Int. J. Human–Computer Studies 190 (2024) 103304

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Efficient VR-AR communication method using virtual replicas in XR remote collaboration

Eunhee Chang [a,1], Yongjae Lee [b,1], Mark Billinghurst [a], Byounghyun Yoo [c,d,*]

[a] Empathic Computing Lab., University of South Australia, Mawson Lakes Blvd, Adelaide, 5095, SA, Australia
[b] Department of Electrical and Computer Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, 21218, MD, United States
[c] Center for Artificial Intelligence, Korea Institute of Science and Technology, 5 Hwarangro14-gil, Seongbuk-gu, Seoul, 02792, Republic of Korea
[d] AI-Robotics, KIST School, Korea National University of Science and Technology, 5 Hwarangro14-gil, Seongbuk-gu, Seoul, 02792, Republic of Korea

ARTICLE INFO

Keywords:
Extended reality
Virtual reality
Augmented reality
Remote collaboration
Interaction interfaces
Virtual replicas

ABSTRACT

When using Virtual Reality (VR) and Augmented Reality (AR) to support remote collaboration, effective communication between a remote expert in VR and a local worker in AR is important for guiding and following task instructions. This is especially crucial for assembly tasks, which require precise identification of parts and clear directions for their combination. Despite the increasing interest in efficient VR-AR communication methods, previous studies have been limited to complex hardware setups and simplified assembly tasks. In this research, we introduce a communication approach for remote collaboration in complex assembly tasks, utilizing simplified hardware configurations. We conducted a user study (n = 30) and compared three interaction interfaces (hand gestures, 3D drawing, and virtual replicas) in task completion time, subjective questionnaires, and preference rank. The results showed that the use of virtual replicas not only enhances task efficiency but also receives strong preference by users. These findings indicate that virtual replicas can provide intuitive instructions to local workers, resulting in a clearer understanding of the expert's guidance.

1. Introduction

Remote collaboration involves the joint efforts of physically distant workers in accomplishing a shared task goal. Recent research on remote collaboration has focused on the potential of using eXtended Reality (XR), which encompasses Virtual Reality (VR) and Augmented Reality (AR) environments. XR remote collaboration includes manipulating virtual or real objects through VR and AR devices. Depending on the specific collaboration scenario, various combinations of realities, such as VR-VR, AR-AR, and VR-AR, can be employed (Ens et al., 2019; Lee and Yoo, 2021).

One of the well-known scenarios in XR remote collaboration is VR-AR communication where there is an asymmetry in knowledge levels between remote (usually on the VR side) and local (usually on the AR side) collaborators (Oda et al., 2015; Gao et al., 2016; Yu et al., 2021). For example, a remote expert in VR can guide a local worker in AR on the procedures for object assembly (Anton et al., 2018). In this type of collaboration, it is crucial to refer precisely to the objects required at each assembly step and to articulate the correct spatial orientation for their combination. Many studies have explored diverse communication methods in assembly scenarios such as hand gestures (Bai et al., 2020), three-dimensional (3D) drawing (Fakourfar et al., 2016), and virtual replicas (Oda et al., 2015).

Traditionally, hand gestures have been the dominant method of non-verbal communication for delivering messages. In XR remote collaboration, the remote expert's hand movements can be captured and perceived by a local worker through a 3D virtual hand model that copies the expert's hands in real time (Kim et al., 2020). Given that certain hand gestures are universally recognized for conveying specific intentions (e.g., pointing or giving a thumbs-up), various systems have employed predefined gesture animations to facilitate communication during collaboration (Le Chénéchal et al., 2016; Piumsomboon et al., 2018a,b).

Sketching methods have also been investigated as an intuitive tool to aid workers in object assembly. Using this approach, a remote expert in VR can directly highlight target areas by instantly generating shaped annotations such as arrows, lines, and circles (Fakourfar et al., 2016; Zillner et al., 2018). These annotations are superimposed onto a local worker's AR display, effectively directing the individual's attention

[*] Corresponding author at: Center for Artificial Intelligence, Korea Institute of Science and Technology, 5 Hwarangro14-gil, Seongbuk-gu, Seoul, 02792, Republic of Korea.
E-mail address: yoo@byoo.org (B. Yoo).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ijhcs.2024.103304
Received 4 October 2023; Received in revised form 6 May 2024; Accepted 3 June 2024
Available online 11 June 2024

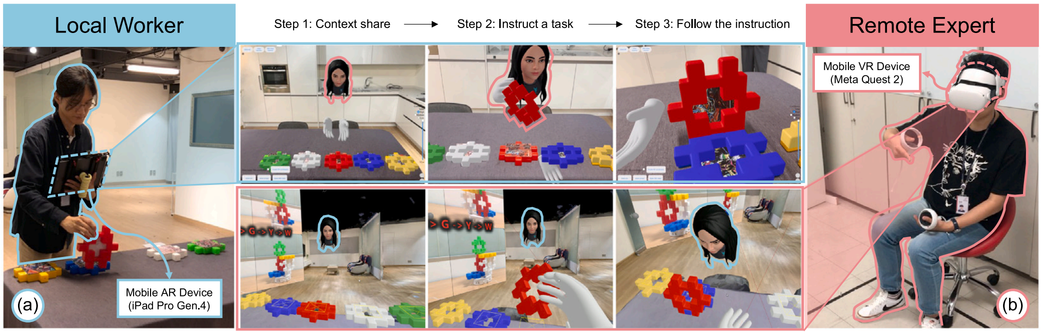


Fig. 1. An example of XR remote collaboration using the virtual replicas interface. A local worker (a) is assembling waffle blocks, following the remote expert's (b) instructions through each step.

to specific points in the real-world environment. For example, when it is necessary for the worker to connect objects, the remote expert can illustrate the correct angles by sketching. However, Katti et al. (2021), Xiang et al. (2021), and El Ammari and Hammad (2019) pointed out that poor drawing skills and the inaccuracy of superimposing virtual sketches can impair communication quality and task performance.

With advances in sensor and vision technology, recent works have adopted virtual replicas which are copies of corresponding physical objects. For example, Oda et al. (2015) enabled a remote expert in VR to explain the assembly process by manipulating virtual replicas. The local worker could then observe this as augmented graphics in their own real-world environment. In several cases, augmented graphics were either positioned close to their respective physical objects (Wang et al., 2021a) or directly overlaid upon them (Lee and Yoo, 2021). This approach has proven to enhance task performance by delivering work procedures in a highly intuitive manner.

Nevertheless, implementing these techniques in real-world scenarios presents several challenges. To provide a better remote collaboration experience, previous studies have required one or multiple cameras, projectors, and head-mounted display (HMD) devices, which is resource-intensive. Moreover, the assembly scenario was often relatively simplified using Lego or Tangram puzzle solving, so there was little room to identify the interaction interface effect on task performance.

In recognition of these gaps, we focus on the following points for research originalities: a streamlined hardware setup, three types of interaction interfaces, and a complex assembly scenario. In our study, using a previous framework (Lee and Yoo, 2021), we create a simpler hardware configuration within a unified software framework capable of accommodating three distinct interaction methods (hand gestures, 3D drawing, and virtual replicas). Employing this integrated system, we conducted a comparative analysis of each interaction method within the context of a complex assembly scenario (Fig. 1). The assessment encompassed an examination of both subjective and objective parameters, with the aim of identifying the optimal interaction interface for a better quality of VR-AR communication.

2. Related work

In recent years, research focusing on the application of XR collaboration technologies has notably increased across various domains, including healthcare (Gasques et al., 2021; Yu et al., 2021, 2022), education (Choi et al., 2017), creativity (D'Angelo and Begel, 2017; Radu et al., 2021), entertainment (Dey et al., 2017; Park et al., 2020), and industry (Aschenbrenner et al., 2018; Vidal-Balea et al., 2020; Wang et al., 2021c; Lee et al., 2023). Recent review papers (Schäfer et al., 2022; Ens et al., 2019; Wang et al., 2021b) have attempted to dissect and categorize XR collaboration systems from multiple perspectives, providing valuable insights into potential future directions.

