A User Study on View-sharing Techniques for One-to-Many Mixed Reality Collaborations

Geonsun Lee*
Department of Computer Science & Engineering
Korea University

HyeongYeop Kang†
Department of Software Convergence
Kyung Hee University

JongMin Lee‡
Department of Computer Science & Engineering
Korea University

JungHyun Han§
Department of Computer Science & Engineering
Korea University

Abstract

In a one-to-many mixed reality collaboration environment, where multiple local users wearing AR headsets are supervised by a remote expert wearing a VR HMD, we evaluated three view-sharing techniques: 2D video, 360 video, and 3D model augmented with 2D video. Through a pilot test, the weaknesses of the techniques were identified, and additional features were integrated into them. Then, their performances were compared in two different collaboration scenarios based on search and assembling. In the first scenario, a local user performed both search and assembling. In the second scenario, two local users had dedicated roles, one for search and the other for assembling. The experiment results showed that the 3D model augmented with 2D video was time-efficient, usable, less demanding and most preferred in one-to-many mixed reality collaborations.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Collaborative and social computing—Collaborative and social computing theory, concepts and paradigms—Computer supported cooperative work

1 Introduction

A notable strength of virtual reality (VR) and augmented reality (AR) techniques is to enable distant users to collaborate with each other in immersive or augmented environments. There also exists an increasing interest in mixed reality (MR) collaborations. Nevertheless, it has been so far largely limited to *one-to-one* collaboration. Its typical scenario consists of a *local user* and a *remote expert*. With the aid of MR devices, the remote expert helps the local user work properly in the local workspace. Such a collaboration is needed when the remote expert has to give instructions to the local user who lacks knowledge and competence to carry out a task, or when the remote expert wishes to check the work progress and accordingly give advice to the local user.

Beyond such one-to-one collaboration, *one-to-many* or even *many-to-many* collaborations are also required in the real world. The practical uses may include training workers in manufacturing plants, training mechanics for industrial maintenance, managing workers in warehouses, and allocating and managing sales clerks in stores. In addition, plausible educational uses are teaching trainee doctors and medical students for practicing surgery or teaching students in science classes and laboratories.

This paper focuses on a one-to-many collaboration scenario, where the local users work in the same space under the supervision of a remote expert. In our study, each local user wears an AR headset and the remote expert wears a VR head-mounted display (HMD). The workspace captured by each AR headset's camera is transmitted to the VR HMD such that the expert can observe the progresses being made by the local users. In such a one-to-many MR collaboration scenario, the remote expert's instructional effectiveness would heavily depend on how the captured workspace is shared with the expert. This paper reports the results of evaluating three *view-sharing techniques* in respect of the remote expert.

The contributions of this paper can be listed as follows:

- Our work is the first earnest attempt to evaluate view-sharing techniques in a one-to-many collaborative MR environment.
- We propose an effective view-sharing technique, which combines live videos with a reconstructed 3D model and also allows to switch between different video display modes.
- Our experiment was made with two plausible collaboration scenarios. Valuable insights gained during the experiment and implementation guidelines based on the results are provided.

2 Related Work

There has been extensive research made in the field of MR collaborations throughout the years [7, 8]. A number of studies focused on a "remote expert collaboration" scenario where a user's workspace is captured and shared with a remotely located expert. In remote collaborations, how the remote expert views the workspace, or how the view is shared, is an important topic [19, 37]. In fact, this subject of "view sharing" was widely discussed in the area of VR collaborative environments [6, 27, 31, 34, 35, 46]. For example, Valin et al. [46] explored under what circumstances view sharing is sufficient for effective collaboration, and Chellali et al. [6] observed users' behavior on view sharing with different reference systems.

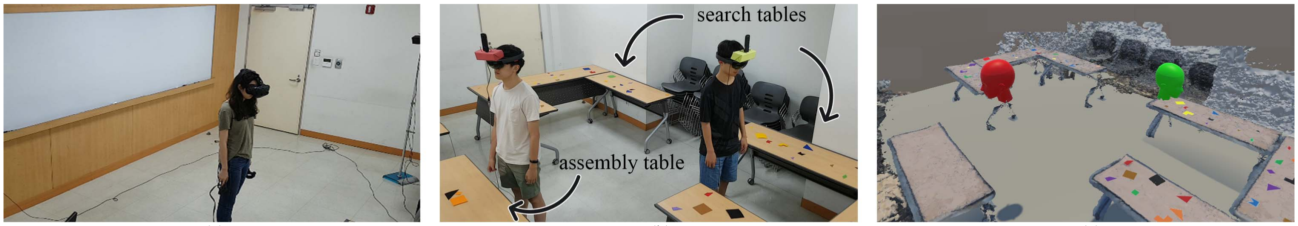
The most traditional way of view sharing in remote MR collaborations is to transmit a live video captured by a standard camera or one built in a hand-held display or an HMD. This was done either by sharing a first-person point of view video captured from a local user with a remote VR user [2, 20] or by providing an independent view by placing an additional camera in the workspace [9, 19].

To provide a more immersive view, 360 cameras were used for view sharing in remote collaborations. Piumsomboon et al. [29] proposed a system where a remote expert is immersed in a live 360 video and a local user wearing an AR HMD can physically manipulate the remote expert's view by moving the 360 camera. Kasahara et al. [17] used multiple cameras around a user's head in different angles to create a 360 spherical video for the remote VR user. Tang et al. [41] developed a system where a remote user views the 360 video captured by a 360 camera mounted on a user's backpack. Lee et al. [22] and Teo et al. [44] mounted a 360 camera on Microsoft HoloLens [24] to share the panorama video with a remote VR user.

Static 3D reconstruction is also popularly used in MR remote collaborations to virtually represent the physical environments [26, 36, 39]. With this method, the remote expert can freely investigate the local user's workspace. Recent studies used HMD as a way to view the reconstructed environment [10, 11, 21, 28, 30, 42]. Nevertheless, static 3D reconstruction limits the collaboration since

*e-mail: g_sun@korea.ac.kr
†e-mail: siamiz88@gmail.com
‡e-mail: leejm8304@gmail.com
§e-mail: jhan@korea.ac.kr

2642-5254/20/$31.00 ©2020 IEEE
DOI 10.1109/VR46266.2020.00054



(a) (b) (c)

Figure 1: The experimental room setup. (a) On one side of the room is the remote expert wearing a VR HMD. (b) On the other side of the room are two local users, each wearing an AR headset. (c) The 3D reconstructed scene where two local users are visualized as head-only avatars.

it does not capture the dynamic changes made in the real environment. Recently, there are increasing interests in exploring a live reconstructed 3D model in VR/AR environments. Although real-time 3D reconstruction has made progress [13, 38, 47], it is not easy to reconstruct a dynamic scene at real time. As a solution, Teo et al. [44] proposed to pre-reconstruct a static scene and augment a live video.

