International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

A User Study on the Comparison of View Interfaces for VR-AR Communication in XR Remote Collaboration

Eunhee Chang, Yongjae Lee & Byounghyun Yoo

To cite this article: Eunhee Chang, Yongjae Lee & Byounghyun Yoo (17 Aug 2023): A User Study on the Comparison of View Interfaces for VR-AR Communication in XR Remote Collaboration, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2023.2241294

To link to this article: https://doi.org/10.1080/10447318.2023.2241294

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2241294

A User Study on the Comparison of View Interfaces for VR-AR Communication in XR Remote Collaboration

Eunhee Chang[a†], Yongjae Lee[b†], and Byounghyun Yoo[c,d]

[a]University of South Australia, Mawson Lakes, Australia; [b]Arizona State University, Tempe, AZ, USA; [c]Korea Institute of Science and Technology, Seoul, Korea; [d]KIST School, University of Science and Technology, Seoul, Korea

ABSTRACT

Previous studies have investigated which context-sharing interfaces are effective in improving task performance; however, consistent results have yet to be found. In this study, we developed a convenient remote collaboration system that provides multiple context-sharing interfaces in a single platform (2D video, 360° video, and 3D reconstruction view). All interfaces can reflect live updates of environmental changes, and we aimed to clarify the effect of the interface on the quality of remote collaboration. Thirty participants were recruited to perform a target-finding-and-placing scenario. Using both objective and subjective metrics, we compared the task performance among the three interface conditions. The results showed that participants completed the task faster and reported a better collaborative experience in the 3D interface condition. Moreover, we found a strong preference for the 3D view interface. These results suggest that providing 3D reconstructed spatial information can enable remote experts to instruct local workers more effectively.

KEYWORDS

Remote collaboration; extended reality; virtual reality; augmented reality; context share and update

1. Introduction

Using extended reality (XR) technology, two or more workers who are geographically distant from each other can collaborate remotely. For example, an expert in a remote location can recognize the context of a local workspace using a virtual reality (VR) head-mounted display (HMD). Simultaneously, the partner in the local workspace can receive the guidance of the remote expert using an augmented reality (AR) device, as if the expert is right next to them. For successful collaboration, it is critical for collaborators to communicate appropriate information in a timely manner. In particular, under navigation or object-finding scenarios, a remote worker's accurate understanding of the spatial configuration of a local area is the key to achieving task goals (Qiu et al., 2023).

Previous studies have investigated which view interface is most effective for context sharing and inducing a better collaborative experience (Table 1). Choi et al. (2018) allowed a local worker to take pictures of the workspace (i.e., 2D images) and share them with a remote expert. In Lee, Kang, et al. (2020) and Piumsomboon, Lee, et al. (2019), a local worker held a spherical camera so that a remote expert could expand their view omnidirectionally. Gao et al. (2020) adopted a fixed 360° camera on the ceiling of a workspace, which could deliver a bird's-eye view of the space. Recently, several studies have shown that three-dimensional (3D) graphics augmented with the physical world can help local workers understand remote workers' intentions in a straightforward visual manner (Bai et al., 2020; Wang et al., 2021).

However, some studies have shown inconsistent results concerning the optimal view interface for facilitating collaboration. For example, Teo et al. (2019) compared 360° video and 3D reconstruction interfaces in an object-finding scenario, and showed that the 360° video condition enhanced understanding of the partner's focus and feeling of being together. On the other hand, Gao et al. (2020); Lee, Kang, et al. (2020) examined the effect of three types of view interfaces and found that participants preferred the 3D reconstruction view condition, which was considered time-efficient, usable, and less demanding. While these studies utilized similar experimental setups, participants indicated mixed reports about the collaborative experience.

This conflict might originate from the lack of balance between the interface conditions and the simplified task design. For example, although the 2D or 360° video view could reflect context changes in the local space (e.g., location changes in objects), the 3D reconstruction view interface consisted of static scenes; therefore, experts had to watch 360° videos to identify the current settings of the local site (Gao et al., 2020). Kim et al. (2020); Lee, Kang, et al. (2020); Teo et al. (2020) implemented an improved 3D reconstruction view interface that can update the texture of 3D

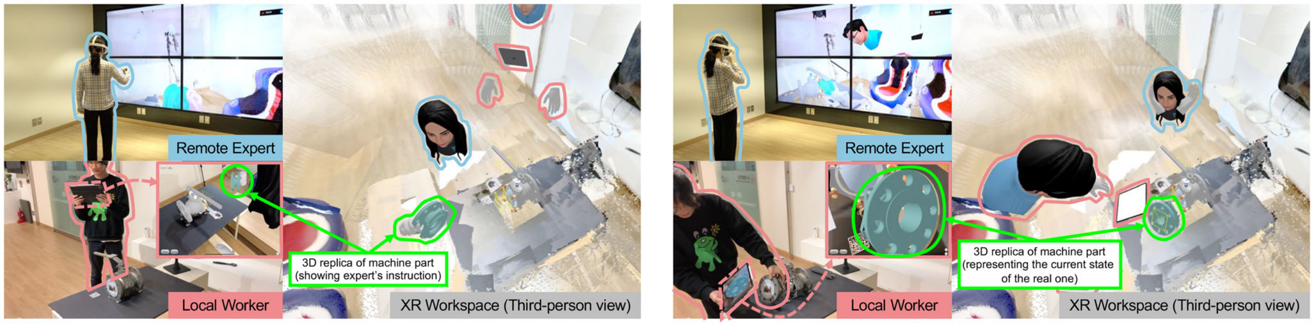
CONTACT Byounghyun Yoo yoo@byoo.net Korea Institute of Science and Technology, Seoul, Korea; KIST School, University of Science and Technology, Seoul, Korea

†These authors contributed equally to this work.

Table 1. Comparison of user studies on view interfaces in XR remote collaboration.

	Context-sharing interfaces (remote expert)	Hardware configuration (local worker)	Live update of 3D spatial information	Workspace type (area)	Task scenario	Shape of working target	Experimental design	Interface comparison result
Teo et al. (2019)	360° (HM) vs. 3D (Recon. mesh)	HoloLens + Wired 360°	Not available (static scene)	Room (6 m × 8 m)	Target finding	Three-dimensional block object	A-P*	Completion time (3D > 360°), SEQ (360°>3D), SoPQ (360°>3D), Preference (360°>3D)
Gao et al. (2020)	2D (FPV) vs. 360° (Fixed) vs. 3D (Recon. PC)	Magic Leap + Wired 360°	Not available (static scene)	Room (12 m × 6 m)	Target-finding-and-placing	Three-dimensional electronic gadgets	P-P**	Preference (3D > 360°>2D)
Lee, Kang, et al. (2020)	2D (FPV) vs. 360° (HM) vs. 3D (Recon. mesh)	HoloLens + Wired 360°	2D texture only***	Room (4 m × 7 m) (mainly work on tables)	Target-finding-and-placing	Flat object	A-P*	Completion time (2D > 360°>3D), SoPQ (3D > 360°>2D), SUS (3D > 360°>2D), NASA-TLX (2D > 360°>3D), Preference (3D > 360°>2D),
Anton et al., (2018)	2D (Top view) vs. 3D (Top view, Fixed, 3D Display, Recon. mesh) vs. 3D+ (Interaction)	Kinect + Webcam + Projector + Monitor	2D texture + 3D geometry	Tabletop	Target-finding-and-placing	Three-dimensional block object	P-P**	UEQ (2D > 3D+ >3D), Preference (3D+ >2D > 3D),
This research	2D (HHV) vs. 360° (HH) vs. 3D (Recon. mesh)	iPad + Wireless 360°	3D transformation (position + orientation)	Room (11.1 m × 8.4 m)	Target-finding-and-placing	Three-dimensional block object	A-P*	Completion time (360°>2D > 3D), Moving distance (360°>2D > 3D), GCE (3D > 360°>2D), SSQ-O (360°>2D > 3D), NASA-TLX (2D > 360°>3D) Preference (3D > 360°>2D)

FPV: First Person View; GCE: General Collaborative Experience; HH: Handheld; HHV: Handheld View; HM: Head Mounted; NASA-TLX: NASA Task Load Index; Recon. mesh: Reconstructed mesh; Recon. PC: Reconstructed Point Cloud; SEQ: Single Ease Question; SoPQ: Social Presence Questionnaire; SSQ-O: Simulator Sickness Questionnaire-Oculomotor; SUS: System Usability Scale; UEQ: User Experience Questionnaire.
*A-P means an actor conducts a local worker role, and a participant conducts a remote expert role.
**P-P means two participants conduct local worker and remote expert roles.
***2D texture only means updating the texture of 3D information is available, but geometry is not.



(a) (b)

Figure 1. An example of XR remote collaboration. (a) A remote expert in VR (cyan-highlighted woman and avatar) instructs the way of dissembling a machine part. (b) A local worker in AR (pink-highlighted man and avatar) performs the task following the expert's guide.

information; however, the transformation information of the 3D object was not updated. In the case of Teo et al. (2019), the collaboration task did not require the induction of spatial changes in the target objects, resulting in limited observation of the interface effect.

In addition, most collaborative scenarios are performed in a simple or restricted space, such as a tabletop (Gao et al., 2016, 2017; Grandi et al., 2019; Wang et al., 2021) or a single room (Jing et al., 2021; Piumsomboon et al., 2017; Piumsomboon, Dey, et al., 2019). Although this setup simplifies the workspace, distinguishing the characteristics of each interface in terms of context sharing may be difficult. When a scenario occurs in a complex area, there is more room to identify the view interface that is effective for task performance. Assuming that a real-world collaborative scenario requires mutual interactions in a complicated space, we set up a space of higher spatial complexity with several walls. Using this setup, we aimed to clarify the effect of context-sharing interface on remote collaboration.