Understanding the scenario in which an XR collaboration application is to be deployed is a pivotal aspect for the system designers. This enables them to pinpoint the essential technical features required by the given scenario. For instance, in scenarios focused on object-finding, the way in which remote experts can effectively perceive spatial information of a shared workspace is known to be one of the key factors impacting task performance (Gao et al., 2020; Li et al., 2020; Chang et al., 2023b; Cho et al., 2021). In brainstorming scenarios, the participants' ability to intuitively express ideas and experience a sense of co-presence can significantly influence the quality of the collaborative experience (Zillner et al., 2014; Higuchi et al., 2015; Orts-Escolano et al., 2016).

In assembly scenarios, remote collaborators typically report difficulty in maintaining attention on the object being combined and communicating about the assembly process. Prior research revealed that the level of co-presence can be improved by visualizing awareness cues (e.g., field of view, head direction, eye gaze) when handling the assembly objects (Piumsomboon et al., 2019; Jing et al., 2021). Additionally, non-verbal cues such as hand gestures, 3D drawings, and virtual replicas have been proposed to support communication about three-dimensional information for assembly. Yet, when assessing their effectiveness, it is not straightforward to gain insight into which is superior; they are often evaluated under task conditions with limited degrees of freedom (DOF) and continuity, constraints that reduce the number of possible assembly combinations and thereby simplify the tasks.

In the following sections and in Table 1, we have summarized previous studies that used and evaluated interaction methods in assembly scenarios.

2.1. Hand gestures

Conveying three-dimensional information is integral to guiding the assembly of complex objects. Hand gestures are one of the significant non-verbal communication methods for doing this. Gestures are particularly useful for conveying the remote expert's hand movements. Capturing these movements via video is a common practice (Speicher et al., 2018). In early studies, hand gestures were extracted from video frames using computer vision techniques and then superimposed onto a local worker's display or workspace (Pauchet et al., 2007; Izadi et al., 2007; Sakong and Nam, 2006; Higuch et al., 2016; Huang et al., 2019; Young et al., 2019). These studies showed that participants can perform tasks effectively, similarly to when collaborating in co-located situations. However, a notable limitation was the difficulty in indicating an object's vertical position (Sakong and Nam, 2006; Alem and Li, 2011).

Recently, innovative methods using an RGBD camera or a stereo camera to capture *live point clouds* of hand gestures have been explored. Research by Huang et al. (2013) and Gao et al. (2016) showed systems that can capture the remote expert's hand movements as

Table 1
A summary of related works investigating the effects of interaction interfaces on assembly tasks. The interaction interfaces are categorized as hand gestures (HDG), 3D drawing (3DD), or virtual replicas (VRP).

Reference	Task configuration			Display configuration		Input hardware configuration	
	Scenario (task target[a])	Translation DOF[b] (Continuity[c])	Rotation DOF[b] (Continuity[c])	Local	Remote	Local	Remote
Alem and Li (2011)	Assembly (lego block)	0	2 (Continuous)	2D monitor	2D monitor	2 cameras	2 cameras, keyboard, mouse
Kim et al. (2014)	Assembly (tangram+)	1 (Continuous)	1 (Continuous)	HMD (Vuzix Wrap 1200 DX-VR) or Tablet (Microsoft Surface Pro)	2D monitor	Camera (Logitech QuickCam)	Mouse
Wang et al. (2019)	Assembly (water pump)	0	0	Projector	HMD (HTC VIVE with base stations)	Camera (Basler)	Hand tracker (Leap Motion)
Zhang et al. (2022)	Assembly (lego block+) Assembly (vise)	2 (Discrete) 0	1 (Discrete) 0	HMD (HoloLens)	HMD (HTC Vive Pro 2)	RGBD camera (Intel RealSense D435i)	Hand tracker (Leap Motion)
Tian et al. (2023)	Assembly (soma cube+)	2 (Discrete)	3 (Discrete)	HMD (HoloLens 2)	HMD (Meta Quest)	3 RGBD cameras (Microsoft Azure Kinect)	VR controllers
Present study	Assembly (waffle block+)	3 (Discrete)	3 (Continuous)	Tablet (iPad Pro 4)	HMD (Meta Quest 2)	Tablet (Built-in sensors)	VR controllers

GCE: General Collaborative Experience, HMD: Head-Mounted Display, NASA-TLX: NASA Task Load Index, SSQ: Simulator Sickness Questionnaire, SUS: System Usability Scale.

[a] '+' mark indicates that the assembly parts have an undetermined mating condition, thus increasing the assembly complexity.

[b] The degrees of freedom are estimated based on the figures presented in the relevant paper and are characterized by the number of translational and rotational freedoms available.

[c] Continuity refers to whether the angles and positions for the mating are predetermined or not. Depending on the type of task target, parts can be positioned either discretely in predefined positions or continuously in unlimited cases.

Reference	Interaction type			Live update method		Conditions[d]	Exp. design[e]	Measures
	HDG	3DD	VRP	Local scene	Remote interaction			
Alem and Li (2011)	v			2D live video	HDG: 2D live video	*Hand gestures* vs. Pointer	P-P	Completion time Number of mistakes Perception of the quality of collaborative effort Satisfaction Preference
Kim et al. (2014)		v		2D live video	3DD: Drawing on a virtual plane	None (voice only) vs. *Pointer* vs. *Drawing*	P-P	Likert scale ratings Participants' activity Preference
Wang et al. (2019)	v	v		2D live video	HDG: Hierarchical skeletal mesh; 3DD: Drawing on an object's surface	3D drawing vs. *Hand gestures*	P-P	Completion time SUS Collaborative experience Preference
Zhang et al. (2022)	v		v	Live point cloud	HDG: Hierarchical skeletal mesh; VRP: Static 3D mesh+transformation	Hand gestures vs. *Hand gestures +virtual replicas*	P-P	Completion time GCE NASA-TLX Preference
Tian et al. (2023)		v	v	Static 3D mesh +live point cloud	3DD: Free drawing in 3D; VRP: Static 3D mesh+transformation	3D drawing vs. *Virtual replicas*	P-P	Social presence Spatial presence Completion time NASA-TLX SUS Preference
Present study	v	v	v	Static 3D mesh +transformation	HDG: Set of predefined gestures+transformation; 3DD: Free drawing in 3D; VRP: Static 3D mesh+transformation	Hand gestures vs. 3D drawing vs. *Virtual replicas*	A-P	Completion time GCE NASA-TLX SSQ SUS Preference

[d] An italic-font condition represents the proposed method in their paper.

[e] Experimental design: While P-P denotes a *participant-participant* pair for the respective roles of remote expert and local worker, A-P denotes an *actor-participant* pair.

point clouds, depicting the specific manipulations required for block assembly. This point cloud data is subsequently overlaid onto a local worker's see-through video. User studies affirmed a greater preference for hand gestures as task complexity increases. However, this method necessitates substantial data bandwidth and demands additional image processing for hand segmentation (Gao et al., 2016, 2017).

A different approach (Sodhi et al., 2013; Wang et al., 2014; Lee et al., 2017, 2018; Kim and Seo, 2023) is to use a hand tracker such as the Leap Motion, which uses optical sensor data to analyze the position of hand joints and produces *a hierarchical skeletal mesh* of finger bones. It is possible to recognize specific hand gestures through the relative positions of each finger joint. Sodhi et al. (2013) confirmed that even sharing of pointing gestures can increase communication efficiency and proposed a method to use the recognized hand gestures as an interaction interface for controlling 3D objects. Lee et al. (2017, 2018) reported that non-verbal communication, including gestures, aids collaborators in understanding the intentions of collaboration counterparts. Sharing gestures using hand trackers enables people to convey the natural movements of both hands during the collaboration. However, ensuring these gestures carry contextually appropriate meaning in physical tasks requires a sophisticated setup of the trackers and a hardware-specific calibration process (Lee et al., 2018; Sasikumar et al., 2019; Genest et al., 2013).