Teo et al. [44] was the first to compare the user experience between 360 live panorama and 3D reconstruction based MR remote collaboration systems. The comparison results indicated the method with 360 video performed better in task completion time and social presence. Teo et al. [43] merged the two methods to enhance users' collaboration experience. The study was conducted only in a one-to-one collaboration scenario.

While the aforementioned works focused on one remote user and one local user, there are collaboration cases where multiple users may be co-located and distributed. Mixed Presence Groupware (MPG) [40] defines such collaboration systems that support synchronous work in a shared workspace by collaborators that are both co-located and distributed. Tang et al. [40] observed two problems on MPG: The disparity in displays of heterogeneous devices and the disparity in users' perception of collaborators' presence. MPG was researched for collaborations based on tabletop applications [33, 45], heterogeneous large-scale displays [23], and VR systems [16]. Norman et al. [25] initiated the investigation of MPG using MR systems. The experiment involved two local users using MR HMDs with one remote user using a desktop-based AR interface. They examined how different role assignment affected the remote user's usage of visual communication cues.

To the best of our knowledge, there were no previous researches made to compare the view-sharing techniques of a remote expert in MPG. Our work addresses this issue in a one-to-many MR collaboration scenario where one remote expert has to instruct multiple co-located users. This is important since many real-world applications for MR collaborations involve more than two users [7].

3 Workspace Setup

In our one-to-many collaboration system, each local user is wearing an AR headset, a Microsoft HoloLens [24], on which a Ricoh Theta S 360 camera [32] is mounted. HoloLens' field of view is 34° and its camera resolution is 1408×792. The 360 camera captures 1920×1080 resolution panorama video. The remote expert is wearing a VR HMD, HTC Vive. Its field of view is 110° and the OLED panel has the resolution of 1080×1200 per eye.

Consider a one-to-many collaboration environment, where a local user looks for an object, locates it, and works on or with it under the supervision of a remote expert. A good example is an auto repair shop, where multiple mechanics work on a car. In order to investigate the view-sharing techniques in such an environment, we devised a simplified scenario with a tangram-like puzzle, where the local users look for flat polygons, called *tans*, to assemble a target configuration. The local users' actions are all directed by the remote expert.

Our experiment setup is shown in Figure 1. In a room, the remote expert (Figure 1-(a)) and the local users (Figure 1-(b)) are separated by a partition so that the users cannot see the expert. In this setup the expert can talk to the users without a microphone. The local users' workspace is approximately $4m \times 7m$. The tans were scattered over six tables that surrounded the local users, and a separate table was provided for assembling the target configuration with tans. They are called *search tables* and *assembly table*, respectively. The target configuration was known only to the expert, and the expert instructed one local user at a time to pick up a tan from the search tables or assemble the target configuration on the assembly table.

While there exist MR collaboration studies that utilize a desktop setup for the remote expert [25], we used a VR setup. Not only does it have an advantage of allowing the expert to watch the workspace in an immersive manner [22], but it also benefits the expert with intuitive spatial interactions, for example, simply turning the head to see the left is more convenient than using a keyboard or a mouse.

4 Pilot Test

Before the main experiment, we conducted a pilot test with three view-sharing techniques that were commonly used in the previous "one-to-one" MR collaboration systems. The goal of the pilot test was to identify the weaknesses of the view-sharing techniques, improve them based on participants' feedback, and also adapt them to fit to our "one-to-many" collaboration scenarios.

4.1 View-sharing Techniques

Three view-sharing techniques used in the pilot test are named *2D Video*, *360 Video*, and *3D Model*.

2D Video This is the simplest view-sharing technique [12, 20]. The live video stream from the local user's HoloLens camera is shared with the remote expert. Obviously, the video is not affected by the expert's head movement.

Using the VR controller, the expert can switch from a user's view to another's. The fade-in and fade-out effects are made when switching between local users to prevent the expert from feeling sick. The expert can also point at a specific object in the video to share the visual cue with the local user. Figure 2-(a) shows a user's view shared with the expert, where the blue line represents the expert's pointing direction. Assuming two local users, named A and B, they are color-coded as red and green, respectively. In Figure 2-(a), the video's frame is colored in red, which implies that it is A's view.

360 Video The stream of live panorama video is captured by the 360 camera mounted on HoloLens. It is shared with the remote expert. As was done in previous works [9, 19, 22, 29], we counter-rotate the panorama video when the HoloLens rotates. This ensures that the expert's view is not bound to the local user's head movement.

The local user's view direction is depicted with a color frame (e.g., the red rectangle in Figure 2-(b)), and therefore the expert can always identify the local user's attention. With respect to switching between users, pointing, and color coding, the discussions made for *2D Video* also apply to *360 Video*. For example, the red frame in Figure 2-(b) implies that it describes the view direction of A.

(a) (b) (c)

Figure 2: Three view-sharing techniques. (a) *2D Video*: The light blue line points at a dark blue tan. (b) *360 Video*: The red frame indicates A's attention. (c) *3D Model*: The avatars of A and B are in red and green, respectively.

3D Model If the workspace is reconstructed in 3D, the remote expert can freely walk through the reconstructed space (unlike in *2D Video* and *360 Video*) to direct the local users more effectively. Figure 1-(c) shows the reconstructed workspace, where each local user is visualized as a color-coded head-only avatar. (We will use the term, avatar, only for the local users, not for the remote expert.)

Despite the tremendous progress in real-time 3D reconstruction, however, there is still a long way to go to precisely reconstruct a dynamic scene at real time. To get around the limitation, Teo et al. [44] pre-reconstructed a static scene but displayed the dynamic scene captured along a user's view direction at a *live-video frame* attached to the front of the user's avatar. See Figure 2-(c), where the tilted rectangle perpendicular to an avatar's view direction shows a video frame. Let us call the video frame simply *billboard*.

Similar to Teo et al. [44], our workspace is pre-reconstructed using Agisoft PhotoScan [1]. It is then smoothed using Meshmixer [3]. The tans are pre-reconstructed together with the tables. The live video in the billboard is captured by HoloLens' camera, as is the case for *2D Video*. We call this view-sharing technique of combining the pre-reconstructed workspace and live video simply *3D Model*.

All avatars' movements were synchronized with the local users' using Microsoft Azure's Spatial Anchors [4] built in HoloLens. The expert, whose pose is also represented as a 6DOF head, is visible to *all* local users, and so is the expert's pointing line. In Figure 2-(c), for example, both the red and green avatars can see the blue line.

The remote expert's tracking volume is set with the Room-scale mode of HTC Vive. Its dimensions are about $3.1m \times 4.2m$. Since it is smaller than the local users' walkable space, which is approximately $3.5m \times 4.7m$, the expert's movement is visually scaled by 110%. It means that, for example, the remote expert will move $1.1m$ in the virtual world when moving $1m$ in the real world.