In this study, we developed three interchangeable context-sharing interfaces in addition to our previous remote collaboration system, the Webized eXtended Reality (WXR) (Lee, Moon, et al., 2020; Lee & Yoo, 2021; Lee et al., 2021). The upgraded system offers three different context-sharing interfaces: 2D video, 360° video, and 3D reconstruction view. To clarify the interface effect on task performance, all view interfaces synchronously reflect spatial changes in the local workspace, ensuring that each interface can provide changes in the details of the workspace in real time during collaboration. In other words, live updates of 3D spatial information in a comparatively spacious and complex workspace, which have been limited in previous studies (Gao et al., 2020; Kim et al., 2020; Lee, Kang, et al., 2020; Teo et al., 2019, 2020), were available in this study. Additionally, a remote expert can freely move their viewpoint to gain contextual information about the local workspace, regardless of the type of interface used.

Furthermore, we designed a user experiment with an actor–participant pair. According to Teo et al. (2019), a proactive local worker can perform a collaborative task regardless of guidance from a remote expert, which may confound the effect of the view interface during the experiment. In the studies of Teo et al. (2019) and Lee, Kang, et al. (2020), a participant served as a remote expert while an actor performed as a local worker. Using this experimental design, the local worker (i.e., actor) only behaved in line with the expert's (i.e., participant's) instructions. Based on this background, this study chose the actor-participant pair design to clarify the effectiveness of each context-sharing method without intervention from uncontrolled factors.

2. Related work

2.1. XR remote collaboration in VR-AR environment

XR remote collaboration refers to working together on virtual data (e.g., 3D CAD) or real objects (e.g., machine maintenance) through various types of reality. During XR remote collaboration, a combination of VR-VR, AR-AR, and VR-AR environments can be applied, depending on the context (Ens et al., 2019; Lee & Yoo, 2021). When the collaboration involves manipulating real-world objects, the typical environment combination consists of a remote expert using VR and a local worker using AR. Because the worker must physically manipulate the objects with their hands, it is effective for a local worker to use AR while a remote expert gives instructions using VR. According to Grandi et al. (2019), the VR-AR combination was the preferred type of collaboration, and participants showed significantly better task performance under the VR-AR combination when manipulating 3D objects compared to the AR-AR and VR-VR combinations.

Figure 1 illustrates an example of VR-AR remote collaboration. Two collaborators in different XR environments worked together in the same XR workspace. The remote expert in VR (cyan-highlighted woman and avatar) shows how to detach a machine part (a flange), and the local worker in AR (pink-highlighted man and avatar) learns the procedure through the overlaid graphics on his display (Figure 1(a)). As the local worker follows the instructions (Figure 1(b)), the expert can observe context changes through the 3D reconstructed space. The 3D replica of the machine part within this space serves as a representation of the real object.

2.2. Context-sharing methods

Various systems have been proposed for efficient XR remote collaboration. Typically, they are based on one or more of the three major methods of sharing the context information of a local workspace.

2.2.1. 2D video

The 2D video view interface copies and streams the local worker's display to the remote expert or directly delivers live videos captured by cameras that are installed or held by the worker (Fussell et al., 2003). Based on the information drawn from these frame sequences, the remote expert makes decisions or gives instructions to the local worker. Early studies favored this interface because of its simple setup and abundance of corresponding off-the-shelf products (Bauer et al., 1999; Bottecchia et al., 2009; Gurevich et al., 2012; Kikuchi et al., 2022). Even today, this interface still plays a role in XR remote collaboration, as the confined view angle helps collaborators focus on their tasks (Gauglitz et al., 2014; Higuch et al., 2016). However, this confined view makes it difficult for the remote expert to comprehensively grasp the local workspace. The remote expert may need to request multiple camera adjustments from the local worker to achieve the desired viewpoint. Additionally, camera movements that result in unstable video can distract the remote expert from the task at hand.

2.2.2. 360° video

The 360° video view interface provides omnidirectional live capture and streaming of a workspace during collaboration. Many studies have adopted wired cameras installed in certain spaces, such as ceilings, displays, and rooms (Gao et al., 2020; Jing et al., 2021; Lee et al., 2023; Li et al., 2020; Speicher et al., 2018). In this hardware setup, the viewpoint of the context-sharing interface depends on the position of the camera. Thus, the VR remote expert was not able to change the viewpoints of the 360° video. Otherwise, the local user can attach the 360° camera to an AR device, or hold it in their hand and transmit the video to the expert in VR (Lee et al., 2017; Piumsomboon, Lee, et al., 2019). In this case, the 360° video is rendered egocentrically to the expert; therefore, the VR remote expert can voluntarily check the local situation by changing the viewing direction.

2.2.3. 3D reconstruction

The 3D reconstruction view interface offers a virtual workspace where the remote expert can gain an understanding of the local context through virtual objects that replicate the real ones at the local site. This interface allows the remote expert to freely select a favorable viewpoint and have a clear view. Typically, we can categorize the type of reconstruction into static and partial or full live 3D reconstruction. In the static reconstruction, the virtual objects are generated prior to collaboration and remain fixed in shape and arrangement throughout the collaboration process (Gao et al., 2018, 2020; Gauglitz et al., 2014; Teo et al., 2019). To overcome the limitation, some studies have utilized live video (Gao et al., 2018) or re-texturing techniques (Gauglitz et al., 2014).

Compared to static reconstruction, partial or full live reconstruction can contribute to the facilitation of accurate spatial understanding. Bai et al. (2020); Gao et al. (2016); Tecchia et al. (2012) shared lively captured color and depth information from the local site using one or multiple RGB-depth (RGBD) cameras. Adcock et al. (2013); Stotko et al. (2019) have developed a method that utilizes a single RGBD camera to share the local context, covering a wide area by accumulating previous frames. Furthermore, Tait and Billinghurst (2015) proposed a collaboration system that provides the remote expert with a pre-reconstructed local site, updating only the object poses through tracking to enhance the experience within the virtual workspace.

2.3. Originality of this research

As summarized in Table 1, previous research had several limitations in adequately controlling confounding factors when comparing the impact of context-sharing interfaces in XR remote collaboration. For example, the previously introduced 3D reconstruction interface was limited to implementing a static scene of the local workspace while other interfaces could provide the live stream of the spatial information. Regarding the viewpoint change availability, the 360° video interface often constrained the view angle of the remote expert to a specific position, which might result in a poorer collaboration experience compared to other interfaces. Lastly, previous user experiments were conducted in a relatively restricted area, so there might be few chances of observing differences in context-sharing effect between each interface.

Taken together, this research considered the following points to overcome the limitations mentioned above. In addition, regardless of the types of context-sharing interface, we provided sufficient visual information such as virtual replicas of task objects and avatars of collaborators to uphold originality in this research.

1. Real-time update of the local workspace: To clarify the effect of the context-sharing interface on task performance, we designed all interfaces to provide real-time spatial information about the local workspace. Each interface can synchronously show changes in the local context, such as positional and rotational updates of objects. To fully evaluate the potential of each interface, we introduced volumetric objects (i.e., waffle blocks) as task targets, which were designed to cause geometrical changes according to the manipulation of local workers.
2. Flexibility in viewpoint changes: We ensured all three interface conditions can provide the experience of viewpoint changes by equipping each local worker with an AR tablet and a wireless 360° camera. Most previous research used a fixed 360° camera, which only allows a stationary viewpoint (i.e., bird's-eye view) during collaboration in the 360° video condition. As remote experts can voluntarily select their perspectives in the 2D and 3D conditions, we allowed the experts to freely change viewpoints in the 360° video condition as well.

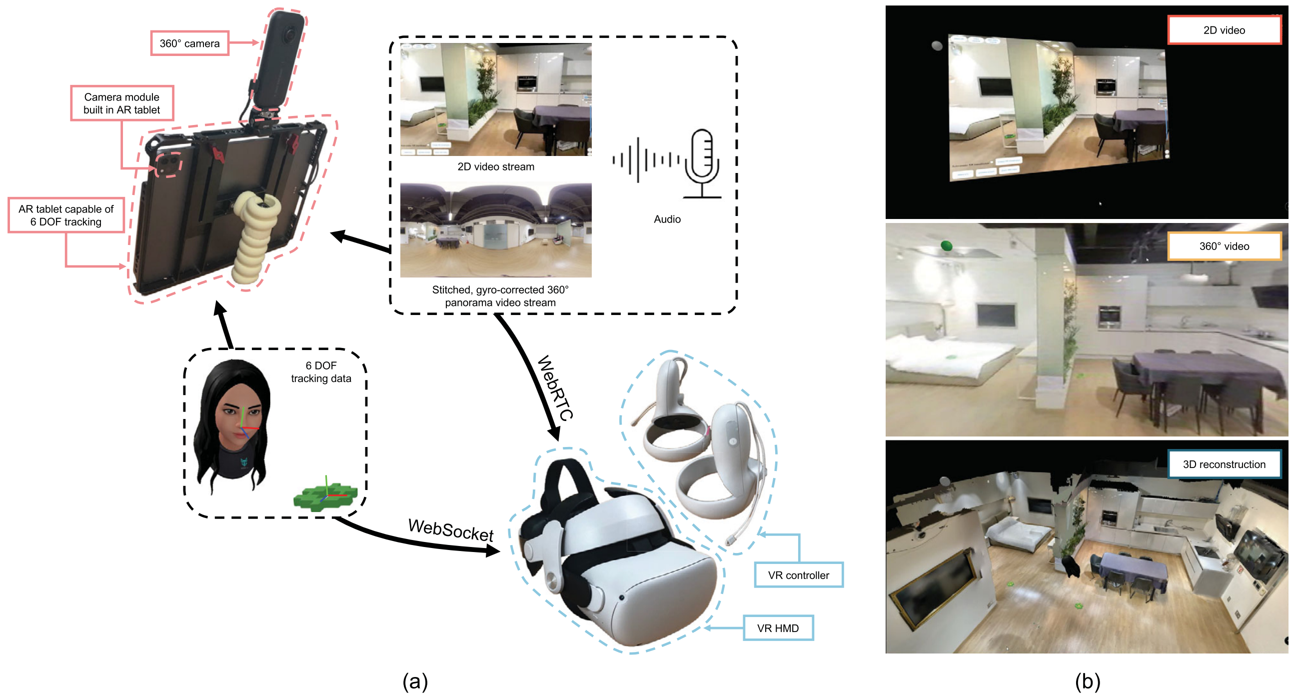


Figure 2. (a) Data communication protocols between the iPad and Meta Quest 2. (b) Three types of view interface: 2D video, 360° video, and 3D reconstruction.