Interestingly, it appears that a precise virtual replication of remote expert's hand gestures is not essential in XR remote collaboration (Piumsomboon et al., 2019). Le Chénéchal et al. (2016) and Teo et al. (2022) have included a feature whereby pressing buttons on controllers results in a change of the hand shape into a predefined hand gesture. Taking a similar approach in our study, we provide a remote expert with *a set of predefined hand gestures*. During the collaboration, the expert is able to freely select and use the most suitable gestures from this set, depending on the task progress. This approach obviates the need for dedicated imaging sensors or trackers and facilitates comparative experiments using a standardized VR controller interface across all interface conditions.

2.2. 3D drawing

Drawing is a commonly utilized annotation method. Earlier research allowed remote experts to freely sketch curves on shared videos (Tang and Minneman, 1990; Fussell et al., 2004; Kim et al., 2013; Zillner et al., 2014). Given the widespread human familiarity with sketching or writing on flat surfaces, He et al. (2020) and Kasahara et al. (2012) provided a two-dimensional (2D) *virtual plane*, like a canvas or sketch board, even within 3D virtual environments. Depending on the task, a virtual plane can be pre-established at specific locations (Gauglitz et al., 2012, 2014; Fakourfar et al., 2016), or it can be dynamically generated within the local workspace by detecting flat surfaces using image feature analysis (Kim et al., 2014). To mitigate errors in annotations due to the local worker's movements while the remote expert is drawing, momentarily pausing the video frames has been suggested as a solution (Kim et al., 2015). However, if the virtual plane is situated at a distance from the target object, parallax problems can still occur, potentially leading to erroneous pointings when the local worker alters their viewpoint.

When a local worker is operating within a Spatial Augmented Reality (SAR) environment, a remote expert can communicate their own intentions more accurately by *drawing directly on real objects' surfaces* (Roo and Hachet, 2017). In SAR, virtual objects are projected on physical surfaces, thereby enabling drawings to adhere faithfully to the 3D contours of the physical objects (Palmer et al., 2007). If the projector is fixed, the expert's drawings remain steadfast and in their intended places (Wang et al., 2019; Adcock et al., 2013). However, if a mobile projector is mounted on the local worker's shoulder or arm, any movement by them may lead to inaccuracies in the projected drawings' positioning (Gurevich et al., 2012). To address this issue, Gunn and Adcock (2011) presented a technique that adjusts the rendering location of drawings in response to camera movement, thereby maintaining the original positions of the drawings irrespective of any movement by the local worker.

Presenting 3D drawings in their inherent form, as opposed to projecting them onto a 2D surface, offers a more comprehensive conveyance of information. For instance, Ogawa et al. (2005) devised a system that allowed a remote expert to place virtual line segments into the air, using the geometric information from a pre-created 3D model of the workspace. Further, Jo and Hwang (2013) introduced a method that generates 3D annotations using a tablet interface. This method maps the 2D coordinates of drawings on the video frame into the 3D space of the workspace, a process accomplished through sensor fusion. However, simple drawings might misrepresent the expert's intentions due to changes in how they appear based on the worker's viewpoint. Nuernberger et al. (2016) proposed a technique that recognizes targets of annotations and subsequently re-renders annotations in an undistorted manner in accordance with the worker's viewpoint change.

Prior research has explored diverse methods of generating drawings across different combinations of realities and hardware interfaces. However, creating drawings on a virtual plane or an object's surface does not always lead to significant improvements in task performance or usability, especially in assembly scenarios that necessitate conveying intricate, three-dimensional information (Wang et al., 2019). Moreover, creating high-quality drawings using a 2D interface, such as a tablet, can be challenging. Consequently, we have adopted a more effective approach of *free drawing in 3D*. We provide a remote expert with an interaction interface that leaves 3D lines in the trajectory of a VR controller tracked with 6 DOF, a technique similar to the approach used in Tian et al. (2023). These drawings are then accurately augmented at the expert's intended location within the local worker's workspace.

2.3. Virtual replicas

The most straightforward way to explain a complex assembly procedure is for an expert to directly demonstrate the process firsthand to a worker (Oda et al., 2015; Orts-Escolano et al., 2016). In remote collaboration, where collaborators are geographically separated, using virtual replicas of physical objects in the local workspace can effectively communicate task know-how (Tait and Billinghurst, 2015). One simple method of guiding the local worker using virtual replicas is to position these virtual replicas in their correct assembly locations (Bottecchia et al., 2010; Tait and Billinghurst, 2015). However, in the studies from Bottecchia et al. (2010) and Tait and Billinghurst (2015), the inherent limitations associated with using a desktop interface made it challenging for a remote expert to show free and natural movements of virtual replicas. In contrast, studies that provided a remote expert with an immersive VR experience (Elvezio et al., 2017; Oda et al., 2015; Piumsomboon et al., 2018b; Tian et al., 2023) found that the expert could simulate the entire work process in a detailed manner. Additionally, Okajima et al. (2009) reported that leaving an afterimage of the trajectory of virtual replicas can enhance the local worker's understanding of the path that the real objects should follow.

In certain circumstances, 3D CAD models of physical objects can serve as virtual replicas (Bottecchia et al., 2010; Wang et al., 2021a; Seo et al., 2021), but obtaining the 3D CAD models may not always be feasible. An alternative is to use a 3D scanner or modeling software to reconstruct the real object's 3D model before the collaboration begins (Yang et al., 2015). However, this approach has limitations such that it confines the scope of instructions to what can be represented using pre-existing 3D models. To circumvent this, Zillner et al. (2018) proposed a system providing an interface enabling a remote expert to instantaneously generate virtual replicas of any specified objects from the reconstructed 3D model of the local workspace.

Recently, Zhang et al. (2022) investigated the effectiveness of different interfaces in assembly tasks. They compared an interface that relied solely on shared gestures with another interface that integrated both gestures and virtual replicas. Their experimental results revealed that when virtual replicas were incorporated into the interface, participants not only performed better in their collaborative tasks but also expressed a more pronounced preference for this setup. Moreover, the study reported that when the replicas were available, participants found the task less difficult and the interface more useful. Similarly, Tian et al. (2023) conducted a comparative study between the virtual replica and

3D drawing methods. Consistent with Zhang et al. (2022)'s findings, participants achieved faster completion times and expressed higher satisfaction with the replica as compared to the drawing. In post-experiment interviews, they described VR-AR communication facilitated by virtual replicas as the most direct and unambiguous form of interaction.

2.4. Originality of this research

As outlined in Table 1, numerous prior studies have investigated effective interaction interfaces, relying on complex hardware configurations, often involving combinations of one or more cameras, projectors, and HMD devices. Such configurations render the systems resource-intensive, potentially impeding the broader adoption of these technologies in real-world applications. Moreover, the focus on simplified assembly scenarios has left a research gap concerning the effect of interaction interfaces on task performance in more intricate settings.

Addressing these limitations, our work focuses on the following novel points:

1. We have established a streamlined hardware setup consisting solely of a single tablet and a single HMD, both of which are supported by a unified software system (see Section 3).
2. This system seamlessly integrates three distinct interaction methodologies—hand gestures, 3D drawing, and virtual replicas—within a single architectural design (see Section 3).
3. Using this system, we compared each interaction method under a complex assembly scenario to identify the optimal interface for enhanced VR-AR communication (see Section 4).

3. System overview

In this paper, we extend our previous system (Lee and Yoo, 2021) by introducing three interaction interfaces: Hand gestures, 3D drawing, and virtual replicas. This improvement aims to enhance communication capability between users during remote collaboration, while also leveraging the flexibility and accessibility afforded by web-based platforms for streamlined experiment setup. Our extension utilizes two significant benefits from our base system. Firstly, unlike traditional tools like Unity (Unity Technologies, 2024), which support distinct control flows for VR and AR entities, forcing developers to duplicate VR and AR scenes for identical content, our system supports unified control flows across various realities (AR/VR/MR/XR). Secondly, our web-based system enables the deployment process to be simple, requiring only a webpage reload for users to retrieve updated systems.