For the purpose of ensuring real-time communications between the expert and users, the live video captured by HoloLens (for both *2D Video* and *3D Model*) is down-sampled to the resolution of 704×396 to be sent to the expert at 15fps. Each 360 camera mounted on HoloLens is connected to a laptop using a $5m$ USB 3.0 extension cable, and the 1920×1080 resolution panorama video is streamed at 15fps. In all of *2D Video*, *360 Video*, and *3D Model*, the live videos and local users' positions are shared with the expert through WiFi and then are rendered using Unity game engine.

The video stream in *2D Video* will take the entire field of view of the VR HMD if the *full-screen mode* is turned on. If it is turned off, the video size will be reduced to match the resolution and reveal the black background. See Figure 2-(a). In the pilot test, we used four techniques in total: (1) *2D Video* with the full-screen mode on, (2) *2D Video* with the full-screen mode off, (3) *360 Video*, and (4) *3D Model*.

4.2 Participants and Procedure

In both the plot test and the main experiment, we employed two *actors* as local users. This is to minimize the performance gaps caused by individual differences [44,48]. We recruited eight volunteers (one female, seven males) for the remote expert, and a participant spent 15 minutes for each technique.

Eight tans were placed on the search tables. The participants were asked to freely instruct local users to take a tan from the search tables and assemble the target configuration in the assembly table. After each participant tried out all of four techniques, an open-ended interview was conducted. The questions focused on what modifications or additional features would be desired for the view-sharing techniques in order to improve the expert's instructional effectiveness.

4.3 Feedback

Out of eight participants, denoted as P1 through P8, five pointed out that the full-screen mode in *2D Video* made it rather difficult to understand the scene at a glance. P2 and P4 commented "The full screen is unnecessarily big and seems to bring less information than *360 Video*. I would prefer a smaller screen."

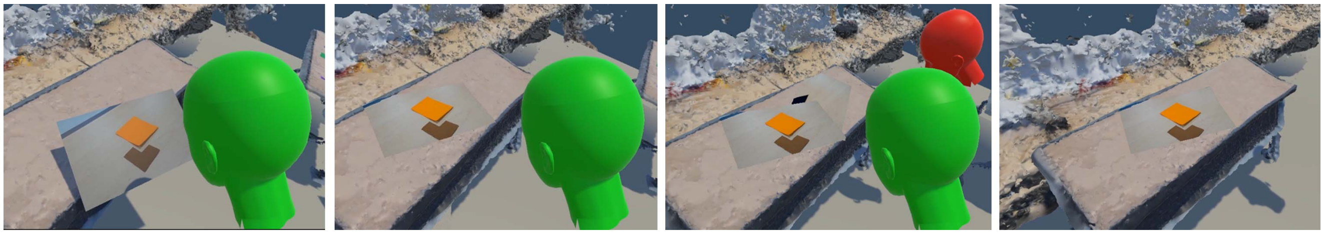
As for *360 Video*, seven participants complained about the image distortion caused by the video projected onto a sphere. Thus it was hard for them to identify the distant tans. P1, P2 and P7 added "If we can zoom in, the distant objects can be identified easily and the performance will be increased."

As for *3D Model*, the majority of the comments were twofold. First, seven participants complained that the pre-reconstructed 3D scene did not show the change made in the real workspace. Second, all of eight participants complained that the billboard was clearly visible only when their view direction was close to the avatar's and this feature significantly degraded the overall performance. Suppose that two avatars, A and B, work face to face. If the expert finishes advising A and wants to advise B, the expert has to go round in a half circle to observe the billboard located in front of B. P3 said "I always had to stick to one of the avatars to see its billboard." P7 complained "It was inconvenient to switch between the avatars." The billboard proved to be unsuitable for one-to-many collaboration scenarios.

4.4 Modifications and Improvements

Based on the feedback, we modified the view-sharing techniques for the main experiment:

- *2D Video*: The full-screen mode is disabled so that the video is displayed always with the black background, as shown in Figure 2-(a).
- *360 Video*: We add zoom-in and zoom-out functions. They are made along the expert's view direction and implemented by simply translating the virtual camera's position back and forth.
- *3D Model*: We add a function to switch between the video billboard and what we call *video projector*, which are illustrated in Figure 3-(a) and -(b), respectively. The latter represents the video projected as a texture onto the scene objects and is implemented using *projective texture mapping* in Unity. When projected onto the tables, it is visible from a wide range of view directions, freeing the expert from sticking to an avatar.



(a) (b) (c) (d)

Figure 3: Video display in *3D Model*. (a) Video billboard. (b) Video projector. (c) Overlapped video projectors. (d) Frozen video projector.

Interesting issues ensue from projecting the videos. When the avatars' fields of view overlap, the video projectors may also overlap. Then, the video projector of the avatar *closer* to the expert is brought forward. Figure 3-(c) shows the expert's view, and the green avatar, *B*, is closer to the expert. Consequently, *B*'s video projector overlays the red avatar *A*'s. This is a reasonable choice in that the expert would have come closer to *B* so as to advise *B*.

Even though the video projector is a substantial improvement over the video billboard, it does not change the 3D workspace model at all. Suppose that an avatar assembles a target configuration in the assembly table. The video projector enables the change being made on the table to be visible from an arbitrary position of the expert around the table. If the avatar moves to the search tables to pick up another tan, however, the assembly table will return to the initial state.

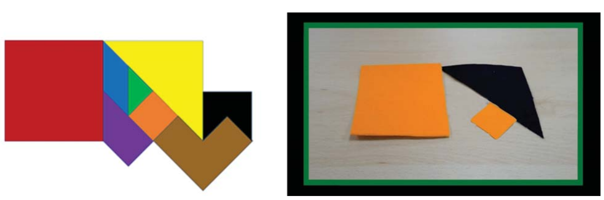
In order for the expert to save the change made by an avatar, a *freeze* function is added. If the function is invoked, the current frame of the video projector is textured to the scene objects and remains fixed. See Figure 3-(d). The green avatar is not in front of the assembly table any longer, but the frozen projector remains fixed at the table.

5 Main Experiment

5.1 Participants

In the main experiment, 17 subjects (4 females, 13 males) participated as the remote expert. They aged between 19 and 30 ($\mu = 24.59$, $\sigma = 2.67$). Fifteen subjects had experiences in VR. No subject was color-blind or had color amblyopia. Each subject was paid 30 USD.

5.2 Design



(a) **(b)**

Figure 4: Tangram-like puzzle. (a) A target configuration. (b) Three tans initially placed at the assembly table (*2D Video* view).

We used 64 unique tans (8 shapes $\times$ 8 colors). The subject was given a target configuration composed of 8 tans, each with distinct shape and color. An example is shown in Figure 4-(a). The target configuration was rendered in the subject's VR screen, and the subject could remove it temporarily to secure a clear view of the workspace.

