way, the *View arrow* was useful for 'finding' where their partner was looking at when the users were looking away from each other. In comparison, the remote guests were using a wide field of view VR HMD, hence it was easier to notice and find the *View frame*. We also note that the *View frame* indicated where exactly their partner is looking at, which could be more useful in wide field of view compared to the *View arrow* which only gave the direction to turn.

7.3 Limitations

While the user study helped evaluating the system in a controlled environment with an experimental task, it needs to be further tested in various real-world application scenarios and with longer term usage. Different tasks and work environments would involve a different amount of head movement of the local host user which could either increase or decrease the necessity of view independence. As suggested by the participants, there could be also certain cases where switching between the view modes is more beneficial. Likewise, different MR visual cues could be more useful in certain use cases. While the *Virtual hand* was perceived as not very useful in the collaborative visual search task, other use cases that need communicating complex hand motions (e.g. manipulating physical objects or equipment) or social gestures could certainly benefit from sharing hand gestures.

Another limitation of the user study is the relatively small sample size ($N=12$). While the number of participants was enough to derive statistically significant results, some of the measures (especially the spatial presence questionnaire) showed a p -value that is close to the significance level which could have turned out to be statistically significant given more samples. To avoid misrepresenting such cases by simply stating them as non-significant results, we tried to capture and report the trend as well as the p -values of the non-significant results. Future studies with a larger number of participants with more variants in their background and relationships would be beneficial to draw stronger and more generalisable conclusions.

The prototype system implementation also had certain limitations that would need improvements in the future studies. The most obvious problem was the jittery motions due to off-sync panorama levelling and stabilisation. Using better 360 cameras with an integrated gyro sensor for image stabilisation (e.g., Insta 360 One [21] or Garmin VIRB360 [11]) could be a relatively easy way for improving the system, while adding computer vision based image stabilisation [22] could be another approach. Another limitation is that the MR cues are visualised assuming the user's focus being at a fixed distance. While this was not a huge problem in the user study setup, it is necessary to solve this issue for using the system in a real world scenario. We plan to solve this problem by using the depth sensor and spatial mapping features on the HoloLens. Supporting audio streaming as well as improving video streaming quality (e.g. latency, resolution, etc.) would be other necessary enhancements to further investigate if the system would be usable in real world scenarios where two users are located in further distance.

7.4 Design Implications

Here we summarise the lessons learnt as design implications for future research and development.

1) Use the independent view as default but provide an option to switch to the dependent view as needed. While the independent view is preferred and improves co-presence in MR collaboration, there are certain cases where users can benefit from switching to the dependent view. Therefore, giving an option to switch between the independent and dependent views would be helpful for the users.

2) View awareness cues are not only helpful but important to have. When sharing live 360 video using an independent view, without a view awareness cue, it becomes hard for the users to know where their partner is looking at. View awareness cues not only help users to understand each other's focus and be more aware of each other's activity, but also could be utilised as a pointer. Hence the design of a view frame could pursue such double purpose by not only marking the boundary of the view but also adding a cross mark or similar to the centre.

3) Provide screen space registered cues when using HMDs with a narrow field of view. While visual cues shown in the world space (e.g., *View frame*) are useful for indicating absolute position or direction, when they are out of the user's field of view it becomes useless. This is especially problematic when using a display with a narrow field of view which is still common in optical see-through HMDs. Screen space registered cues (e.g., *View arrow* and halo with the *Virtual hand*) can be helpful in such cases as they remain within the user's view, yet it should be carefully designed so it does not become too much of a clutter.

4) Virtual hands should be considered for more than pointing purpose. While one of its use cases is for making deictic gestures, there are other alternatives that can be better for pointing in terms of accuracy and speed (e.g., in the study the *View frame* was). Virtual hands should rather be used for making more complex gestures to express non-verbal communication cues (e.g. object manipulation or social interactions).