The design objectives of our XR remote collaboration system—termed the Webized eXtended Reality (WXR) system—are twofold: (1) to provide a collaborative workspace where geographically distributed collaborators are connected and work together on complex physical tasks (Section 3.1), and (2) to enable the remote expert to use various interaction interfaces (Section 3.2). Note that, in addition to our previous system achieving the first objective, the new system incorporates the critical functionality of the second objective for our experiment: enabling local workers to receive information through various interaction interfaces. In this section, we describe in more detail the key features of our upgraded WXR system while ensuring a consistent user experience.

3.1. The WXR system

Fig. 2 shows an illustration of the overall configuration of the WXR system. At a high level, it supports common XR remote collaboration settings for physical tasks with independent viewing experiences (i.e., a remote expert independently navigates the local site and provides instructions through a mobile VR device, while a local worker receives the instructions from the expert via a mobile AR device).

To enable the remote expert to understand the local context, such as the object arrangement, we generate a virtual replica of the local site (i.e., mirrored world). This can be created through 3D modeling in advance or coarse scanning (Lee et al., 2021) just before collaboration. While previous studies (Gao et al., 2016; Bai et al., 2020; Lindlbauer and Wilson, 2018) prefer to employ live point cloud reconstruction during collaboration, we adopt a pre-mesh reconstruction approach because point cloud reconstruction requires large data bandwidth and high computation, which is unfavorable for most real-world practices.

In our mirrored world, we semantically divide objects into the foreground (target objects) and background (ambient objects). The foreground objects are the focus of collaboration and change state during the collaboration, while the background objects remain mostly static and provide spatial context to the remote expert. The 3D models of the target objects in the mirrored world are linked to their physical counterparts in the real world, mirroring their positions and orientations. This allows the remote expert to perceive real-time changes in the local context. The local worker uses their AR device to detect and track the foreground physical objects and streams this tracking information over the network.

On the local side, the 3D models of the target objects are augmented on the local worker's display. This feedback confirms to the local worker that the target objects' states are accurately shared. At the same time, the 3D models are used for conveying the remote expert's work information. For instance, when the remote expert demonstrates how to connect two target objects in the mirrored world by manipulating their corresponding 3D models using VR controllers, this action is also shown to the local worker as an animation, augmented onto the local worker's video frames.

Unlike other remote collaboration systems (Tian et al., 2023; Zhang et al., 2022), the WXR system focuses on the simple hardware configuration (see Table 1). Complex configurations often prove challenging to implement in real-world scenarios of remote collaboration; therefore, simpler configurations are considered more pragmatic and feasible. For example, systems such as Tian et al. (2023) entail reinstallation of a set of sensors and a cumbersome camera calibration process whenever collaborating in a new space.

In the local site, the worker uses a mobile AR device which supports a standard web browser capable of AR functions such as image tracking and device localization. Although popular see-through HMDs, such as the Microsoft HoloLens, are also available as long as they support standard web features (WHATWG, 2023a; W3C, 2022), we adopted a hand-held device for the local worker in our experiment because there have been reports (Fang et al., 2023) suggesting that prolonged use of HMD-based AR can cause discomfort to users, making it difficult for long-term use. Additionally, current cutting-edge see-through HMDs still have a narrow field of view (FOV).[2] Hence, to avoid any negative impact on the usability evaluation, we chose to use a hand-held device for the local worker.

To initiate collaboration, the local worker loads a web page shared with the remote expert. This web page hosted by a web server (the WXR Server) includes static assets for the collaboration (e.g., 3D models of foreground and background objects) as well as javascript client application code, called the WXR Library.

The rendering engine of the WXR Library initializes the XR scene, which is a common scene graph representing different realities (Lee et al., 2020), and renders it conditioned on the user's environment. Specifically, it renders only foreground objects in AR and foreground/background objects in VR. The WXR Library constantly communicates with the WXR Server to keep the XR scene up-to-date with the remote expert.

The remote expert uses a mobile VR device which supports web standards such as device localization and VR controller input. Likewise on the local side, the expert loads the shared web page, and the loaded application renders the same XR scene, providing the expert with a fully-immersive VR experience.

[2] Microsoft Hololens 2 has about 52 degrees of FOV (https://www.microsoft.com/en-us/hololens/hardware) whereas iPad Pro 4 has 125 degrees (https://support.apple.com/kb/SP882).

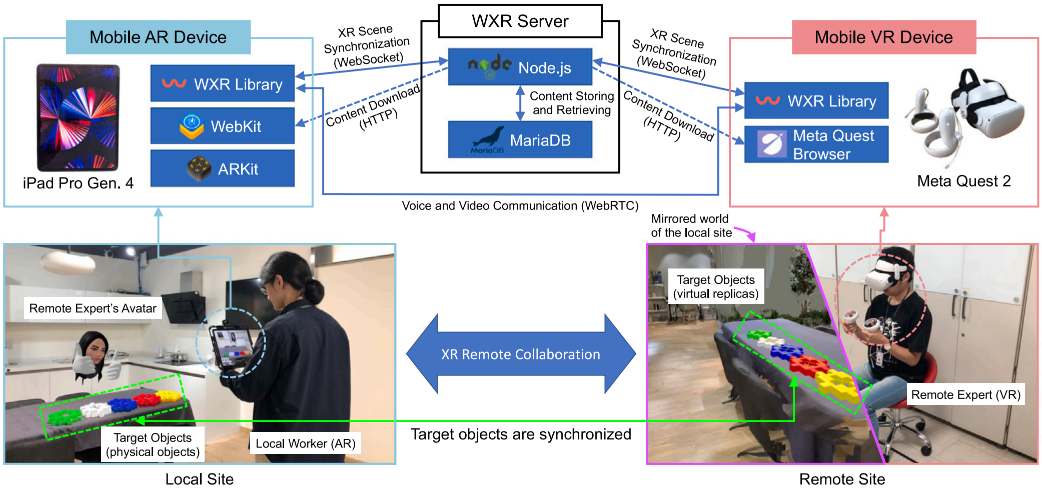
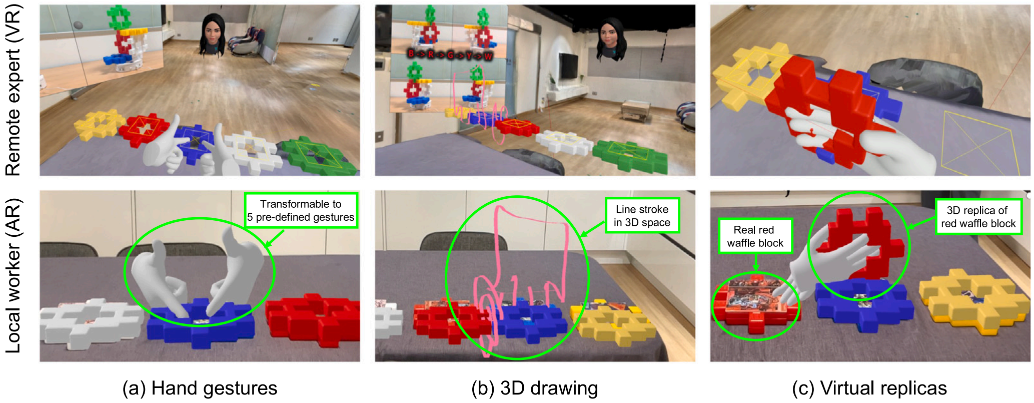


Fig. 2. Overview of the WXR System. The upper row shows the architecture of the WXR system and the lower row shows the common setting of XR remote collaboration. Note that the collaborators are equipped with widely used off-the-shelf products: An iPad for the local worker and the Meta Quest 2 for the remote expert.



(a) Hand gestures (b) 3D drawing (c) Virtual replicas

Fig. 3. Three interaction interfaces viewed from a remote expert (upper row) and a local worker (lower row): (a) hand gestures, (b) 3D drawing, and (c) virtual replicas.