3. A spacious local workspace with a complex structure: Existing collaboration spaces are relatively simple or restricted to table areas. We set a realistic residential space surrounded by walls to increase the spatial complexity. Using this experimental setup, we attempted to determine which interface is most robust and effective for context sharing.

3. User study

3.1. Participants

Thirty participants (including 15 women) were recruited for the user study. The mean age of participants was 27.0 (± 3.9). Each participant was paired with an actor and performed a remote collaboration scenario: a target-finding-and-placing task as a remote expert (see more details in Section 3.3). The actor was trained to perform consistent behaviors based on prepared scenarios. To clarify the effect of context-sharing interfaces on task performance, we intentionally recruited participants without any prior spatial knowledge of the local workspace. All experiments were performed under the guidelines of the institutional review board of the author's institution (IRB-2021-036).

3.2. Setup

The present remote collaboration system was implemented based on immersive web[1] technologies. In our previous studies, we developed the WXR system, which allows users to have the same experience regardless of the XR devices used if the device's browsers support web standards (Lee & Yoo, 2021). The server program was implemented with Node.js and MariaDB and ran on a PC (Intel Xeon E5-2699 v4 2.2 GHz CPU, 64 GB DDR4 RAM) under the Microsoft Windows 10 operating system. The client program was implemented using A-Frame (v1.2.0) (Marcos et al., 2023), an open-source web framework for creating VR scenes. The client program runs in all browsers supporting web standards, and users can connect any device to the browsers using the WebXR Device API (World Wide Web Consortium, 2022). Because of the absence of an object tracking function for the WebXR Device API, we additionally implemented a custom web browser for iOS using the WebKit and ARKit frameworks.

During the XR remote collaboration, we adopted the iPad Pro 12.9-inch (4th generation)[2] as an AR device for the local worker and the Meta Quest 2[3] as a VR device for the remote expert. The iPad communicated with the Quest 2 directly (using WebRTC) or through a WXR server (using WebSocket) over Wi-Fi (Figure 2(a)). Depending on the condition, the participants received spatial information about the local workspace through one of the three view interfaces: 2D video, 360° video, or 3D reconstruction (Figure 2(b)).

In the 2D video interface, the expert can collaborate while watching the local worker's display, rendered in a 2D rectangle in a virtual space with a black background. The local worker can transmit the 2D video in the direction and position that the expert requires. The video (with audio) stream is communicated wirelessly through WebRTC at a resolution of 960 × 720 @ 60 fps.

For the 360° video view interface, an Insta360 One X[4] was mounted on top of the tablet, and the iPad stitches, gyro-corrects, and streams the 360° video to the Quest 2. In line with the 2D condition, the expert can achieve the desired viewpoint by asking the local worker to change the camera position. The 360° video (with audio) is also wirelessly streamed through WebRTC at a resolution of 1440 × 720 @ 30 fps.

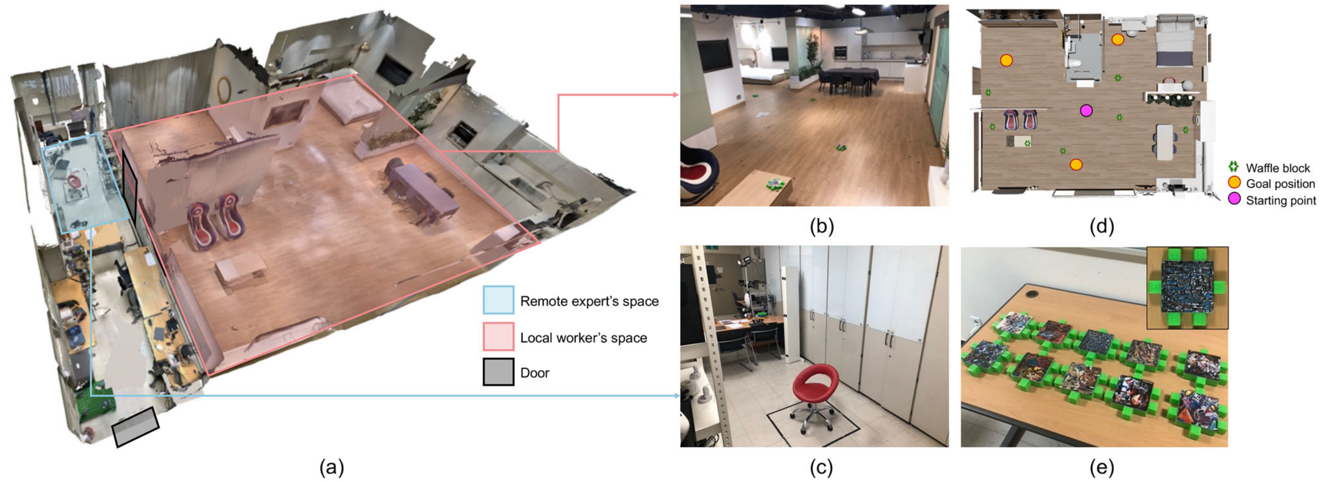


Figure 3. (a) Floor plan of the experimental space. (b) Pink-highlighted area refers to the local worker's space where the target-finding-and-placing scenario is conducted. (c) Cyan-highlighted area refers to the remote expert's space. (d) Seven waffle blocks are scattered in the local workspace, and three of them have to be placed in a designated location. (e) Examples of waffle blocks used in this research.

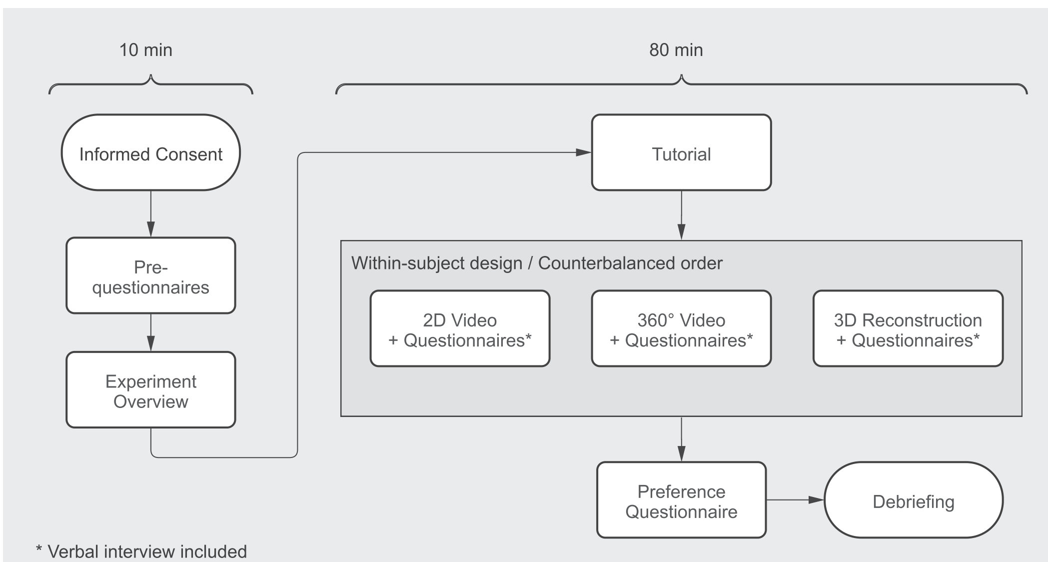


Figure 4. Diagram of the experimental procedures.

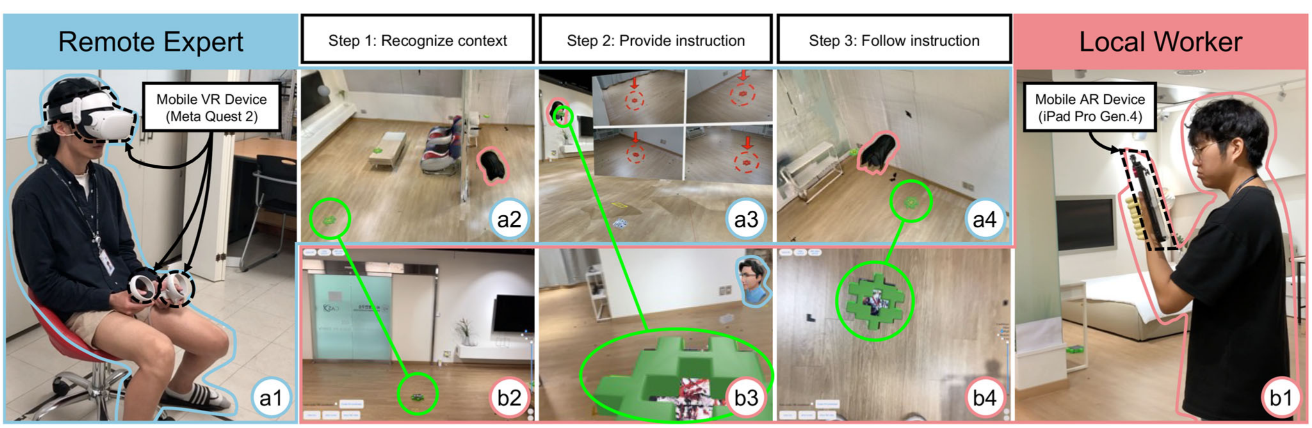


Figure 5. Detailed procedures during the collaborative task. (a1) A remote expert was located in the operation area wearing an HMD. (a2–a4) A remote expert's view in the HMD (in this figure, 3D reconstruction view interface). Pink-highlighted avatar represents a local worker. (b1) A local worker holding a tablet followed the expert's instructions. (b2–b4) A local worker's view in the tablet. Waffle blocks and the avatar of the expert (cyan-highlighted) were augmented on the local worker's screen (avatar model courtesy: Wolf3D (https://wolf3d.io/)).