8 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this paper we reported on a user study evaluating SharedSphere, a remote collaboration system with shared live 360 panorama video and a Mixed Reality interface. The system allows for enriched two-way collaboration by supporting non-verbal communication cues, including view awareness cues and hand gestures. A user study was conducted to evaluate the benefits of providing an independent view with shared live 360 panorama video, and compare various visual cues overlaid onto the view. The results showed that participants preferred the independent view over the dependent view, as it improved their sense of being together with their partner. Participants reported that the main benefit of sharing a live 360 panorama video view was providing remote guest users with the freedom to control their own view, enabling them to work in parallel and independently from the local host. This not only helped the remote guest users, but the local hosts also reported improved communication as they collaborated with their partner. The results showed that the view awareness cues were very useful, while the hand gestures was not used much for the experimental task. Further evaluation of the system with different types of tasks would be beneficial to investigate how hand gestures could be used in different context.

In the future, we plan to further explore various real-world applications scenarios and conduct user studies to evaluate the system in a real world context. We also plan to improve the system to support richer non-verbal communication cues, such as better hand gesture support, drawing annotations, sharing facial expression and eye gaze tracking. Extending the concept to supporting one-to-many social sharing experiences would be another interesting direction for future investigation. We expect that conducting further studies with these system improvements would lead us to fuller understanding of how shared live 360 panorama video based MR collaborations could be useful for telecommunication and remote collaboration in the future.

REFERENCES

- [1] J. Amores, X. Benavides, and P. Maes. ShowMe: A Remote Collaboration System that Supports Immersive Gestural Communication. In *Proceedings of the 33rd Annual ACM*

Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15). ACM, New York, NY, USA, pp. 1343-1348, 2015.

- [2] S. Barathan, G. A. Lee, M. Billinghurst, and R. W. Lindeman. Sharing Gaze for Remote Instruction. In *Proceedings of ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pp. 177-184, 2017.
- [3] P. Baudisch and R. Rosenholtz. Halo: a technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*, pp. 481-488, 2003.
- [4] M. Billinghurst, A. Nassani, and C. Reichherzer. Social panoramas: using wearable computers to share experiences. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*, Article 25, 2014.
- [5] M. Cai, S. Masuko, and J. Tanaka. Gesture-based Mobile Communication System Providing Side-by-side Shopping Feeling. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion (IUI '18 Companion)*, ACM, New York, NY, USA, Article 2, pp. 1-2, 2018.
- [6] S. E. Chen. QuickTime® VR – An Image-Based Approach to Virtual Environment Navigation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques 29*, Annual Conference Series, pp. 29-38, 1995.
- [7] S. R. Fussell, L. D. Setlock, and R. E. Kraut. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'03)*, Ft. Lauderdale, Florida, USA, 05-10 April 2003. New York: ACM, pp. 513-520, 2003.
- [8] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, and E. Mauer. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction* 19, 3: 273-309, 2004.
- [9] L. Gao, H. Bai, G. Lee, and M. Billinghurst. An oriented point-cloud view for MR remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications on - SA '16*: 1-4, 2016.
- [10] L. Gao, H. Bai, T. Piumsomboon, G. A. Lee, R. W. Lindeman, and M. Billinghurst. Real-time Visual Representations for Mixed Reality Remote Collaboration. In *Proceedings of ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pp. 87-95, 2017.
- [11] Garmin VIRB360. <http://virb.garmin.com>
- [12] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12)*. ACM, New York, NY, USA, pp. 241-250, 2012.
- [13] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. 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, pp. 449-459, 2014.
- [14] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. In Touch with the Remote World: Remote Collaboration with Augmented Reality Drawings and Virtual Navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (VRST '14)*, pp. 197-205, 2014.
- [15] K. Gupta, G. A. Lee, and M. Billinghurst. Do you see what I see? the effect of gaze tracking on task space remote collaboration. *IEEE Transactions on Visualization and Computer Graphics* 22, 11: 2413-2422, 2016.
- [16] C. Harms and F. Biocca. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annual International Workshop: Presence 2004*: 246-251, 2004. Retrieved from <http://cogprints.org/7026/>
- [17] K. Higuchi, R. Yonetani, and Y. Sato. 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*: 5180-5190, 2016.
- [18] G. A. Lee, S. Kim, Y. Lee, A. Dey, T. Piumsomboon, M. Norman, and M. Billinghurst. [POSTER] Mutually Shared Gaze in Augmented Video Conference. In *Proceedings of 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 79-80, 2017.
- [19] G. A. Lee, S. Kim, Y. Lee, A. Dey, T. Piumsomboon, M. Norman, and M. Billinghurst. Improving Collaboration in Augmented Video Conference using Mutually Shared Gaze. In *Proceedings of ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pp. 197-204, 2017.
- [20] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications (SA '17)*, Article 14, pp. 1-4, 2017.
- [21] Insta360 ONE. <https://www.insta360.com/product/insta360-one>
- [22] S. Kasahara, S. Nagai, and J. Rekimoto. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE Transactions on Visualisation and Computer Graphics* 23, 3: 217-225, 2015.
- [23] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and Michael G. Lilienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3: 203-220, 1993.
- [24] S. Kim, G. A. Lee, N. Sakata, A. Dünser, E. Vartiainen, and M. Billinghurst. Study of augmented gesture communication cues and view sharing in remote collaboration. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality 2013 (ISMAR 2013)*, pp. 261-262, 2013.
- [25] S. Kim, G. Lee, N. Sakata, and M. Billinghurst. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality - Science and Technology 2014 (ISMAR 2014)*, pp. 83-92, 2014.
- [26] S. Kim, M. Billinghurst, and G. Lee. The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration. *Computer Supported Cooperative Work (CSCW)*, pp. 1-39, 2018.
- [27] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi. Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *Proceedings of 2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pp. 9-12, 2016.
- [28] Leap Motion. <https://www.leapmotion.com>
- [29] R. Masoni, F. Ferrise, M. Bordegoni, M. Gattullo, A. E. Uva, M. Fiorentino, E. Carrabba, and M. Di Donato. Supporting Remote Maintenance in Industry 4.0 through Augmented Reality. *Procedia Manufacturing*, 11, 1296-1302, 2017.
- [30] Microsoft HoloLens. <https://www.microsoft.com/en-us/hololens>
- [31] P. Milgram and F. Kishino. Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems* E77-D, 12: 1321-1329, 1994.
- [32] J. Müller, T. Langlotz and H. Regenbrecht. PanoVC: Pervasive telepresence using mobile phones. In *Proceedings of 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Sydney, NSW, pp. 1-10, 2016.
- [33] B. Nuernberger, K. Lien, T. Höllerer, M. Turk. Interpreting 2d gesture annotations in 3d augmented reality. In *Proceedings of 2016 IEEE Symposium on 3D User Interfaces (3DUI 2016)*, pp. 149-158, 2016.
- [34] B. Nuernberger, M. Turk, and T. Höllerer. 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, pp. 1-11, 2017.