3.2. The interaction interfaces

Our system provides not only full-duplex voice communication but also three interaction interfaces (Fig. 3): Hand gestures, 3D drawing, and virtual replicas. The rationale underlying the development of these interfaces is centered on the challenge of drawing conclusions regarding interface superiority based on previous studies, which often employed disparate hardware configurations. Each of the aforementioned interfaces affords unique visual cues, necessitating diverse technical implementations. For example, hand gestures were captured through the camera (Alem and Li, 2011), 3D drawing was generated through the mouse (Fakourfar et al., 2016), and replicas were manipulated through the VR controllers (Tian et al., 2023). Consequently, the observed interface effects in prior studies may be influenced by variations in hardware configuration, thereby complicating direct comparisons.

In this study, we focus on ensuring a consistent user experience across various interaction interfaces within the system. The functions of each interface are mapped to different buttons on the VR controller, allowing the remote expert to freely use all three interfaces within a single session. These interactions are synchronized and displayed in real time on the local worker's AR device.

The **Hand Gestures** interface permits the remote expert to alter the shape of their virtual hands. By holding down specific buttons on the VR controller, the expert can transition from the default gesture (Open) to one of five predefined gestures (Fig. 4). During the collaboration, the remote expert provides the local worker with concrete information by forming an intended gesture alongside the ambiguous verbal information (e.g., pointing at a specific part of the target object while saying "here"). This method is intuitive and straightforward, but since there is no common protocol for gesture interpretation across all societies and cultures, its meaning can easily become unclear.

The **3D Drawing** interface allows the remote expert to leave 3D lines and sketches in the air. These lines are created when the remote expert holds down the trigger button on the VR controller, thus they can generate free-formed 3D annotations such as arrows, circles, or even a draft of the completed shape of target objects. One advantage of this method is that it enables the remote expert to generate on-the-fly annotations during collaboration. However, the limitation is that the communicative power of the annotations is heavily dependent on the drawing skills of the remote expert.

The **Virtual Replicas** interface enables the remote expert to manipulate 3D replicas of the target objects and to demonstrate exemplary behavior to the local worker on how to perform tasks. On the local worker side, the 3D replicas are overlaid on the video frames that capture the physical target objects, allowing the local worker to understand how to manipulate the physical objects. This method is particularly intuitive as it does not require the interpretation of the remote expert's actions, unlike the other interfaces.

3.3. Implementation

The WXR system is built upon immersive web technologies (W3C, 2022). The WXR Server consists of two main modules: A web server implemented in Node.js and a MariaDB database for storing and retrieving

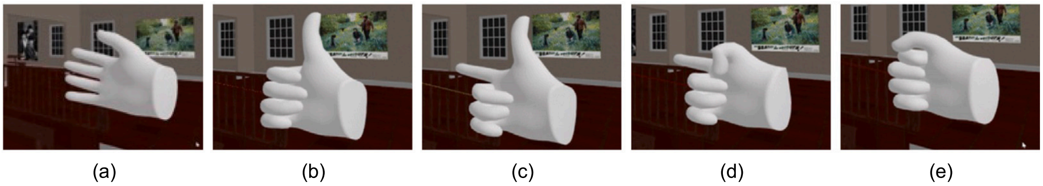


Fig. 4. Five hand gestures supported by the WXR system: (a) open, (b) thumb up, (c) pointing thumb up, (d) point, and (e) fist.

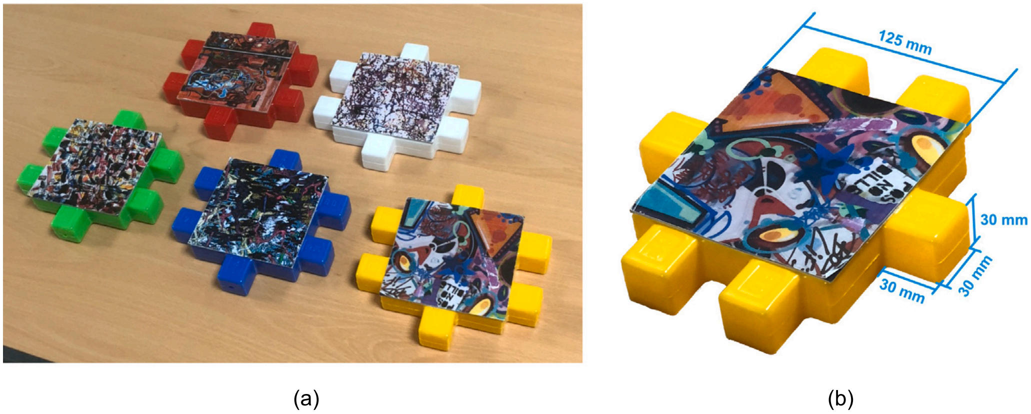


Fig. 5. Target objects used in our experiment. (a) Five distinctive image markers are attached to one flat side of each block. (b) Dimensions of a waffle block.

web resources as well as XR scene contents. The WXR Server runs on a desktop computer (Intel Xeon E5-2699 v4 2.2 GHz CPU, 64 GB DDR4 RAM) running Microsoft Windows 10.

For local worker use, we employ a tablet PC (iPad Pro 12.9-in 4th generation). The remote expert wears an HMD with VR controllers (Meta Quest 2). Our client software (the WXR Library) is perfectly operable on Meta Quest 2 alone. For the purpose of monitoring and recording the remote expert's view, we tether the expert's HMD over Wi-Fi to a desktop computer (AMD Ryzen 9 3900X 3.79 GHz CPU with 12 Cores, 64 GB DDR4 RAM, NVIDIA Quadro RTX 5000 GPU) running Microsoft Windows 10.

We use A-Frame (v1.2.0) (Marcos et al., 2023) to implement the WXR Library, which depends on the WebXR Device API (W3C, 2022) to retrieve the pose information of the HMD and controllers at 6 DOF. For the purpose of image tracking, we developed a custom web browser for the iPad based on the ARKit and Webkit frameworks, as this functionality is not natively supported by the WebXR Device API. Therefore, the custom browser provides pose information of the iPad and tracked images at 6 DOF to the WXR Library, instead of the WebXR Device API. Every data transmission between the WXR Server and the clients (the local worker's tablet PC and the remote expert's HMD) is done over Wi-Fi. Updates to the XR scene between the local worker and the remote expert are transmitted by the WebSocket API (WHATWG, 2023b), while voice communication is achieved by the WebRTC API (W3C, 2023).

To set up the VR space for the remote expert, we scan and reconstruct the 3D model of the local site (11.1 m × 8.4 m), using the 3D Scanner App™ from Laan Labs (Laan Labs, 2023). We craft the 3D models of waffle blocks, which serve as the target objects in our experiment, using Blender (Blender Foundation, 2023). A waffle block is a rectangular waffle-shaped object measuring 12.5 cm in width and length and 3 cm in height, with two, one, two, and one 27 cm^3 cubes attached to its four narrow sides, respectively (Fig. 5). Distinctive images are affixed to each waffle block to facilitate their identification and tracking. To visually represent the expert and the worker within the workspace, avatars generated through Ready Player Me (Ready Player Me Inc., 2023) are employed.

4. User study

4.1. Aim

Existing studies have shown that various types of interaction interfaces (e.g., Hand Gestures (HDG), 3D Drawing (3DD), and Virtual Replicas (VRP)) were applied during VR-AR communication and improved user experience in remote collaboration (Wang et al., 2019; Tian et al., 2023). Taking note of these findings, the research questions for this study are as follows:

- RQ1: Which interaction condition is the most effective communication method in terms of task completion time?
- RQ2: How does each condition affect the quality of remote collaboration, such as task load and system usability?
- RQ3: Is there a significant difference in user preferences between using the HDG, 3DD, and VRP interaction methods in an assembly task?

We hypothesize that the VRP interaction will outperform the other two methods in terms of task completion time, subjective quality of experience, and preference rank. We conducted a user study to test the impact of each interaction method on communication between experts and novices in a remote assembly task using multiple experimental measures (Table 2).

4.2. Participants

We recruited thirty participants (15 females) aged 23 to 37 years (M = 27.47, SD = 3.95). In the user study, we designed an actor-participant pair experiment in which each participant performed an object assembly task as a local worker while a trained actor served

Table 2
Summary of experimental measures.