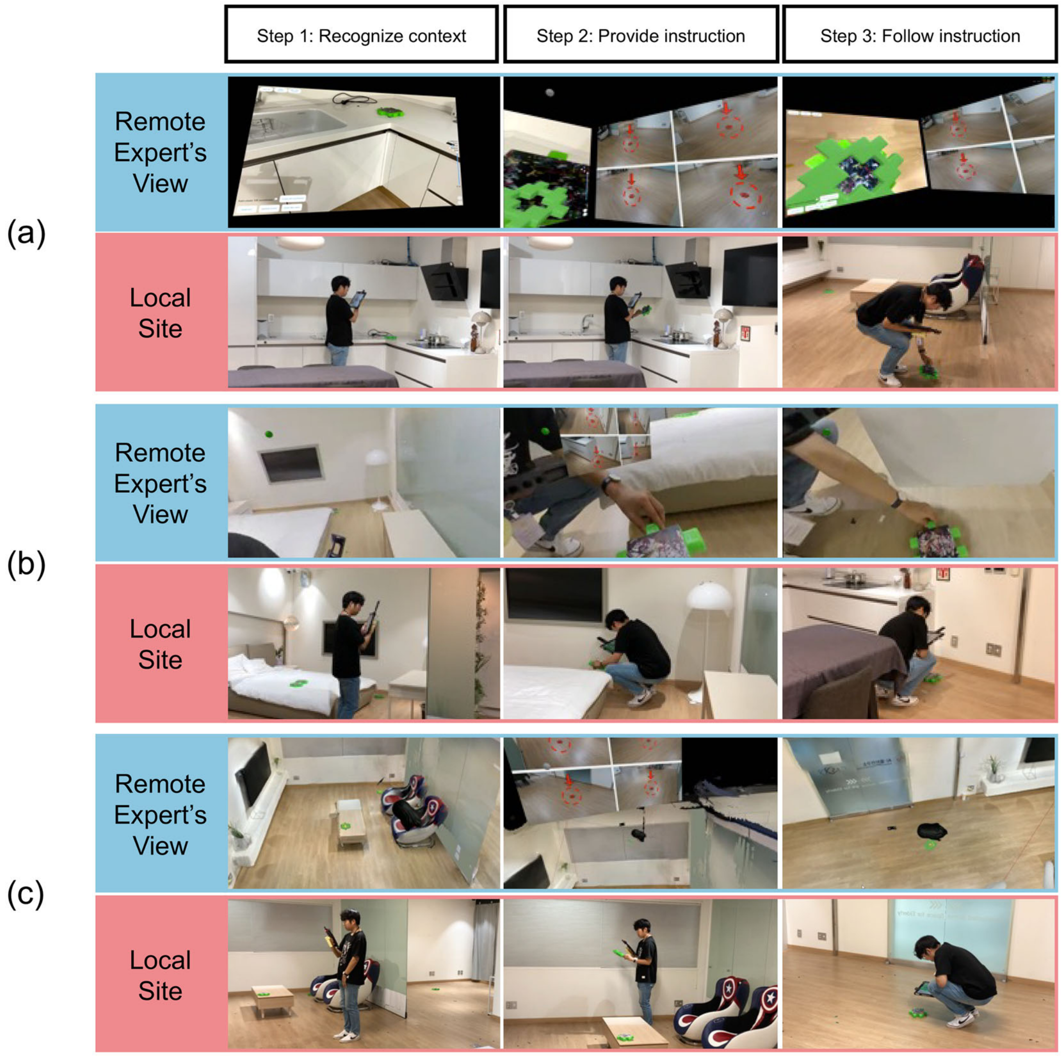


Figure 6. Comparison between three interfaces during the target-finding-and-placing scenario. While the local worker navigated around the workspace (pink-highlighted box), the remote expert instructed him to find and place the target using each interface (cyan-highlighted box): (a) 2D video, (b) 360° video, and (c) 3D reconstruction view interface.

In the 3D reconstruction view interface, we divided the local workspace into background and foreground objects to support real-time updates in this interface. Background objects, such as walls or furniture, were not of interest during collaboration and should remain stationary. These objects were scanned and 3D reconstructed using the 3D Scanner App™ from Laan Labs.[5] Foreground objects (the waffle blocks in our experiment) can be relocated in the workspace during the collaboration, and 3D models of these objects were created using Blender. The 3D waffle block models were managed by a web server and downloaded when the collaboration occurred. To correctly show the locations of the local worker and waffle blocks in the virtual space of the remote expert, we tracked their trajectories using ARKit, which offers the pose of the device in six degrees of freedom through its customized LiDAR-Visual-Inertial simultaneous localization and mapping (SLAM) algorithm. The ARKit also offers the pose of tracked images attached to the waffle blocks through its image tracking algorithm. The tracking information is delivered to the remote expert through WebSocket via the WXR server to log the local worker's movement. In this interface, no video but audio information was streamed through WebRTC, and the remote expert can change the viewpoint by manipulating the VR controllers.

Regardless of the view interfaces, the local worker's AR device rendered 3D waffle block models upon the video feed when detecting the physical blocks. This feedback aimed to assure the local worker that the system accurately identified the objects and displayed them within the virtual workspace. Additionally, all interfaces included representations of avatars for both the local worker and remote expert on their respective devices, fostering a stronger sense of co-presence. As each view interface operates on the same WXR architecture, most collaboration features, such as camera pose tracking, object pose tracking, and voice communication, were shared across all interfaces.

We used the "Living Lab" to simulate the spacious experimental area (Figure 3(a)). The Living Lab consists of a residential area (11.1 m × 8.4 m) (Figure 3(b)) and an operational area (2.4 m × 1.8 m) (Figure 3(c)). The residential area is similar to a real-world housing space, consisting of a

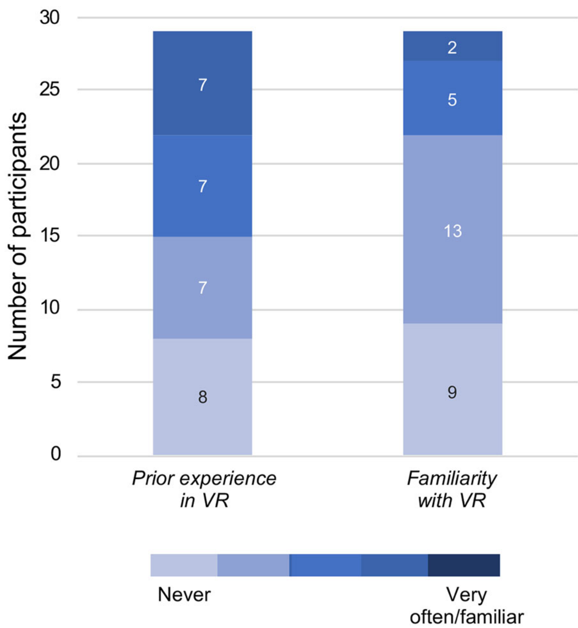


Figure 7. Self-report on prior experience in VR and familiarity with VR.

living room, bedroom, kitchen, and video wall. The operational area is a separate room from which the residential area can be monitored. A remote expert in the operational area (i.e., participant) delivered instructions to a local worker in the residential area (i.e., actor). Specifically, the expert instructed the local worker to find waffle blocks (Figure 3(d), green objects) distributed in the residential area and move them to designated places (Figure 3(d), yellow circles). We prepared ten waffle blocks of the same color and randomly selected seven for each interface experiment (Figure 3(e)). Our purpose was to set up a spacious and complex area to clarify the effect of context-sharing interfaces on task performance and user preferences.

3.3. Procedure

Figure 4 illustrates the procedure of the user experiment. Before the experiment, each participant signed an agreement to participate and answered pre-questionnaires. They then performed a tutorial to familiarize themselves with the collaborative scenario using each interface. In the tutorial, a remote expert (i.e., participant) instructed a local worker (i.e., actor) to locate two waffle blocks in the operational area and move them to an appropriate location.

After the tutorial, a remote expert was located in the operation area wearing an HMD (Figure 5(a1)), while a local worker moved to the residential area holding a tablet (Figure 5(b1)). When the experiment began, the two collaborators began searching for seven target objects (i.e., waffle blocks) scattered in the residential area to relocate three of them to a designated location, one by one (i.e., they conducted seven sub-tasks in a row). In Step 1 (Figure 5(a2) and (b2)), the remote expert inspects and recognizes the current state of the local site. After finding a block, the remote expert instructs the local worker on both the approach direction and the distance, based on the worker's current location represented by his or her avatar (pink-highlighted woman avatar). As the local worker follows the expert's instructions and detects the target block, an image pops up in front of the expert's view, showing where to move the block (Step 2, Figure 5(a3) and (b3)). Then, the expert further guides the local worker to finish the relocation of the block (Step 3, Figure 5(a4) and (b4)). In Figure 5, the green lines connect the 3D replicas seen by the remote expert (highlighted with the cyan lines) to the corresponding real blocks seen by the local worker (highlighted with the pink lines).

In all conditions, we restricted the remote worker to use explicit directional and longitudinal directives based on the local worker's perspective (e.g., "go forward 2 steps" or "turn right approximately 45 degrees"). This limitation aimed to ensure consistent communication from the expert when guiding the local worker to specific locations. If the target's locations were described relative to the remote expert's view (e.g., "over here" or "to the left of me"), it would require the local worker's knowledge to interpret the expert's directives based on the partner's point of view. This could potentially introduce confounding factors when evaluating the effect of the view interface on collaboration. By eliminating this additional interpretation process, the local worker was allowed to move directly from their current location to the target point by following the expert's instructions.

Note that Figure 5 depicts the example of the 3D reconstruction view interface, but the same scenario was also adopted for the 2D video (Figure 6(a)) and 360° video (Figure 6(b)) view interfaces. In the 2D and 360° video conditions, the local worker had to navigate the residential area to deliver spatial information to the remote expert, after which the remote expert instructed the local worker. However, in the 3D reconstruction view interface (Figure 6(c)), the remote expert could directly instruct the local worker, because the expert could navigate the reconstructed workspace using VR controllers.

3.4. Measurements

We measured the quality of remote collaboration using both objective and subjective metrics.

3.4.1. Task performance

3.4.1.1. Completion time per sub-task. We defined the completion time per sub-task as the total time of remote collaboration divided by the number of accomplished sub-tasks (i.e., the number of waffle blocks relocated). This approach was introduced to count only the number of successful tasks that the participants completed, depending on the interface. Both voice and video recordings were used to calculate this index.