Initially, three tans were placed on the assembly table, as shown in the *2D Video* case in Figure 4-(b). Their shapes and positions were identical to those given in the target configuration. However, two of them had different colors and therefore they had to be replaced by the correct tans. This imitates a real-world scenario where the wrong components of an object have to be replaced.

The remaining 61 tans (64 unique tans minus 3 tans on the assembly table) were distributed to the search tables. The expert had to instruct the local users to replace two incorrect tans and also assemble the target configuration using another five tans. The instructions can be made either verbally or with line pointing. The local users (the employed actors) did not make any movement unless instructed by the expert.

Out of the seven tans to be looked for, four were placed on a group of three connected search tables, and the other three were on another group of three tables. (Two groups of search tables can be found in Figure 1-(b).) The expert was not informed in advance of the tans' distribution.

Only after the cycle of "$\langle 1 \rangle$ looking for a tan, $\langle 2 \rangle$ picking it up, $\langle 3 \rangle$ bringing it to the assembly table, and $\langle 4 \rangle$ placing it to the desired position of the target configuration" is completed, the next tan will be looked for. The steps of $\langle 1 \rangle$, $\langle 2 \rangle$ and $\langle 3 \rangle$ make up what we call the *search procedure* whereas $\langle 4 \rangle$ is called the *assembly procedure*.

5.3 Experiment Procedure

The experiment was composed of two scenarios. In the first scenario (henceforth, S1), a local user can be involved in both search and assembly procedures, and the expert can freely choose between the two users for either procedure.

Unlike in S1, where the local users work independently, the users' roles are divided in the second scenario (henceforth, S2). One user is dedicated to the search procedure whereas the other user is dedicated to the assembly procedure. We wanted to see whether the view-sharing techniques would be perceived differently depending on the collaboration scenarios.

In both S1 and S2, a subject went through three trials, each with a distinct view-sharing technique, resulting in a total of six trials through out the experiment. The order of the techniques was counterbalanced between subjects. The three trials were randomly matched with three different target configurations in each scenario, the order of which was also counterbalanced between subjects.

Before starting the experiment, each subject filled out a demographic survey, and the experiment procedure was explained in detail. Prior to each trial, a tutorial including five-minute training was given. There was a five-minute break between the trials.

5.4 Evaluation Items

During the experiment, all trials of the subjects were recorded in videos. We then measured the times and usage frequencies of the evaluation items from the videos.

By default, the initial position of a local user is right in front of the assembly table. The *search time* is defined as the time consumed by the search procedure, i.e., the time it takes for a local user to leave the assembly table, find a tan from the search tables, and then bring it to the assembly table. Similarly, the *assembly time* is what the assembly procedure consumes. During the experiment, we measured the search and assembly times separately.

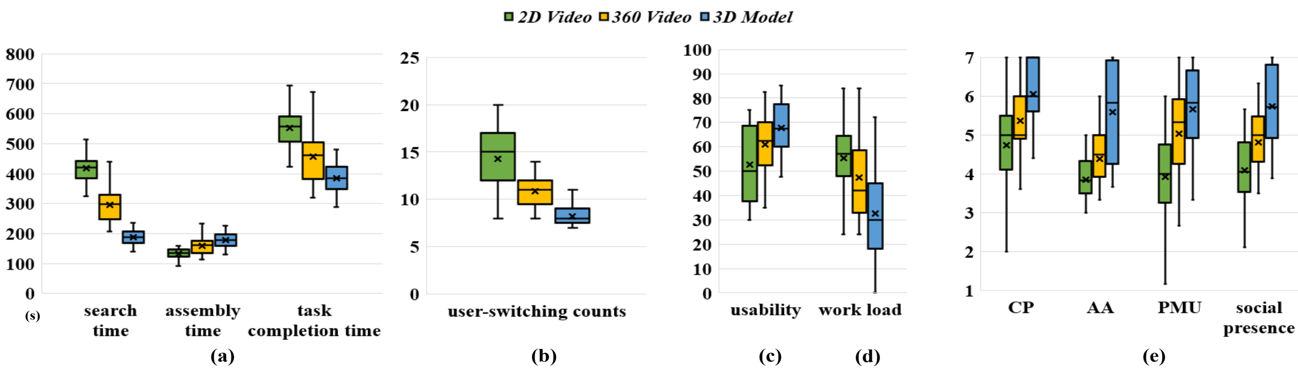


Figure 5: S1 analysis results.

In each view-sharing technique, we counted how many times the subject switched between local users. The number was counted only when specific instructions were given to the local users. If the expert switched from one user to the other but did not give any instructions, for example, it was not counted.

We also counted and measured a few features that are specific to *360 Video* and *3D Model*:

- In *360 Video*, we counted how many times "zooming" was made.
- In *3D Model*, we counted how many times "switching between the video billboard and the video projector" was made.
- In *3D Model*, we counted how many times "freezing the video projector" was made.
- In *3D Model*, we measured the time the video billboard was active and that the video projector was active.

The subjects were asked to fill out a survey after each trial: SUS questionnaire [5] for system usability, Raw TLX [15] for work load, a social presence questionnaire [14], and SSQ [18] for sickness. After finishing the entire experiment, the subjects ranked the view-sharing techniques in order of preferences and selected the most effective technique for each of search and assembly procedures.

5.5 Hypotheses

We formulated hypotheses about the experiment results based on previous works and our observations made in the pilot test:

H1 *360 Video* will take the shortest search time.

H2 *2D Video* will take the shortest assembly time.

H3 The usability will be the highest in *3D Model*.

H4 The work load will be the highest in *2D Video*.

H5 The social presence will be the highest in *3D Model*.

H1 was based on the results of a previous work [44] that compared a method using 360 camera and that using a pre-reconstructed static 3D model. The comparison was made with a search task in a one-to-one collaboration scenario. The results showed that the 360 camera method took less time.

H2 was formulated as we assumed that *2D Video* would help the subjects focus on the assembly procedure by not showing unnecessary environmental information which *360 Video* or *3D Model* would show. **H3** was formulated since *3D Model* would allow the subjects to freely move in the environment and instruct the local users with more ease. **H4** was based on the assumption that in *2D Video*, the remote expert would repeatedly instruct the local users to change their views. **H5** was formulated since *3D Model* would allow the remote expert to feel co-located with both local users at the same time.

6 Scenario 1: Independent Tasks

In S1, each local user can be involved in both the search procedure and the assembly procedure. For the search procedure, a user was allowed to access only a group of three connected tables in the room setup of Figure 1-(b), i.e., the right three tables are accessible only by the actor wearing a black t-shirt and the left three are only by the white t-shirt actor.

This section analyzes the results of the experiment made in S1. The Shapiro-Wilk test ($p < 0.05$) showed that all data were not normally distributed. Therefore, we used a non-parametric statistical test, the Friedman test, throughout the analysis.