Name	Details	Reference
Task completion time	The index measures the time duration from the moment the remote expert issues the initial instruction to the local worker until the completion of the final stage of block assembly. Both voice and video recordings were used for calculating this index.	Tian et al. (2023)
Numbers of instructions and Instruction time	Based on voice and video recording data, we quantified both the frequency and duration of each instruction issued by the remote expert to the local worker for describing the assembly procedure. We posited that a lower frequency of instructions coupled with a shorter duration would indicate a more efficient interaction interface for collaboration.	
General Collaborative Experience (GCE)	The GCE aims to measure the quality of collaborative experience. The questionnaire consists of 10 questions for either a remote expert or a local worker. In this experiment, we used a set of questions for a local worker to provide a subjective evaluation of the collaborative experience. Each question is answered on a 7-point Likert scale, and the score increases when participants have better user experiences.	Wang et al. (2021a)
NASA Task Load Index (NASA-TLX)	The NASA-TLX survey was used to evaluate the task load of the worker. The index considers six types of task load: Mental demand, Physical demand, Temporal demand, Performance, Effort, and Frustration. Each type is assessed on a 21-point Likert scale. While a lower score on the Performance index indicates better task results, higher scores on the remaining five indices suggest a more challenging task.	Hart and Staveland (1988)
System Usability Scale (SUS)	Using the SUS, we aimed to assess the usability of each interface. A total SUS score exceeding 80 is generally considered indicative of a system with excellent usability.	Brooke (1996)
Simulator Sickness Questionnaire (SSQ)	The SSQ can measure a subjective level of discomfort while using AR/VR devices. We used the SSQ to investigate whether participants felt simulator sickness during the collaboration. Participants filled out SSQ immediately after the collaboration ended. To quantify each symptom, the SSQ includes three subscales (i.e., nausea, oculomotor, and disorientation). A higher score indicates more severe discomfort.	Kennedy et al. (1993)
Preference rank	Participants ranked each interface in terms of overall preference as well as specific aspects of the collaboration, such as communication, focus, and presence.	Teo et al. (2019)
In-depth interview	After experiencing each interface, we conducted interviews asking about the participants' collaboration experiences, as well as the pros and cons of each interface.	Tian et al. (2023)

as the remote expert. Given that the study aimed to investigate the effectiveness of different interaction interfaces in understanding assembly procedures, we focused on task performance and the quality of collaboration from the local worker's perspective. To this end, we employed an actor who was well-trained in providing consistent and clear instructions across different interface types. This design choice was made to minimize the possibility that observed differences in the effectiveness of the interfaces could be attributed to variations in the experts' instructional capabilities.

During the experiments, participants experienced all three types of interaction interfaces (a within-subject design) in a counterbalanced manner to mitigate order effects. All experiments were conducted in accordance with the guidelines of the Institutional Review Board of KIST (IRB-2021-036).

4.3. Procedure

Fig. 6 outlines the entire sequence of our experimental procedure. Firstly, all participants completed an informed consent form and filled out questionnaires about demographic characteristics. Prior to each task, participants underwent a tutorial to familiarize themselves with each interface. Specifically, we had them experience the recognition of markers on the blocks using the AR device and briefly explained the graphical representations that would appear on the worker's display, such as the partner's avatar and interactions.

During the experiment, participants engaged in assembly tasks. Compared to previous studies, we considered the following points to make the assembly procedure complex (Table 1). First, we used 3D waffle blocks which can create various shapes with depth information. Second, all parts were identically shaped and could be continuously translated and rotated according to the assembly scenario. Lastly, the final assembled blocks were intentionally designed with an abstract structure, making it challenging for the local worker to deduce the end product's shape without step-by-step guidance from the expert.

Each assembly task required the participants to combine five blocks (i.e., 4 stages of block assembly) in a specific color sequence. The colors of the blocks were green, white, red, blue, and yellow respectively. Participants were positioned standing in front of a table, on which the five blocks were placed (Fig. 7(a)). The remote expert (i.e., actor) viewed the 3D scanned environment of the local area in VR. Upon connecting to the WXR workspace, the remote expert was shown images depicting the final assembly configuration of the waffle blocks. These images indicated four different viewpoints and the sequence of assembly by the initial letter of the color (Fig. 8).

Each assembly stage started with the expert guiding the participant about which block should be joined to which part of the other block. With the hand gestures (HDG) interface, the expert explained the assembly process by pointing directly at the sections that needed to be joined (Fig. 7(b)). When the 3D drawing interface (3DD) was used, the remote expert drew lines or arrows to indicate which parts should be connected or even drew a representation of the final assembled state roughly (Fig. 7(f)). Lastly, when the virtual replicas interface (VRP) was employed, the remote expert simulated the assembly process by manipulating 3D virtual models of the waffle blocks (Fig. 7(j)).

After understanding the instructions of the expert, which were augmented on the iPad screen, the participant initiated the physical assembly process (Fig. 7(c), (g), (k)). Upon completion of the assembly, the participant scanned the markers on the assembled blocks using the iPad, and the current placement state of the blocks was synchronized in the WXR workspace (Fig. 7(d), (h), (l)). Through this synchronization, the remote expert was able to check the progress of the physical assembly and correct errors (Fig. 7(e), (i), (m)). When the expert confirmed that all waffle blocks were assembled correctly, the task was finished. A video clip further illustrating the methods of this study can be found online in Appendix A at https://doi.org/10.1016/j.ijhcs.2024.103304.

4.4. Experiment measures

The study collected the following measures (Table 2) to quantify the user experience while using three interaction interfaces and answer the research questions.

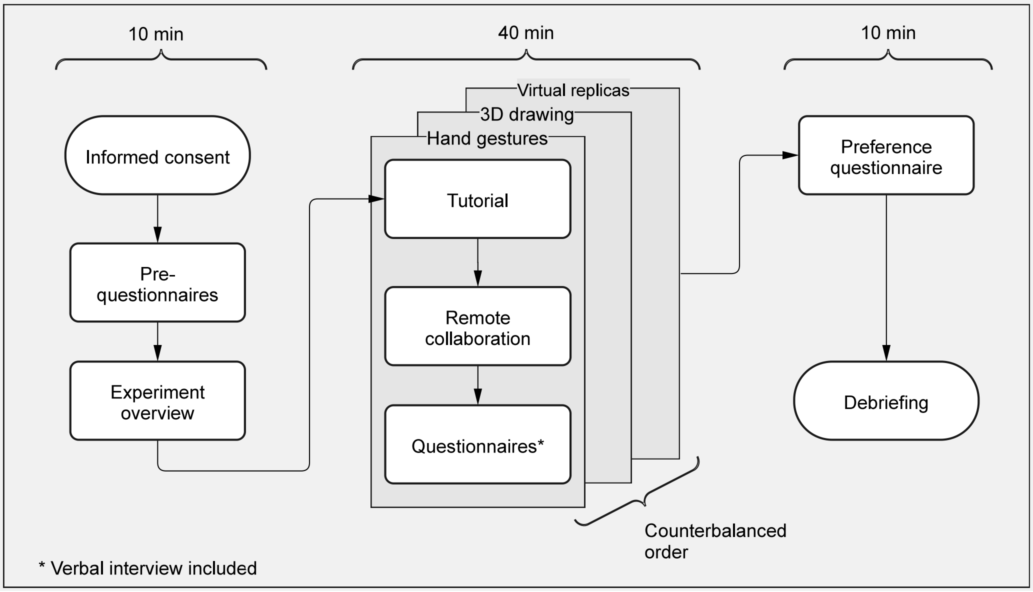


Fig. 6. A diagram of experimental procedures.