3.4.1.2. Moving distance. For the purpose of measuring local worker's movement, we logged their trajectories in all three interfaces. Using the camera tracking function of the ARKit, we computed the distance by summing the entire trajectory for each interface condition (see Section 3.2). We assumed that the local worker's physical movement would differ depending on the context-sharing interface. That is, while 2D and 360° video require the local worker to wander through more expansive spaces to provide spatial information to the expert, the 3D reconstruction view interface enables the local worker to reduce physical activity because the remote expert can directly navigate the reconstructed local workspace.

3.4.2. Questionnaires

3.4.2.1. General collaborative experience (GCE). GCE is a 7-point Likert scale that evaluates the quality of collaboration (Wang et al., 2021). The index contains eight common questions for both remote and local sites and two additional questions for each site. A higher score indicates a better collaborative experience.

3.4.2.2. NASA task load index (NASA-TLX). NASA-TLX is a multidimensional index that assesses a participant's workload by considering six aspects of task performance: Mental demand, Physical demand, Temporal demand, Performance, Effort, and Frustration (Hart & Staveland, 1988). Each subscale is a 21-point Likert scale, and for each except for the Performance scale, a higher score indicates a more demanding task. For the Performance scale, a lower score indicates a participant successfully accomplishes the task.

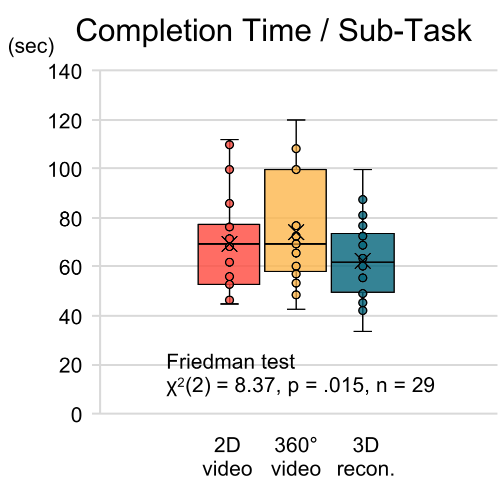


Figure 8. Average completion time per sub-task depending on the view interface.

3.4.2.3. System usability scale (SUS). We used the SUS score to evaluate the usability of each interface (Brooke, 1996). The SUS consists of 10 questions regarding various aspects of the system's usability. The overall value of the SUS ranges from 0 to 100.

3.4.2.4. Simulator sickness questionnaire (SSQ). The SSQ was used to measure the level of cybersickness after remote collaboration (Kennedy et al., 1993). Using this index, we attempted to quantify the severity of discomfort and demonstrate any differences between the context-sharing interfaces.

3.4.3. Preference rank

We instructed the participants to rank the interfaces that they had experienced from best to worst. After experiencing all interfaces, they completed a preference questionnaire adapted from Teo et al. (2019). The questionnaire consists of seven questions covering the overall user experience during collaboration.

3.5. Data analysis

The results of the Shapiro-Wilk test indicated violations of the normal distribution. Therefore, we performed the Friedman test for task performance (completion time and moving distance) and questionnaires (GCE, NASA-TLX, SUS, and SSQ) as well as the chi-squared test for preference rank. All statistical analyses were performed using R (version 4.0.3), and the significance level (α) was 0.05.

4. Results

Although we recruited 30 participants, one was excluded due to an incorrect understanding of our instructions. Thus, the data from 29 participants were used in the analysis.

4.1. Characteristics of participants

Before the experiments, the participants completed several pre-questionnaires: the motion sickness susceptibility questionnaire (MSSQ), prior VR experience (0: never to 5: very

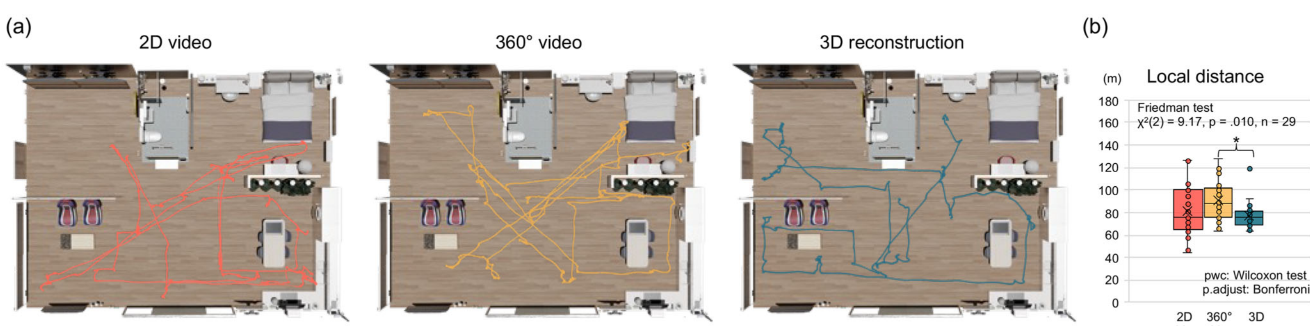


Figure 9. (a) Trajectories of the local worker in 2D, 360° video, and 3D reconstruction view interfaces, respectively. (b) Average moving distance between interfaces.

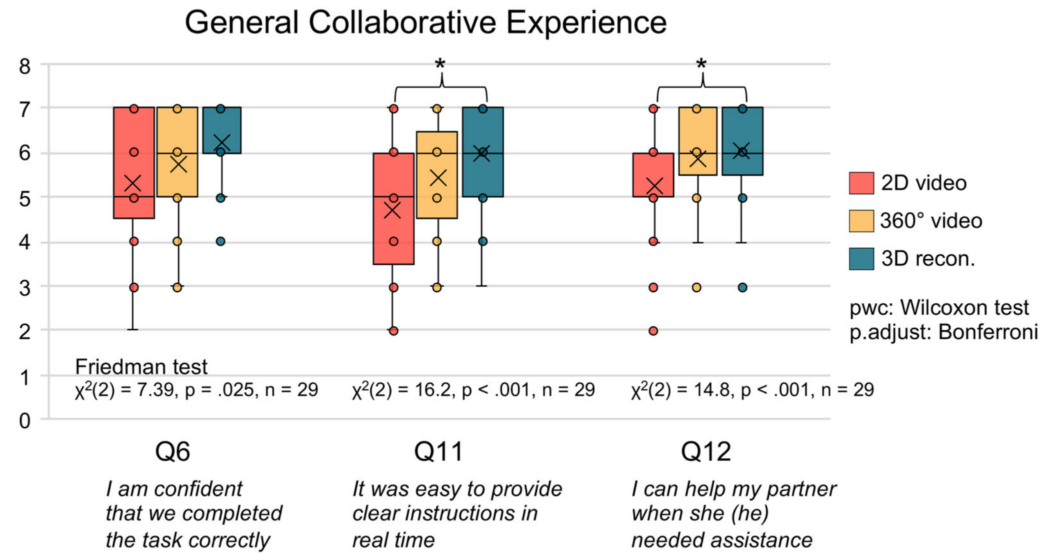


Figure 10. Results of general collaborative experience (*: statistically significant after Bonferroni correction).

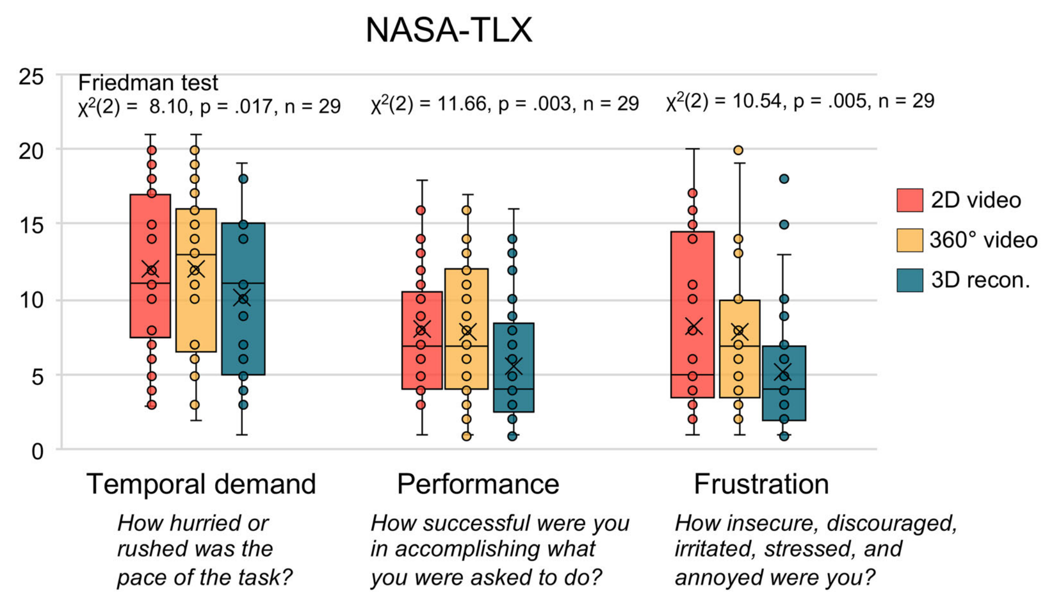


Figure 11. Results of NASA Task Load Index.