Search and assembly times The results for the search and assembly times are illustrated in Figure 5-(a), where the *task completion time* is the sum of the search and assembly times. The Friedman test showed significant differences between the three techniques in search time ($X^2(3) = 34.111, p < 0.001$), assembly time ($X^2(3) = 14.778, p = 0.001$), and task completion time ($X^2(3) = 25.000, p < 0.001$). The results of the post hoc analysis using Wilcoxon signed-rank test, applied with a Bonferroni correction (significance level set at $p < 0.0167$), are presented in Table 1.

evaluation item		*2D Video - 360 Video*	*2D Video - 3D Model*	*360 Video - 3D Model*
search time	*Z*	-3.574	-3.621	-3.621
	p	< .001	< .001	< .001
assembly time	*Z*	-2.627	-3.621	-1.965
	p	.009	< .001	.049
task completion time	*Z*	-2.959	-3.621	-2.769
	p	.003	< .001	.006

Table 1: S1 post hoc analysis results for search and assembly times.

Switching between users The statistical results for user-switching counts are illustrated in Figure 5-(b). The average user-switching counts were in the decreasing order of *2D Video* (14.24 times), *360 Video* (10.82 times) and *3D Model* (7.94 times). The Friedman test showed a significant difference between the three techniques ($X^2(3) = 23.493, p < 0.001$), and the post hoc analysis revealed significant differences between *2D Video* and *360 Video* ($Z = -3.113, p = 0.002$), between *2D Video* and *3D Model* ($Z = -3.581, p < 0.001$), and between *360 Video* and *3D Model* ($Z = -3.564, p < 0.001$).

Features specific to *360 Video* and *3D Model* Listed below are the averages of the counts and measurements:

- In *360 Video*, "zooming" was made 6.08 times.
- In *3D Model*, "switching between the video billboard and the video projector" was made 7.48 times.
- In *3D Model*, "freezing the video projector" was made 6.72 times.

- In *3D Model*, the video billboard was active during 47.61% of the task completion time, and the video projector was active during the rest, 52.39%.

Usability The SUS questionnaire was answered on a 5-point Likert scale, and the statistical results are shown in Figure 5-(c). The Friedman test showed a significant difference between the three techniques ($X^2(3) = 15.121, p < 0.001$). The post hoc analysis revealed significant differences between *2D Video* and *3D Model* ($Z = -2.771, p = 0.006$) and between *360 Video* and *3D Model* ($Z = -2.525, p = 0.012$).

Work load The statistical results for work load are shown in Figure 5-(d). The Friedman test showed a significant difference between the three techniques ($X^2(3) = 16.646, p < 0.001$). The post hoc analysis revealed significant differences between *2D Video* and *3D Model* ($Z = -3.054, p = 0.002$), and between *360 Video* and *3D Model* ($Z = -3.224, p = 0.001$).

Social presence The social presence questionnaire has three subscales: co-presence (CP), attentional allocation (AA), and perceived message understanding (PMU). The questionnaire was answered on a 7-point Likert scale, and the statistical results are shown in Figure 5-(e).

The Friedman test showed significant difference between the three techniques in CP ($X^2(3) = 8.033, p = 0.018$), AA ($X^2(3) = 12.030, p = 0.002$), PMU ($X^2(3) = 18.317, p < 0.001$), and total social-presence score ($X^2(3) = 16.149, p < 0.001$). The results of the post hoc analysis are presented in Table 2.

evaluation item		*2D Video - 360 Video*	*2D Video - 3D Model*	*360 Video - 3D Model*
CP	Z	-1.983	-2.764	-2.539
	p	.047	.006	.011
AA	Z	-1.685	-3.002	-3.317
	p	.092	.003	.001
PMU	Z	-3.339	-3.439	-2.241
	p	.001	.001	.025
total social presence	Z	-2.511	-3.206	-3.362
	p	.012	.001	.001

Table 2: S1 post hoc analysis results for social presence.

Sickness We evaluated the sickness between the pre- and post-SSQs for each view-sharing technique. The statistical results are shown in Table 3. The Wilcoxon signed-rank test revealed that there were no significant differences between the pre- and post-SSQ scores for all three view-sharing techniques ($p > 0.05$).

techniques	pre-SSQ		post-SSQ		Z	p
	μ	σ	μ	σ		
2D Video	0.91	1.07	0.92	1.06	-1.41	0.16
360 Video	0.78	1.00	0.78	1.06	-1.89	0.59
3D Model	0.64	1.04	0.66	1.05	-1.63	1.02

Table 3: S1 analysis results for SSQ.

Preference and effectiveness The statistical results for preference and effectiveness are depicted in Figure 6. The preference was in the order of *3D Model*, *360 Video* and *2D Video*. With respect to the effectiveness, *3D Model* scored the highest for the search procedure (selected by 15 out of 17 subjects), and *2D Video* scored the highest for the assembly procedure (selected by 9 subjects).

7 Scenario 2: Divided Tasks

Unlike in S1, two local users had different roles in S2. One was dedicated to the search procedure, and the other was to the assembly procedure. This means that the user in charge of the search procedure can access all six search tables shown in Figure 1. Note that the

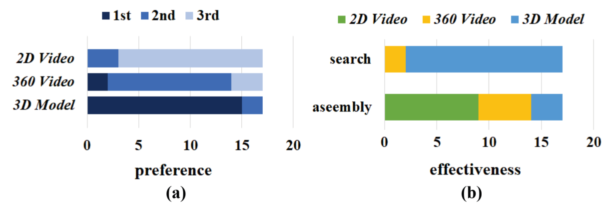


Figure 6: S1 analysis results for (a) preference and (b) effectiveness.

remote expert had to instruct the users to interact with each other for giving and taking the tans.

S2 also used three target configurations, but they were completely different from those in S1. The participants in S1 were re-recruited and S2 followed the same procedure of S1. The Shapiro-Wilk test ($p < 0.05$) showed that all data were not normally distributed, thus we used the Friedman test throughout the analysis.

Search and assembly times The statistical results for the search and assembly times are illustrated in Figure 7-(a). The Friedman test showed significant differences between the three techniques in search time ($X^2(3) = 28.778, p < 0.001$) and task completion time ($X^2(3) = 15.444, p < 0.001$) while there were no differences in assembly time ($X^2(3) = 4.111, p = 0.128$). The results of the post hoc analysis are presented in Table 4.

evaluation item		*2D Video - 360 Video*	*2D Video - 3D Model*	*360 Video - 3D Model*
search time	Z	-2.343	-3.621	-3.645
	p	.019	< .001	< .001
assembly time	Z	-0.024	-2.154	-1.775
	p	.981	.031	.076
task completion time	Z	-0.876	-3.385	-2.533
	p	.381	.001	.001

Table 4: S2 post hoc analysis results for search and assembly times.