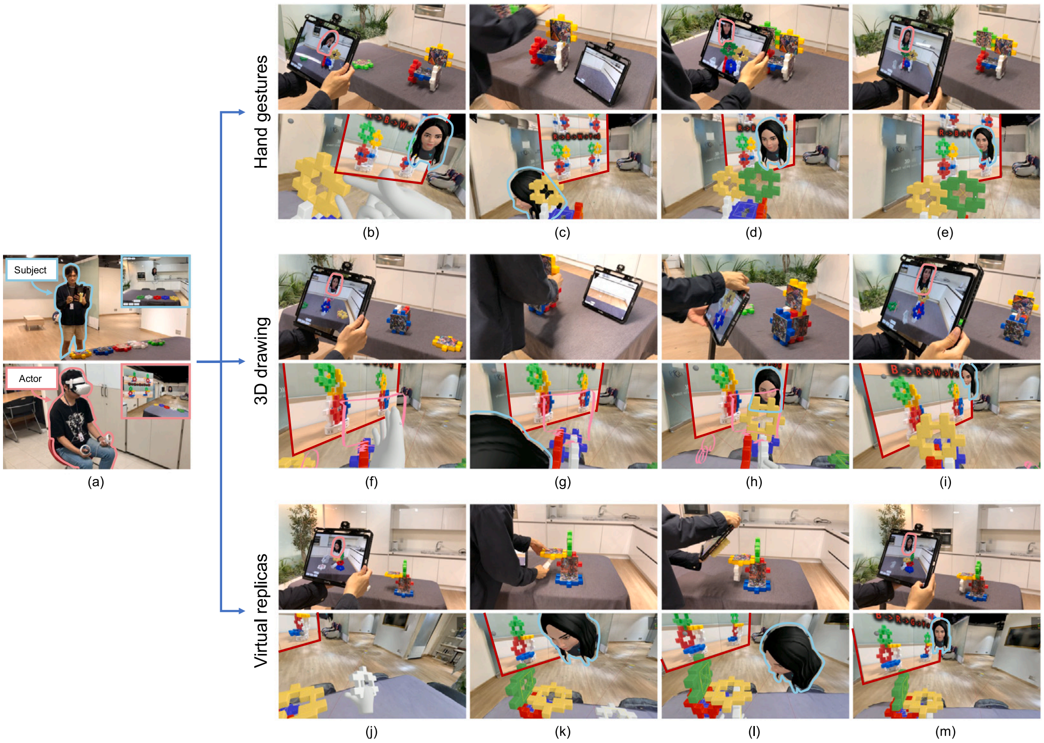


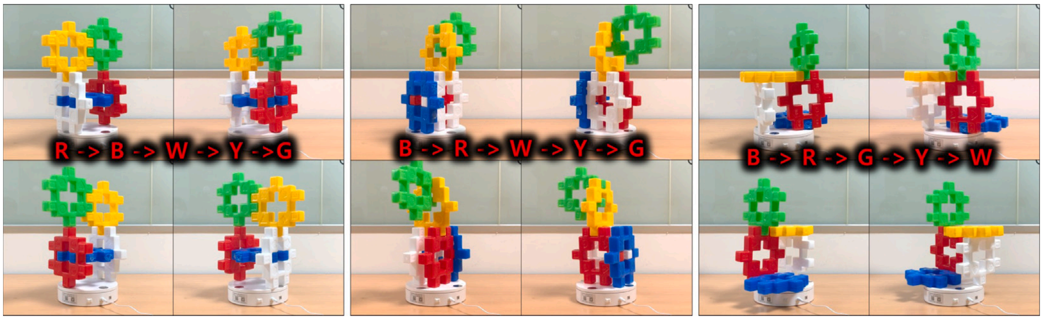
Fig. 7. Experiment setup (a) and example snapshots of conducting an assembly stage using hand gestures (b)–(e), 3D drawing (f)–(i), and virtual replicas (j)–(m). The upper row images of each interface show the local worker assembling the blocks following the expert's instructions, and the lower row images show the virtual workspace from the remote expert side.

4.5. Data analysis

Prior to data analysis, we employed the Interquartile Range (IQR) method to identify and remove outliers. This established method defines outliers as data points falling outside 1.5 times the IQR below the first quartile (Q1) or above the third quartile (Q3). Following outlier removal, we proceeded with the subsequent analyses.

Since we found violations of the normal distribution using the Shapiro–Wilk test, a non-parametric approach was used for the data analysis. We performed a Friedman test for the task completion time and questionnaires (GCE, NASA-TLX, SUS, and SSQ) with a significance level (α) of 0.05. We also calculated the 95% confidence interval (CI) for each variable to confirm the significant differences depending on the interface type.

If a significant interface effect was found in the Friedman test, post hoc analysis was conducted using Wilcoxon signed-rank tests with Bonferroni correction to avoid false positives due to the multiple comparisons. Since we compared three different pairs (HDG vs. VRP, 3DD vs. VRP, HDG vs. 3DD), adjusted p was .017 (=.05/3). Then, we calculated Cohen's d for the effect size.



(a) Hand gestures (b) 3D drawing (c) Virtual replicas

Fig. 8. Images of the correctly completed assembly in four different viewpoints according to each interface: (a) hand gestures, (b) 3D drawing, and (c) virtual replicas. Since the images were shown on the remote expert's view in VR, the expert referred to the pictures when instructing the assembly stage. The color initials (Red, Green, Blue, Yellow, and White) embedded in the center of the images denote the blocks involved at each assembly stage.

Lastly, for the preference rank, we conducted a Chi-squared test with a significance level (α) of 0.05. All statistical approaches were performed using an R (v4.0.3). The data supporting the findings of this study are openly available at https://doi.org/10.17632/h6y2md2fsk.1.

5. Results

For the completion time, a significant difference was found between the interface conditions ($\chi^2(2) = 9.41$, $p = .009$) (Fig. 9(a) left). The average task completion time of each interface was 162.8 s (hand gestures), 177.2 s (3D drawing), and 138.8 s (virtual replicas), respectively. According to the CI analysis (Fig. 9(a) right), 95% CIs between 3D drawing and virtual replicas conditions were not overlapped, suggesting that the difference between the two conditions is significant. When we performed a post hoc analysis with Wilcoxon signed-rank tests, the results indicated that the completion time of the virtual replicas condition was significantly faster than that of the hand gestures condition ($Z = 2.72$, $p = .007$, $d = 0.73$, 95% CI [0.23, 1.31]) and the 3D drawing condition ($Z = 3.09$, $p = .002$, $d = 0.93$, 95% CI [0.35, 1.61]).

We further divided the completion time into *instruction time* and *assembly time*. This separation aims to clarify how the interface impacted the specific workflow either instruction or assembly. Using the voice and video recording data, we measured the duration the remote expert needed to give instructions to the local worker (i.e., instruction time). A higher number of instructions indicates participants needed more guidance during the task. The rest (i.e., completion time subtracted by instruction time) was defined as the assembly time. Based on this, we analyzed whether the number of instructions, instruction time, and assembly time varied depending on interaction interfaces. We observed a significant effect of the interface on both the number of instructions ($\chi^2(2) = 11.37$, $p = .003$) and instruction time ($\chi^2(2) = 16.55$, $p = .0003$) (Fig. 9(b) left and (c) left). While the hand gestures and 3D drawing interfaces needed 6.6 and 6.8 times of instructions on average, the virtual replicas interface required an average of 5.0 instructions. The 95% CIs of the number of instructions and instruction time indicated that the virtual replicas interface showed fewer numbers and faster time in giving directions compared to the rest of the interfaces (Fig. 9(b) right and (c) right). Post hoc analysis also confirmed that significant differences between the virtual replicas interface and the other two interfaces (virtual replicas vs. hand gestures: $Z = 3.06$, $p = .002$, $d = 0.91$, 95% CI [0.37, 1.54]; virtual replicas vs. 3D drawing: $Z = 3.53$, $p = .0004$, $d = 1.29$, 95% CI [0.66, 2.05]) in terms of the number of instructions. When the virtual replicas interface was employed, the remote expert spent significantly less time providing assembly instructions compared to the hand gestures ($Z = 3.46$, $p = .0005$, $d = 1.15$, 95% CI [0.58, 1.86]) and 3D drawing interfaces ($Z = 3.23$, $p = .0012$, $d = 1.00$, 95% CI [0.40, 1.73]). However, no significant differences were observed in assembly time—the time taken to assemble the blocks—across the interfaces based on both the Friedman and CI tests (Fig. 9(d)).