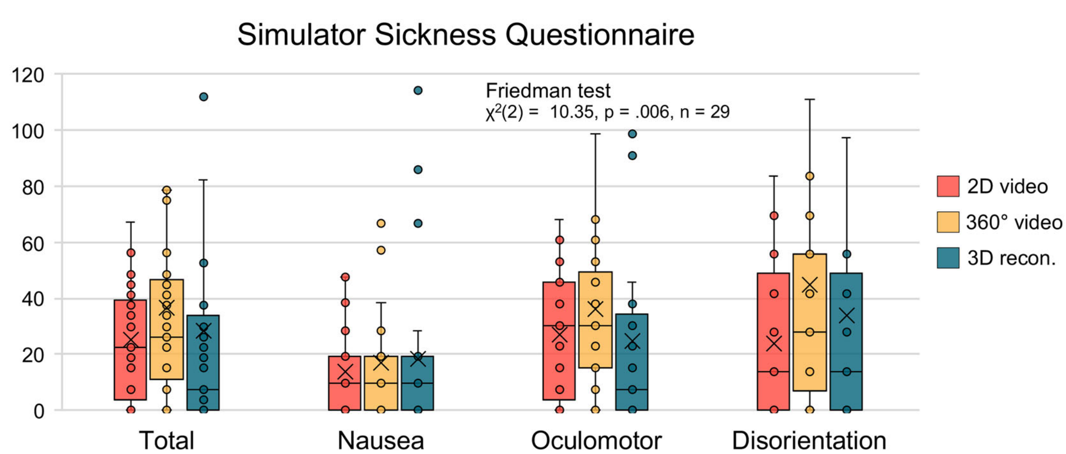


Figure 12. Results of Simulator Sickness Questionnaire.

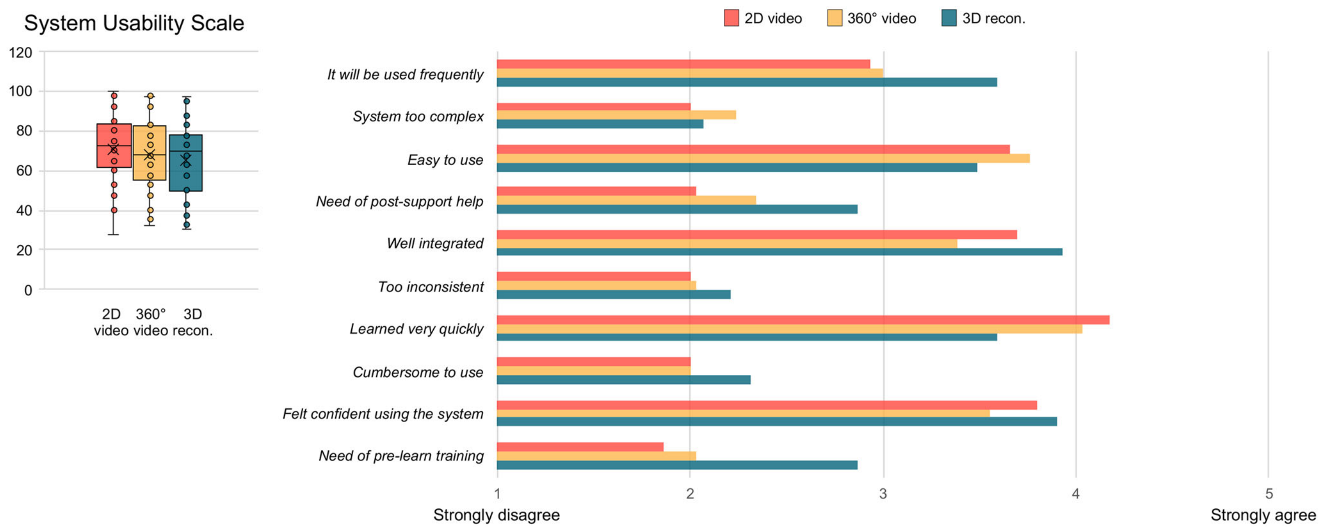


Figure 13. Results of System Usability Scale and user reports on each question.

Figure 14. (a) A radar graph of user preference of remote experts. (b) Rank distribution between interfaces according to remote experts (*: statistically significant).

often), and familiarity with using VR devices (0: never to 5: very familiar). The mean MSSQ score was 10.04 (± 10.58). More than half of the participants (15/29) reported a score lower than 3, indicating that they did not have much experience with VR (Figure 7). Though seven participants reported that they often experienced VR (score 4), only one participant regarded himself as skillful in handling the devices. Most participants felt that they were unfamiliar with using VR devices.

4.2. Context-sharing effect on remote collaboration

4.2.1. Task performance

The Friedman test was used to examine the effect of context-sharing interfaces (2D vs. 360° video vs. 3D reconstruction view) on either task completion time per sub-task or moving distance. We found a significant difference in completion time per sub-task depending on which type of interface was used by the remote expert ($\chi^2(2) = 8.37$, $p = 0.015$) (Figure 8). Post hoc analysis using Wilcoxon signed-rank tests was performed with Bonferroni correction ($p < 0.017$). However, the results did not show a significant difference between each pair. The order of average completion time per sub-task according to the interface was the 3D reconstruction view (62.4 ± 15.3 s), 2D (69.2 ± 19.7 s), and 360° video interface (74.2 ± 20.7 s), respectively.

Moreover, the results showed a significant interface effect on moving distance ($\chi^2(2) = 9.17$, $p = 0.010$). The total distance covered by a local worker during remote collaboration differed significantly depending on the expert's interface (Figure 9). Post hoc analysis indicated that the local moving distance in the 3D condition was significantly lower than that in the 360° video condition ($Z = 3.13$, $p = 0.002$).

4.2.2. Questionnaires

In line with the objective metrics, Friedman tests were performed for each questionnaire. For the GCE, we found that the interface had a significant effect on questions 6, 11, and 12 (Figure 10). Participants reported they were more confident about task completion when they used the 3D reconstruction view interface (Q6; $\chi^2(2) = 7.39$, $p = 0.025$). Moreover, participants in the 3D reconstruction view condition felt that it was easier to provide instructions to a local worker in real time (Q11; $\chi^2(2) = 16.2$, $p < 0.001$). In particular, post hoc analysis revealed a statistically significant decrease in providing real-time instructions in the 2D vs. 3D conditions ($Z = -3.28$, $p = 0.001$). Lastly, the 3D reconstruction view condition demonstrated the highest score in terms of giving immediate feedback to the partner (Q12; $\chi^2(2) = 14.8$, $p < 0.001$). Post hoc analysis showed that participants regarded the 3D condition as more helpful than the 2D condition ($Z = -2.64$, $p = 0.008$).

From the NASA-TLX results, we found a significant interface effect in the temporal demand, performance, and frustration subscales (Figure 11). Participants showed a clear

Figure 15. A screenshot of P16's viewpoint during collaboration. P16 flew up immediately after starting the experiment to scan the overall space.

difference in the level of effort required to achieve the task goal according to the interface types (temporal demand; $\chi^2(2) = 8.10$, $p = 0.017$). Remote experts using the 3D interface felt more successful in accomplishing the work (performance; $\chi^2(2) = 11.66$, $p = 0.003$). Note that a lower score on the performance subscale indicates better performance (1: perfect to 21: failure). In addition, participants showed different levels of stress depending on the type of interface (frustration; $\chi^2(2) = 10.54$, $p = 0.005$). Post hoc analysis on each subscale did not show a significant difference between the pair of interfaces.

Regarding physical discomfort during collaboration, participants showed a significant difference in the level of oculomotor-related symptoms depending on the interfaces ($\chi^2(2) = 10.35$, $p = 0.006$) (Figure 12). Remote experts showed a lower SSQ-O score when using the 3D reconstruction view interface. However, post hoc analysis did not show a significant difference between each pair of interfaces. The other subscales did not show any statistically significant differences among the conditions.

In terms of the usability of each interface, we did not find a significant difference between the conditions ($\chi^2(2) = 1.93$, $p = 0.38$) (Figure 13). The average SUS scores for each interface were 70.9 (2D), 67.7 (360° video), and 65.4 (3D reconstruction views).

4.2.3. Preference rank

We performed a chi-squared test to determine the preference rank of the three different interfaces. The results showed that participants exhibited a significant difference in preference for the type of interface, except for understanding their partner's focus (Figure 14). In terms of communication, remote experts showed the strongest favor for the 3D condition (Q1; $\chi^2(4) = 20.69$, $p < 0.001$). Moreover, participants ranked the 3D interface first regarding guidance (Q4; $\chi^2(4) = 17.59$, $p = 0.001$), spatial awareness (Q5; $\chi^2(4) = 53.79$, $p < 0.001$), task completion (Q6; $\chi^2(4) = 79.03$, $p < 0.001$), and overall preference (Q7; $\chi^2(4) = 25.66$, $p < 0.001$). For the feeling of being together, the 360° video condition showed the strongest preference (Q3; $\chi^2(4) = 12.0$, $p = 0.017$).

5. Discussion

We observed a significantly better quality of collaborative performance using a 3D reconstruction view interface from both objective and subjective perspectives. The remote experts reduced the task completion time when they viewed the 3D reconstructed local space in a target-finding-and-placing scenario. Moreover, local workers exhibited fewer physical movements when the expert provided the target direction using the 3D interface. In line with the objective results, the subjective reports based on questionnaires showed a clear preference for the 3D reconstruction viewing condition. Compared to the 2D and 360° video interfaces, remote experts indicated a better collaborative experience, lower task demand, and lower level of eye-related discomfort in the 3D reconstruction view condition.

5.1. Context-sharing interfaces

In this research, we found a significantly faster completion time per task in the 3D reconstruction view interface than in the other interfaces. The local workers completed the task (finding waffle blocks and moving them to the target location) 12 s faster on average in the 3D interface than in the

360° video interface. Moreover, the local workers saved more physical effort in completing the collaborative goal, thereby indicating the lowest moving distance in the 3D condition. This result may have originated from differences in the manner in which the spatial information of the local workspace is acquired using each interface. That is, while a remote expert should ask a local worker to navigate to a specific location to obtain exact context information through the 2D or 360° video interfaces, the 3D interface enables the expert to search the area by themselves using VR controllers. Owing to this difference in the system, the expert, rather than the worker, could efficiently identify the spatial information and provide clear directions.

These results can be interpreted in terms of view independence. According to Tait and Billinghurst (2015), the viewpoint of the remote expert can be broadly categorized as either dependent or independent of the local worker, and experiments showed that a higher level of view independence led to a shorter completion time. Similarly, Kim et al. (2018) found that independent views provided several benefits to the remote expert by enabling the expert to freely explore the local workspace. In our study, only the 3D interface offered the remote expert an independent view, which resulted in a better collaboration experience for the user.