Switching between users The statistical results for user-switching counts are illustrated in Figure 7-(b). The average user-switching counts were in the decreasing order of *2D Video* (15.59 times), *360 Video* (15.33 times) and *3D Model* (15.11 times). The Friedman test showed no significant differences between the three techniques ($X^2(3) = 4.333, p = 0.115$).

Features specific to *360 Video* and *3D Model* Listed below are the averages of the counts and measurements:

- In *360 Video*, "zooming" was made 5.34 times.
- In *3D Model*, "switching between the video billboard and the video projector" was made 4.63 times.
- In *3D Model*, "freezing the video projector" was made 3.42 times.
- In *3D Model*, the video billboard was active during 73.54% of the task completion time, and the video projector was active during the rest, 26.46%.

Usability The statistical results for usability are shown in Figure 7-(c). The Friedman test showed a significant difference between the three techniques ($X^2(3) = 21.031, p < 0.001$). The post hoc analysis revealed significant differences between *2D Video* and *3D Model* ($Z = -3.153, p = 0.002$) and between *360 Video* and *3D Model* ($Z = -3.553, p < 0.001$).

Work load The statistical results for work load are shown in Figure 7-(d). The Friedman test showed a significant difference between the three techniques ($X^2(3) = 15.594, p < 0.001$). The post hoc analysis revealed significant differences between *2D Video* and *3D Model* ($Z = -3.145, p = 0.002$), and between *360 Video* and *3D Model* ($Z = -3.258, p = 0.001$).

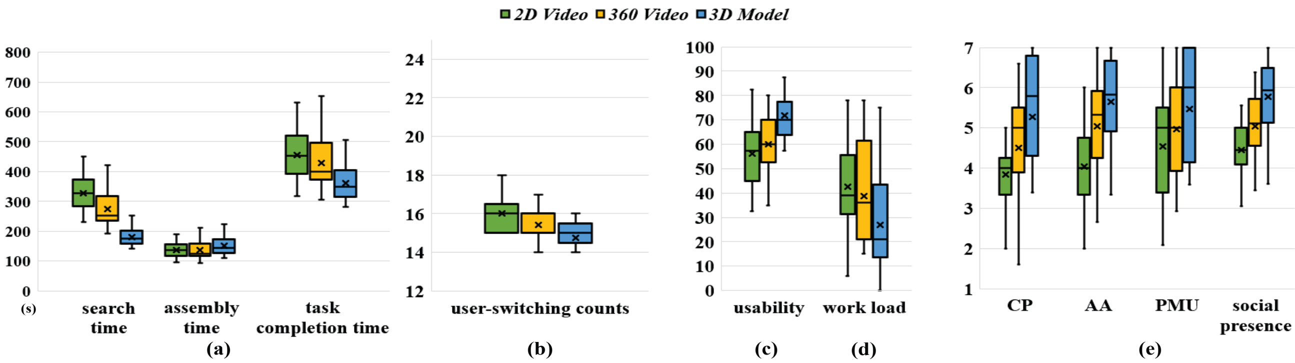


Figure 7: S2 analysis results.

Social presence The statistical results for social presence are shown in Figure 7-(e). The Friedman test showed significant difference between the three techniques in CP ($X^2(3) = 20.868, p < 0.001$), AA ($X^2(3) = 15.180, p = 0.001$), PMU ($X^2(3) = 16.305, p < 0.001$), and total social-presence score ($X^2(3) = 14.800, p = 0.001$). The results of the post hoc analysis are presented in Table 5.

evaluation item		*2D Video - 360 Video*	*2D Video - 3D Model*	*360 Video - 3D Model*
CP	*Z*	-2.770	-3.102	-2.050
	p	.006	.002	.040
AA	*Z*	-3.140	-3.439	-2.241
	p	.002	.001	.025
PMU	*Z*	-1.837	-2.712	-2.708
	p	.066	.007	.007
total social presence	*Z*	-3.183	-3.196	-2.612
	p	.001	.001	.009

Table 5: S2 post hoc analysis results for social presence.

Sickness We evaluated the sickness between the pre- and post-SSQs for each view-sharing technique. The results are shown in Table 6. The Wilcoxon signed-rank test revealed that there were no significant differences between the pre- and post-SSQ scores for all three techniques ($p > 0.05$).

techniques	pre-SSQ		post-SSQ		Z	*p*
	μ	σ	μ	σ		
2D Video	0.76	1.01	0.78	1.03	-1.73	0.83
360 Video	0.80	1.04	0.83	1.06	-1.21	0.59
3D Model	0.56	1.00	0.57	1.03	-1.41	0.157

Table 6: S2 analysis results for SSQ.

Preference and effectiveness The statistical results for preference and effectiveness are depicted in Figure 8. The preference was in the order of *3D Model*, *360 Video* and *2D Video*. With respect to the effectiveness, *3D Model* scored the highest for the search procedure (selected by 12 out of 17 subjects), and *2D Video* scored the highest for the assembly procedure (selected by 7 subjects).

8 Discussion

8.1 Search and assembly times and effectiveness

Table 7 summarizes the analysis results. With respect to the search time, *3D Model* was the best whereas *2D Video* was the worst. This can be explained using the different DOFs of the view-sharing techniques: In *2D Video*, the expert's view is fixed to the current user's, resulting in 0DOF. In contrast, *3D Model* supports 6DOF, which enables the expert to instruct the users where to go and what to do, not necessarily from their views. Moreover, *3D Model* dominated

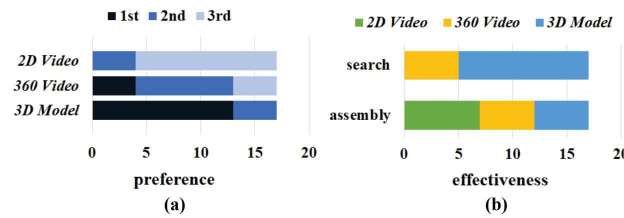


Figure 8: S2 analysis results for (a) preference and (b) effectiveness.

evaluation item	S1	S2
search time	3D < 360 < 2D	
assembly time	2D < 360 = 3D	3D = 360 = 2D
task completion time	3D < 360 < 2D	3D = 360 < 2D
user-switching	2D > 360 > 3D	3D = 360 = 2D
usability	3D > 2D = 360	
work load	3D < 2D = 360	
social presence	3D > 360 > 2D	
preference	3D > 360 > 2D	
effective(search)	3D(15) > 360(2) > 2D(0)	3D(12) > 360(5) > 2D(0)
effective(assembly)	2D(9) > 3D(5) > 360(3)	2D(7) > 3D(5) > 360(5)
billboard/projector	47.61% / 52.39%	73.54% / 26.46%

Table 7: Summary of the experiment results in S1 and S2, where ">" and "<" indicate significant differences, and "=" indicates no significant differences. The votes for effectiveness are in parentheses.

not only *2D Video* but also *360 Video*. Thus, **H1** ("*360 Video* will take the shortest search time.") is violated. This conflicts with the result reported in Teo et al. [44], where the 360 camera method was significantly better than the 3D model method. The conflict can be explained as follows: (1) In *3D Model*, we added the video projector, which is an improvement over the video billboard. (2) The search procedure in our experiment was not performed in the entire room but limited to tables. (3) Most importantly, our experiment was on instructing *multiple* users, and it was observed that the subjects found *3D Model* beneficial since the users' locations could be identified at a glance, helping the expert swiftly move to the next step.