- [35] Oculus Rift. <https://www.oculus.com/rift/>
- [36] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. 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, Paper 46, pp. 1-13, 2018.
- [37] Ricoh Theta S. <https://theta360.com/en/about/theta/s.html>
- [38] B. Ryskeldiev, M. Cohen, and J. Herder. Applying rotational tracking and photospherical imagery to immersive mobile telepresence and live video streaming groupware. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications (SA '17)*. ACM, New York, NY, USA, Article 5, pp. 1-2, 2017.
- [39] J. Sauro and J. Dumas. Comparison of Three One-Question, Post-Task Usability Questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1599–1608, 2009.
- [40] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci. 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, pp. 179-188, 2013.
- [41] M. Tait and M. Billinghurst. The Effect of View Independence in a Collaborative AR System. *Computer Supported Cooperative Work* 24, 6: 563-589, 2015.
- [42] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration in 360° Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, pp. 1327–1339, 2017.
- [43] F. Tecchia, L. Alem, and W. Huang. 3D Helping Hands : a Gesture Based MR System for Remote Collaboration. In *Proceedings of Virtual Reality Continuum and its Applications in Industry (VRCIAI)*, pp. 323–328, 2012.
- [44] Unity 3D. <http://unity3d.com>
- [45] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, F. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, C. Klimmt, J. Laarni, N. Ravaja, A. Sacau, T. Baumgartner, and P. Jäncke. *MEC Spatial Presence Questionnaire (MEC- SPQ): Short Documentation and Instructions for Application*, 2004.
- [46] F. R. Zijlstra. *Efficiency in work behaviour: A design approach for modern tools*. Delft University press, January 1993: 1–186, 1993.