Subjective reports also correspond to objective measures. In the GCE scores, the participants felt more confident about task completion when using the 3D interface. This result could be attributed to the straightforward (GCE Q11) and timely (GCE Q12) instructions provided in the 3D reconstruction view conditions, where the two scores were significantly higher for the 3D condition than for the 2D video interface. These results suggest that it may be better to provide a 3D reconstructed workspace rather than a 2D video for a remote expert to guide a local worker.

Regarding task demand, the participants reported a lower level of temporal demand when using the 3D interface. This suggests that the 3D interface helped the participants better understand the local workspace and find targets in time. Moreover, in line with the GCE results, the participants felt more successful in accomplishing the task when using the 3D interface. This confidence could lead to less frustration during collaboration when viewing a 3D reconstructed space. According to our interviews conducted after the experiment, several participants intuitively obtained the gist of the 3D interface and reported their strategy as follows (Figure 15).

> To me, it was much easier to give instructions when using the 3D interface because I didn't need to imagine the workspace by myself, but the system gave me the exact spatial information. Therefore, I don't need to remember the locations of each waffle block. (P10)

> I preferred the 3D reconstruction view interface because I could zoom out the reconstructed space, so it was easier to identify the locations of the waffles. Also, this approach made me less sick since I didn't need to move a lot to figure out the entire workspace. (P15)

> I used to play Sudden Attack [a 3D first-person shooting game], so I noticed that it would be more efficient to find target objects if I fly up like a bird. (P16)

Interestingly, the participants exhibited fewer eye-related symptoms when viewing the 3D interface. Previous studies have shown that controllability in VR can affect the level of cybersickness (Dong & Stoffregen, 2010; Dong et al., 2011). When a user loses controllability of motion in a virtual environment, they are more likely to experience severe discomfort because the discrepancy between what they expect to see and what they actually experience increases. In this research, the 2D and 360° video interfaces were subjected to the local worker's movements. Therefore, the remote expert could experience greater discomfort due to the motion blur of the local worker. However, in a 3D reconstruction view interface, the remote expert can voluntarily navigate the 3D reconstructed space, which provides higher controllability in VR. This might have induced a significant difference in the SSQ-O scores between the view interfaces. In addition, the participants spent less time completing the search task using the 3D interface. It is well known that a longer VR experience can increase the risk of becoming sick (Chang et al., 2020; Liu & Uang, 2012). Thus, a shorter task completion time in the 3D reconstruction view interface might lessen the chance of experiencing severe cybersickness.

However, regarding the system usability, we did not find significant differences between the interfaces. Specifically, although the highest score of the SUS Q1 (It will be used frequently) was reported in the 3D interface, the results of Q4 (Need of post-support help) and Q10 (Need of pre-learn training) indicated that participants considered the 3D interface more complex than the other two interfaces. These findings may have resulted from the characteristics of the recruited participants. According to the pre-questionnaire data, while the participants showed a wide range of prior VR experience (from never used to frequently used), they regarded themselves as novices in using VR devices. Although the 3D interface provided a better collaborative experience and outcomes, handling the VR controllers may have burdened some participants. Future studies should consider designing more accessible manipulation manuals and providing additional tutorials to help users become accustomed to the system.

Lastly, we found a clear preference for the 3D reconstruction view interface among remote experts. Participants reported the highest ranks in the communication, guiding, spatial awareness, and task completion subscales in the 3D condition. However, this interface did not surpass the others regarding understanding the partner's focus and co-presence. In the 2D and 360° video conditions, the remote expert can acquire spatial information about the workspace by moving the local worker's camera. In contrast, in the 3D reconstruction view interface, the expert already had reconstructed spatial information, so there might have been less chance of interaction between the two collaborators. In addition, we only presented the head of the local worker's avatar in the 3D condition, which might have been less likely to induce a feeling of being together. Similar to the study by Xu et al. (2019), we can mount additional cameras to track the local worker's body and implement a whole-body avatar for future research.

5.2. Limitations

According to our interview, one participant (P18) mentioned a tracking error in the 3D reconstruction view condition, which might have resulted from the misinterpretation of the sensor data from the AR tablet. Since only a limited part of the avatar was represented (i.e., head), a few participants tended to confuse the direction of the partner.

> When I watched my partner's avatar at the corner, half of the head was hidden in the wall, so it was hard to understand the partner's location. Also, it was difficult to provide instructions on how to pick up the block in that situation.

We also observed that some participants had difficulty handling the VR controllers. Although we provided a tutorial for the 3D reconstruction view interface, more practice may be required to familiarize the participants with the device. Future studies should consider providing an extended tutorial or implementing a more intuitive manipulation interface.

Though we adopted the actor-participant pair design to control unexpected proactive behaviors, this approach might reduce dynamic communication between collaborators. Considering the real-world collaboration scenario, further studies will include a participant-participant pair design to confirm the view interface effect.

6. Conclusion and future work

This research contributes to demonstrating the effects of context-sharing interfaces on remote collaboration performance. By improving the WXR system, this study provides three types of view interfaces (2D video, 360° video, and 3D reconstruction view). Each interface reflects the real-time updates of the local workspace and provides flexibility in viewpoint changes. In addition, collaboration was conducted in a spacious local workspace with a complex structure to clarify the effect of the interface on user performance. Based on the experiment, we found that the 3D reconstruction view interface yielded faster completion time, required lower physical demand, and was overall preferred. These results suggest that if all types of view interfaces can provide real-time updates of spatial information, the 3D reconstruction view might enhance task performance in terms of both objective and subjective collaboration experiences.

In future studies, we will add various visual annotations to support non-verbal communication. For example, visualizing the upper body of the avatar (Ahuja et al., 2019) will help clearly indicate the direction where the partner is looking. Sharing gazes or hand gestures would also be a promising approach to improving the quality of remote collaboration. Moreover, we plan to introduce a VR sickness reduction strategy (Lim et al., 2021), which dynamically varies the field of view, along with a reliable prediction method (Chang et al., 2021) to enhance the remote expert's experience. In addition, we will apply the present XR system to other collaborative scenarios, such as object assembly, and implement it in AR glasses to demonstrate its effectiveness in various remote collaboration contexts.

Notes

1. W3C Immersive Web Working Group. Retrieved April 3, 2023, from https://www.w3.org/immersive-web/
2. Apple Inc. *iPad Pro*. Retrieved April 3, 2023, from https://www.apple.com/ipad-pro/
3. Meta Inc. *Meta Quest 2*. Retrieved April 3, 2023, from https://www.meta.com/quest/products/quest-2/
4. Insta360 Inc. *Insta360 ONE X*. Retrieved April 3, 2023, from https://www.insta360.com/product/insta360-onex/
5. Laan Labs. *3D Scanner App*. Retrieved April 3, 2023, from https://3dscannerapp.com/

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Industrial Technology Innovation Program [20012462] funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and the National Research Foundation of Korea (NRF) Grant [NRF-2021R1A2C2093065 and NRF-2021R1A6A3A14039652] funded by the Korea government (MSIT).

ORCID

Eunhee Chang http://orcid.org/0000-0001-9348-0602
Yongjae Lee http://orcid.org/0000-0003-1692-2117
Byounghyun Yoo http://orcid.org/0000-0001-9299-349X

Data availability statement

The data that support the findings of this study are openly available in "Supplementary Data: A User Study on the Comparison of View Interfaces for VR-AR Communication in XR Remote Collaboration" at http://doi.org/10.17632/ykkbth5yyn.4

References

Adcock, M., Anderson, S., & Thomas, B. (2013). RemoteFusion: Real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM Siggraph International Conference on Virtual-Reality Continuum and Its Applications in Industry* (pp. 235–242). Association for Computing Machinery. https://doi.org/10.1145/2534329.2534331

Ahuja, K., Harrison, C., Goel, M., & Xiao, R. (2019, 10). MeCap: Whole-Body Digitization for Low-Cost VR/AR Headsets. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (pp. 453–462). Association for Computing Machinery. https://doi.org/10.1145/3332165.3347889

Anton, D., Kurillo, G., & Bajcsy, R. (2018). User experience and interaction performance in 2D/3D telecollaboration. *Future Generation Computer Systems*, *82*, 77–88. https://doi.org/10.1016/j.future.2017.12.055

Bai, H., Sasikumar, P., Yang, J., & Billinghurst, M. (2020). A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376550

Bauer, M., Kortuem, G., & Segall, Z. (1999). "Where are you pointing at?" A study of remote collaboration in a wearable videoconference system. In *Digest of Papers. Third International Symposium on Wearable Computers* (pp. 151–158). IEEE. https://doi.org/10.1109/ISWC.1999.806696

Bottecchia, S., Cieutat, J.-M., Merlo, C., & Jessel, J.-P. (2009). A new AR interaction paradigm for collaborative teleassistance system: The POA. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *3*(1), 35–40. https://doi.org/10.1007/s12008-008-0051-7

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In *Usability evaluation in industry* (1st ed., pp. 189–194). CRC Press. https://doi.org/10.1201/9781498710411

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual reality sickness: A review of causes and measurements. *International Journal of Human–Computer Interaction*, *36*(17), 1658–1682. Retrieved from https://doi.org/10.1080/10447318.2020.1778351

Chang, E., Kim, H. T., & Yoo, B. (2021). Predicting cybersickness based on user's gaze behaviors in HMD-based virtual reality. *Journal of Computational Design and Engineering*, *8*(2), 728–739. Retrieved from https://doi.org/10.1093/jcde/qwab010

Choi, S. H., Kim, M., & Lee, J. Y. (2018). Situation-dependent remote AR collaborations: Image-based collaboration using a 3D perspective map and live video-based collaboration with a synchronized VR mode. *Computers in Industry*, *101*, 51–66. https://doi.org/10.1016/j.compind.2018.06.006

Dong, X., & Stoffregen, T. A. (2010). Postural activity and motion sickness among drivers and passengers in a console video game. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, pp. 1340–1344). SAGE Publications. https://doi.org/10.1177/154193121005401808

Dong, X., Yoshida, K., & Stoffregen, T. A. (2011). Control of a virtual vehicle influences postural activity and motion sickness. *Journal of Experimental Psychology. Applied*, *17*(2), 128–138. https://doi.org/10.1037/a0024097