The analysis results in the search time are compatible with those in effectiveness (in search). With respect to effectiveness (in search), the *3D Model*'s scores were much higher. Fifteen subjects made the same comment, "Compared to the other techniques, the field of view in *2D Video* was so limited that I had to instruct the user to move around many times to find the tans."

With respect to the assembly time, *2D Video* was the best in S1. Seven subjects stated that the expert's view in *2D Video*, which was fixed to the user's, rather helped them focus on assembling. Interestingly, *2D Video* did not dominate the other techniques in S2. Consider a user that was in charge of assembly. The user's camera pose remained largely still independently of the view-sharing techniques, and therefore the subjects rarely instructed the user to adjust the camera's view. This explains why **H2** ("*2D Video* will take the shortest assembly time.") is supported only in S1.

Note that the results of the assembly time are compatible with those of effectiveness (in assembly): *2D Video*'s score was much higher than the others' in S1, but it was not that higher in S2.

8.2 Switching between users

In S1, *2D Video*'s user-switching count was the highest whereas *3D Model*'s is the lowest. Suppose that, in *2D Video*, a subject looked for a specific tan from a user's viewpoint but failed. Then, the subject switched to another user to continue finding the tan. This was rarely observed in *3D Model*. A subject stated "With *3D Model*, instructing the local users was effortless. It was not the case in both *2D Video* and *360 Video*. When switching to another user, I needed a second to realize where and what the local user was performing."

In S2, there were no significant differences between the view-sharing techniques. Seven subjects pointed out that they were not especially confused in *2D Video* and *360 Video* since the two local users had different roles.

8.3 Usability, work load, and social presence

With respect to usability, work load, and social presence, *3D Model* was the best in both S1 and S2. As for usability, ten subjects commented "Although it took some amount of time to learn the features of *3D Model*, it turned out to be very useful when carrying out a trial." Seven of them added "I especially liked how I could freely move around and supervise the local users." **H3** ("The usability will be the highest in *3D Model*.") is supported. As for work load, there was no significant difference between *2D Video* and *360 Video*. **H4** ("The work load will be the highest in *2D Video*.") is not supported.

In social presence, *3D Model* was significantly higher. Four subjects commented "Since I could actually see the avatars and the video projectors with little restrictions, I really felt like we were working together." **H5** ("The social presence will be the highest in *3D Model*.") is supported.

All these explain why *3D Model* was the most preferred view-sharing technique for both S1 and S2. To summarize, *3D Model* was time-efficient, usable, less demanding (both mentally and physically), and the most preferred in one-to-many MR collaborations.

8.4 Features specific to *3D Model*

The video projector mode took 52.39% of the task completion time in S1. However, it took just 26.46% in S2. The number of switching between video billboard and video projector was proportionate to this time. The advantage of using a video projector was that the expert could share the users' views with less restriction even when their positions were dynamically changing. However, since S2 had one user fixed to the assembly table, the motivation to use the video projector seemed to have decreased.

A similar result was found in "freezing the video projector." It was observed that the video projector was frozen mostly during the assembly procedure. As presented earlier, the freeze count in S2 was about half of that in S1. This is because one user was fixed to the assembly table. The user's live video (regardless of whether billboard or projector) remained largely still, and therefore the subjects did not need to save the change made on the assembly table. This implies that such a video freeze function would be more useful for a scenario similar to S1 or a scenario where the local users are constantly moving.

Even in S1, however, there were subjects who freezed the video projector just once or twice. They stated "In this kind of simple experiment, I could memorize the change without freezing it." It indicates that the freeze feature might be more suitable for more complex applications.

9 Implementation Guidelines

In *2D Video* and *3D Model*, the view orientation that HoloLens' camera is streaming slightly differs from that of the HoloLens user because HoloLens camera is placed slightly above the holographic lenses. *360 Video* has a similar problem since the 360 camera is mounted upon HoloLens. In our experiment, the employed actors were trained to look at the tables with a certain range of orientations and distances so that the remote expert cares less about camera orientation control. In the real-world applications, it is most likely that the local user is untrained. Then, it will be useful to add a rectangular frame to the local user's holographic view that indicates which part of the view is actually being sent to the remote expert.

As discussed in S1, the remote expert tended to be confused while switching between users in *2D Video* and *360 Video*. The problem can be alleviated if additional visual interfaces are added to the expert's VR views. A good candidate would be a workspace's mini-map, which shows the locations of all local users as well as the expert's location.

Consider a related issue. With *3D Model*, the remote expert might find it cumbersome to physically walk to the local users' locations in a larger space. Hence, implementing basic VR teleportation functions or placing visual interfaces (such as buttons) to instantly teleport the remote expert to a local user should be considered.

Our implementation used a pointer to share the remote expert's attention. Many participants took it as an effective spatial interaction tool. Given a more complex task or environment, however, adding a "permanent" visual cue, such as a virtual sticky note, would be helpful. In *3D Model*, it is straightforward to implement such a cue as it can be easily registered to a virtual object.

10 Conclusion and Future work

We presented and evaluated three view-sharing techniques for one-to-many MR collaborations: *2D Video*, *360 Video* and *3D Model*. The study investigated on a specific environment, where the local users wearing AR headsets are instructed by a remotely located expert wearing a VR HMD. First, we adapted the view-sharing techniques used in one-to-one MR collaborations to fit our one-to-many MR collaboration environment. Then, each technique was modified through a pilot test, and experiments were conducted in two collaboration scenarios: Independent tasks and divided tasks. The overall results indicated that *3D Model* was the most preferred technique with the shortest task completion time, smallest work load, highest usability and social presence. We then provided guidelines for implementing view-sharing techniques for practical uses.

Our work has some limitations. In the workspace shown in Figure 1, we often suffered from HoloLens's tracking inaccuracy and loss. This problem was alleviated by placing objects around the environment to add features. Nevertheless, there were still some offsets that caused the visual cue, the pointer, to be presented to the local users with some error. There were also some offsets in the user's avatar position in *3D Model*.

The pre-reconstructed workspace should be as accurate as possible for the video projector in *3D Model* to work properly. Due to reconstruction errors, resulting in bumpy surfaces, the video projector may not always work satisfactorily. In the current implementation, we smoothed the pre-reconstructed surfaces to eliminate bumpy features. We envision however that rapidly developing SLAM technology will be able to produce smooth surfaces in the near future.