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., & Billinghurst, M. (2019). Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, *131*, 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011

Fussell, S. R., Setlock, L. D., & Kraut, R. E. (2003). 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). Association for Computing Machinery. https://doi.org/10.1145/642611.642701

Gao, L., Bai, H., Billinghurst, M., & Lindeman, R. W. (2020). User behaviour analysis of mixed reality remote collaboration with a hybrid view interface. In *32nd Australian Conference on Human-Computer Interaction* (pp. 629–638). Association for Computing Machinery. https://doi.org/10.1145/3441000.3441038

Gao, L., Bai, H., He, W., Billinghurst, M., & Lindeman, R. W. (2018). Real-time visual representations for mobile mixed reality remote collaboration. In *Siggraph Asia 2018 Virtual & Augmented Realitycharacterizing Asymmetric Collaborative Interactions in Virtual and Augmented* (pp. 1–2). Association for Computing Machinery. https://doi.org/10.1145/3275495.3275515

Gao, L., Bai, H., Lee, G., & Billinghurst, M. (2016). An oriented point-cloud view for MR remote collaboration. In *Siggraph Asia 2016 Mobile Graphics and Interactive Applications* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/2999508.2999531

Gao, L., Bai, H., Lindeman, R., & Billinghurst, M. (2017, 11). Static local environment capturing and sharing for MR remote collaboration. In *Siggraph Asia 2017 Mobile Graphics & Interactive Applications* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3132787.3139204

Gauglitz, S., Nuernberger, B., Turk, M., & Höllerer, T. (2014). World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (pp. 449–459). Association for Computing Machinery. https://doi.org/10.1145/2642918.2647372

Grandi, J. G. J. G., Debarba, H. G., & Maciel, A. (2019, 3). Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE Conference on Virtual Reality and 3d User Interfaces (VR)* (pp. 127–135). IEEE. https://doi.org/10.1109/VR.2019.8798080

Gurevich, P., Lanir, J., Cohen, B., & Stone, R. (2012). TeleAdvisor: A versatile augmented reality tool for remote assistance. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems* (pp. 619–622). Association for Computing Machinery. https://doi.org/10.1145/2207676.2207763

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Higuch, K., Yonetani, R., & Sato, Y. (2016, 5). 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* (pp. 5180–5190). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858438

Jing, A., May, K. W., Naeem, M., Lee, G., & Billinghurst, M. (2021). EyemR-Vis: Using bi-directional gaze behavioural cues to improve mixed reality remote collaboration. In *Extended Abstracts of the 2021 Chi Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3411763.3451844

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Kikuchi, N., Fukuda, T., & Yabuki, N. (2022). Future landscape visualization using a city digital twin: Integration of augmented reality and drones with implementation of 3D model-based occlusion handling. *Journal of Computational Design and Engineering*, *9*(2), 837–856. https://doi.org/10.1093/jcde/qwac032

Kim, S., Billinghurst, M., & Lee, G. (2018). The effect of collaboration styles and view independence on video-mediated remote collaboration. *Computer Supported Cooperative Work (CSCW)*, *27*(3–6), 569–607. https://doi.org/10.1007/s10606-018-9324-2

Kim, S., Jing, A., Park, H., Lee, G. A., Huang, W., & Billinghurst, M. (2020). Hand-in-Air (HiA) and Hand-on-Target (HoT) style gesture cues for mixed reality collaboration. *IEEE Access*. *8*, 224145–224161. https://doi.org/10.1109/ACCESS.2020.3043783

Lee, G., Kang, H., Lee, J., & Han, J. (2020, 3). A user study on view-sharing techniques for one-to-many mixed reality collaborations. *In 2020 IEEE Conference on Virtual Reality and 3d User Interfaces (VR)* (pp. 343–352). IEEE. https://doi.org/10.1109/VR46266.2020.1581166222244

Lee, G. A., Teo, T., Kim, S., & Billinghurst, M. (2017, 11). Mixed reality collaboration through sharing a live panorama. In *Siggraph Asia 2017 Mobile Graphics & Interactive Applications* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3132787.3139203

Lee, J.-K., Lee, S., Kim, Y-c., Kim, S., & Hong, S.-W. (2023). Augmented virtual reality and 360 spatial visualization for supporting user-engaged design. *Journal of Computational Design and Engineering*, *10*(3), 1047–1059. https://doi.org/10.1093/jcde/qwad035

Lee, Y., Moon, C., Ko, H., Lee, S.-H., & Yoo, B. (2020, 11). Unified representation for XR content and its rendering method. In *The 25th International Conference on 3d Web Technology* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3424616.3424695

Lee, Y., & Yoo, B. (2021). XR collaboration beyond virtual reality: Work in the real world. *Journal of Computational Design and Engineering*, *8*(2), 756–772. https://doi.org/10.1093/jcde/qwab012

Lee, Y., Yoo, B., & Lee, S.-H. (2021). Sharing ambient objects using real-time point cloud streaming in web-based XR remote collaboration. In *The 26th International Conference on 3d Web Technology*. Association for Computing Machinery. https://doi.org/10.1145/3485444.3487642

Li, Z., Teo, T., Chan, L., Lee, G., Adcock, M., Billinghurst, M., & Koike, H. (2020). OmniGlobeVR: A collaborative 360-degree communication system for VR. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 615–625). Association for Computing Machinery. https://doi.org/10.1145/3357236.3395429

Lim, K., Lee, J., Won, K., Kala, N., & Lee, T. (2021). A novel method for VR sickness reduction based on dynamic field of view processing. *Virtual Reality*, *25*(2), 331–340. https://doi.org/10.1007/s10055-020-00457-3

Liu, C.-L., Uang, S.-T. (2012). A study of sickness induced within a 3D virtual store and combated with fuzzy control in the elderly. In *2012 9th*

International Conference on Fuzzy Systems and Knowledge Discovery (pp. 334–338). IEEE. https://doi.org/10.1109/FSKD.2012.6234149

Marcos, D., McCurdy, D., & Ngo, K. (2023). A-frame. https://aframe.io

Piumsomboon, T., Day, A., Ens, B., Lee, Y., Lee, G., & Billinghurst, M. (2017, 11). Exploring enhancements for remote mixed reality collaboration. In *Siggraph Asia 2017 Mobile Graphics & Interactive Applications* (pp. 1–5). Association for Computing Machinery. https://doi.org/10.1145/3132787.3139200

Piumsomboon, T., Dey, A., Ens, B., Lee, G., & Billinghurst, M. (2019). The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, *6*, 5. https://doi.org/10.3389/frobt.2019.00005

Piumsomboon, T., Lee, G. A., Irlitti, A., Ens, B., Thomas, B. H., & Billinghurst, M. (2019, 5). 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* (pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300458

Qiu, X., Yang, Z., Yang, J., Wang, Q., & Wang, D. (2023). Impact of AR navigation display methods on wayfinding performance and spatial knowledge acquisition. *International Journal of Human–Computer Interaction*, 1–21. https://doi.org/10.1080/10447318.2023.2169524

Speicher, M., Cao, J., Yu, A., Zhang, H., & Nebeling, M. (2018). 360 Anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, *2*(EICS), 1–20. https://doi.org/10.1145/3229091

Stotko, P., Krumpen, S., Hullin, M. B., Weinmann, M., & Klein, R. (2019). 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. http://arxiv.org/abs/1805.03709 https://doi.org/10.1109/TVCG.2019.2899231

Tait, M., & Billinghurst, M. (2015, 12) The effect of view independence in a collaborative AR system. *Computer Supported Cooperative Work (CSCW)*, *24*(6), 563–589. https://doi.org/10.1007/s10606-015-9231-8

Tecchia, F., Alem, L., & Huang, W. (2012). 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). Association for Computing Machinery. https://doi.org/10.1145/2407516.2407590

Teo, T., Lawrence, L., Lee, G. A., Billinghurst, M., & Adcock, M. (2019, 5). Mixed reality remote collaboration combining 360 video and 3D reconstruction. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300431

Teo, T., Norman, M., Lee, G. A., Billinghurst, M., & Adcock, M. (2020). Exploring interaction techniques for 360 panoramas inside a 3D reconstructed scene for mixed reality remote collaboration. *Journal on Multimodal User Interfaces*, *14*(4), 373–385. https://doi.org/10.1007/s12193-020-00343-x

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Wei, S., Xu, G., He, W., Zhang, X., & Zhang, J. (2021). 3DGAM: Using 3D gesture and CAD models for training on mixed reality remote collaboration. *Multimedia Tools and Applications*, *80*(20), 31059–31084. https://doi.org/10.1007/s11042-020-09731-7

Wang, Z., Zhang, S., & Bai, X. (2021). A mixed reality platform for assembly assistance based on gaze interaction in industry. *The International Journal of Advanced Manufacturing Technology*, *116*(9–10), 3193–3205. https://doi.org/10.1007/s00170-021-07624-z

World Wide Web Consortium (2022). WebXR device API. https://immersive-web.github.io/webxr/

Xu, W., Chatterjee, A., Zollhöfer, M., Rhodin, H., Fua, P., Seidel, H.-P., & Theobalt, C. (2019). Mo2Cap2: Real-time mobile 3D motion capture with a cap-mounted fisheye camera. *IEEE Transactions on Visualization and Computer Graphics*, *25*(5), 2093–2101. https://doi.org/10.1109/TVCG.2019.2898650

About the authors

Eunhee Chang is currently a postdoctoral research fellow at the Empathic Computing Laboratory in the University of South Australia. She received her PhD degree in the Department of Psychology from Korea University. Her research interests include cybersickness, self-perception in VR, and human-computer interaction.

Yongjae Lee is currently a first-year PhD student at the Efficient, Secure and Intelligent Computing Laboratory at Arizona State University. He explores cutting-edge technologies to provide humans with a realistic virtual experience. His research areas encompass VR, AR, and Neural Rendering.

Byounghyun Yoo is a principal research scientist at the Korea Institute of Science and Technology. He received his PhD degree in the Department of Mechanical Engineering from the Korea Advanced Institute of Science and Technology. His research interests include VR, AR, and remote collaboration in extended reality environments.