Our experimental setup was simple. The main components were tables and tans, which were all flat. It is challenging to go beyond the flat world. Not only is reconstructing a dynamic scene difficult, but the collaboration system also has to support 3D object segmentation. We are planning to tackle this problem by extracting the dynamic-fusion components from the recent performance capture techniques and then combining them with deep learning.

Our study focused on spatial interactions and visual communications. If the local users were allowed to freely converse with the remote expert, however, it would bring a huge impact on the way they collaborate. To control all variables in such a multi-modal collaboration, the experiment design should be done in a sophisticated manner. On the other hand, there were few communications between local users, and their experiences were neglected in our current study. Evaluating the performance and the amount of cognitive work load in respect of the local users with each technique will be also worth investigating.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (NRF-2017M3C4A7066316 and No. NRF-2016R1A2B3014319). We would like to thank SungIk Cho for discussions on the experiment design, Sangbin Kim and Seongsu Kwon for assistance in system setups, and all members of Korea University Media Lab for their support throughout the entire process.

REFERENCES

[1] Agisoft. Photo scan, 2019. `http://agisoft.com`.

[2] 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*, pp. 1343–1348. ACM, 2015.

[3] Autodesk. Meshmixer 3.5. `https://help.autodesk.com/view/MSHMXR/2019/ENU/`, 2018.

[4] M. Azure. Azure spatial anchor. `docs.microsoft.com/en-us/azure/spatial-anchors/`, 2017.

[5] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[6] A. Chellali, I. Milleville-Pennel, and C. Dumas. Influence of contextual objects on spatial interactions and viewpoints sharing in virtual environments. *Virtual Reality*, 17(1):1–15, 2013.

[7] R. A. J. de Belen, H. Nguyen, D. Filonik, D. Del Favero, and T. Bednarz. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. 2019.

[8] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 2019.

[9] 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*, pp. 513–520. ACM, 2003.

[10] 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*, p. 8. ACM, 2016.

[11] L. Gao, H. Bai, T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst. Real-time visual representations for mixed reality remote collaboration. 2017.

[12] 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*, pp. 449–459. ACM, 2014.

[13] S. Golodetz, T. Cavallari, N. A. Lord, V. A. Prisacariu, D. W. Murray, and P. H. Torr. Collaborative large-scale dense 3d reconstruction with online inter-agent pose optimisation. *IEEE transactions on visualization and computer graphics*, 24(11):2895–2905, 2018.

[14] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence. 2004.

[15] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

[16] A. Johnson, M. Roussos, J. Leigh, C. Vasilakis, C. Barnes, and T. Moher. The nice project: Learning together in a virtual world. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No. 98CB36180)*, pp. 176–183. IEEE, 1998.

[17] S. Kasahara and J. Rekimoto. 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*, pp. 217–225. ACM, 2015.

[18] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[19] 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)*, 27(3-6):569–607, 2018.

[20] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghurst. Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 173. ACM, 2019.

[21] 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 *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pp. 9–12. IEEE, 2016.

[22] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. A user study on mr remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 153–164. IEEE, 2018.

[23] T. Marrinan, J. Leigh, L. Renambot, A. Forbes, S. Jones, and A. E. Johnson. Mixed presence collaboration using scalable visualizations in heterogeneous display spaces. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 2236–2245. ACM, 2017.

[24] Microsoft. Microsoft hololens, 2018. `https://www.microsoft.com/en-Us/hololens.html`.

[25] M. Norman, G. A. Lee, R. T. Smith, and M. Billingurst. The impact of remote user's role in a mixed reality mixed presence system. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, p. 2. ACM, 2019.

[26] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 405–415. ACM, 2015.

[27] L. W. Pettersson, A. Kjellin, M. Lind, and S. Seipel. On the role of visual references in collaborative visualization. *Information Visualization*, 9(2):98–114, 2010.

[28] 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*, p. 46. ACM, 2018.

[29] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 228. ACM, 2019.

[30] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst. Covar: a collaborative virtual and augmented reality system for remote collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies*, p. 3. ACM, 2017.

[31] L. Pouliquen-Lardy, I. Milleville-Pennel, F. Guillaume, and F. Mars. Remote collaboration in virtual reality: asymmetrical effects of task distribution on spatial processing and mental workload. *Virtual Reality*, 20(4):213–220, 2016.

[32] Ricoh. Theta s, 2016. `https://theta360.com/en/about/theta/s.html`.

[33] S. D. Scott, M. S. T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 294–303. ACM, 2004.

[34] D. Snowdon, E. F. Churchill, and A. J. Munro. Collaborative virtual environments: Digital spaces and places for cscw: An introduction. In

Collaborative virtual environments, pp. 3–17. Springer, 2001.
[35] D. Snowdon, C. Greenhalgh, and S. Benford. What you see is not what i see: Subjectivity in virtual environments. *Framework for Immersive Virtual Environments (FIVE'95)*, 1995.
[36] 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*, pp. 179–188. ACM, 2013.
[37] M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar. Wysiwis revised: early experiences with multiuser interfaces. *ACM Transactions on Information Systems (TOIS)*, 5(2):147–167, 1987.
[38] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein. Slamcast: Large-scale, real-time 3d reconstruction and streaming for immersive multi-client live telepresence. *IEEE transactions on visualization and computer graphics*, 25(5):2102–2112, 2019.
[39] M. Tait and M. Billinghurst. The effect of view independence in a collaborative ar system. *Computer Supported Cooperative Work (CSCW)*, 24(6):563–589, 2015.
[40] A. Tang, M. Boyle, and S. Greenberg. Understanding and mitigating display and presence disparity in mixed presence groupware. *Journal of research and practice in information technology*, 37(2):193–210, 2005.
[41] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration in 360° videochat: Challenges and opportunities. Technical report, University of Calgary, 2017.
[42] F. Tecchia, L. Alem, and W. Huang. 3d helping hands: a gesture based mr system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 323–328. ACM, 2012.
[43] T. Teo, A. F Hayati, G. A Lee, M. Billinghurst, and M. Adcock. A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes. In *25th ACM Symposium on Virtual Reality Software and Technology*, p. 23. ACM, 2019.
[44] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 201. ACM, 2019.
[45] P. Tuddenham and P. Robinson. T3: Rapid prototyping of high-resolution and mixed-presence tabletop applications. In *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*, pp. 11–18. IEEE, 2007.
[46] S. Valin, A. Francu, H. Trefftz, and I. Marsic. Sharing viewpoints in collaborative virtual environments. In *Proceedings of the 34th annual Hawaii international conference on system sciences*, pp. 12–pp. IEEE, 2001.
[47] M. Yokozuka, S. Oishi, S. Thompson, and A. Banno. Vitamin-e: Visual tracking and mapping with extremely dense feature points. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9641–9650, 2019.
[48] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinqhurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556. IEEE, 2019.