For the GCE, we observed no significant differences across the interaction interfaces based on both the Friedman and CI tests. For the NASA-TLX, the Friedman test results showed a significant interface effect in *"Performance"* and *"Effort"* subscales (Fig. 10(c) left and (d) left). However, 95% CIs showed that all NASA-TLX subscales showed overlapped CIs between interfaces, indicating no significant differences were found due to the interface (Fig. 10(c) right and (d) right). According to the Friedman test, participants reported that they felt better performance when they were guided through the virtual replicas interface (Performance; $\chi^2(2) = 8.72$, $p = .013$). It should be noted that a lower NASA-TLX score represents a higher feeling of successful accomplishment during the collaboration (1: Perfect ~ 21: Failure). We performed a post hoc analysis and found a significant difference between the virtual replicas and hand gestures conditions ($Z = 2.45$, $p = .015$, $d = 0.61$, 95% CI [0.16, 1.14]). That is, participants reported better performance under the virtual replicas condition compared to the hand gestures condition. In terms of the effort subscale, the Friedman test result showed the lowest effort level in the virtual replicas interface (Effort; $\chi^2(2) = 11.57$, $p = .003$). Post hoc analysis revealed that being instructed through the virtual replicas required a significantly lower level of effort than the 3D drawing condition ($Z = 2.40$, $p = .016$, $d = 0.32$, 95% CI [0.06, 0.61]). Regarding the SSQ, participants did not indicate any statistically significant differences among the conditions based on both the Friedman and CI tests.

For the SUS, while we found a significant interface effect on system usability ($\chi^2(2) = 11.38$, $p = .003$) in the Friedman test (Fig. 11(a)), 95% CIs between interfaces were overlapped (Fig. 11(b)). The average SUS scores for each interface were 76.0 (hand gestures), 80.4 (3D drawing), and 82.8 (virtual replicas), respectively. Post hoc analysis revealed that the SUS score for the virtual replicas condition was significantly higher than that for the hand gestures condition ($Z = -2.49$, $p = .013$, $d = 0.45$, 95% CI [0.07, 0.88]), but no difference with the 3D drawing condition. Fig. 11(c) indicates the mean distributions of each SUS question according to interfaces.

A Chi-squared test was performed for the preference rank. We found a significant interface effect on every subscale of preference rank, that is, participants reported the first rank for the virtual replicas condition in all aspects (Fig. 12). To be specific, participants preferred the virtual replicas interface most in terms of communication (Q1; $\chi^2(4) = 43.60$, $p < .001$), focus (Q2; $\chi^2(4) = 32.20$, $p < .001$), being together (Q3; $\chi^2(4) = 25.60$, $p < .001$), guiding (Q4; $\chi^2(4) = 42.40$, $p < .001$), spatial awareness (Q5; $\chi^2(4) = 39.40$, $p < .001$), task completion (Q6; $\chi^2(4) = 42.40$, $p < .001$), and overall (Q7; $\chi^2(4) = 38.20$, $p < .001$).

6. Discussion

In this study, we investigated the effects of three interaction interfaces on XR remote collaboration. Our focus was on determining

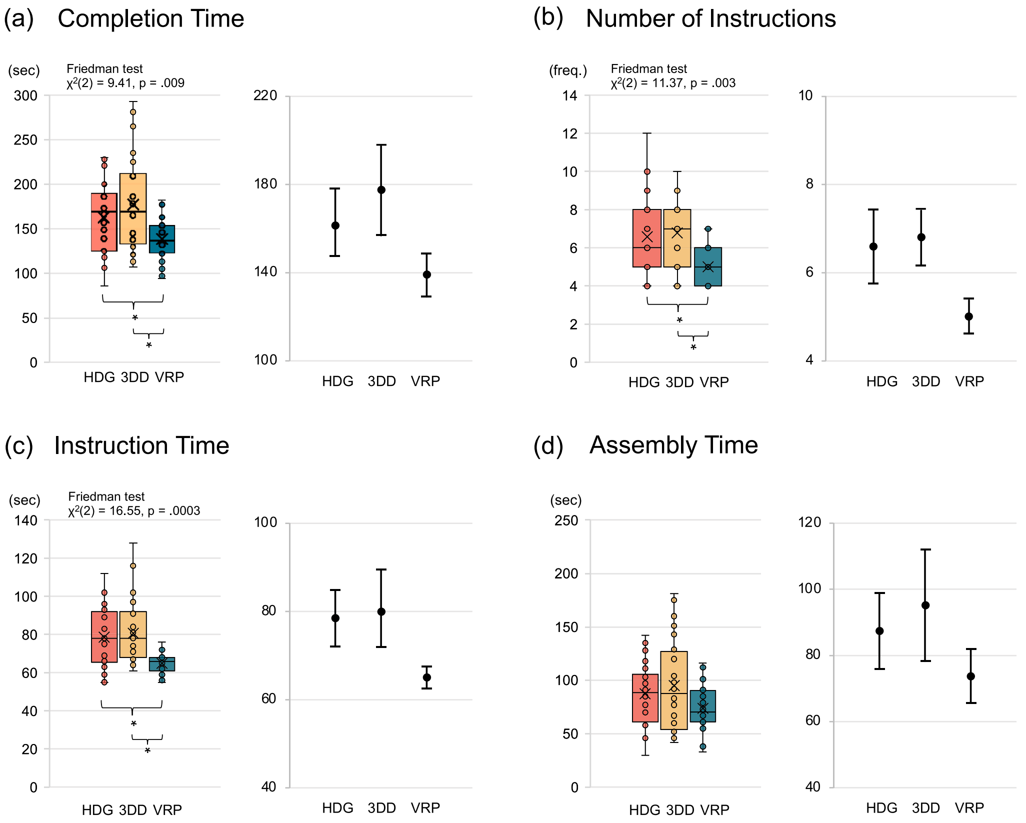


Fig. 9. Box graphs (left) and 95% CIs (right) of task performance (×: mean value of each interface, *: significant pairwise difference, $p_{adjustment}$ < .017). (a) Completion time, (b) Number of instructions, (c) Time for instruction, (d) Time for assembly.

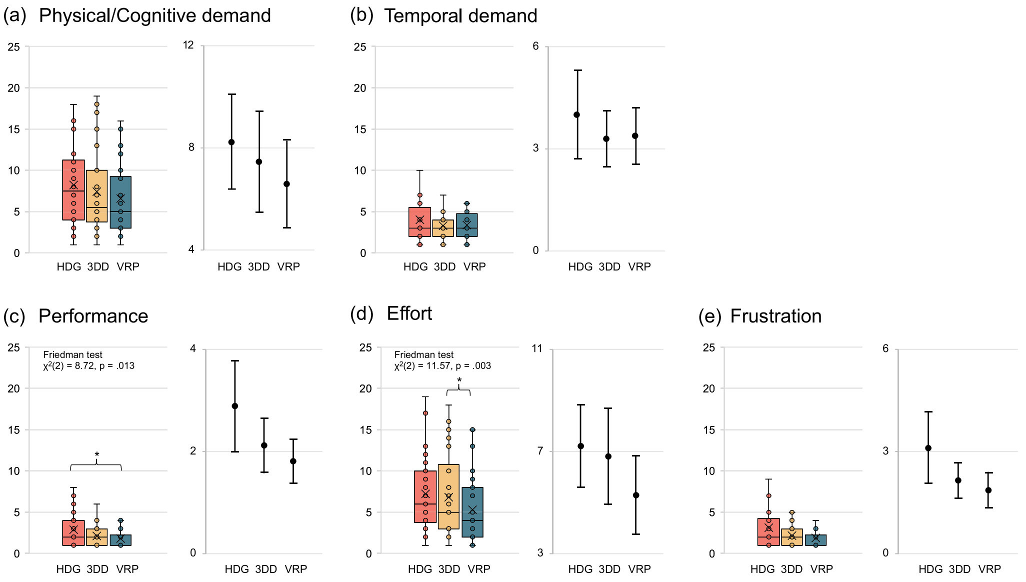


Fig. 10. Box graphs (left) and 95% CIs (right) of NASA-TLX scores (×: mean value of each interface, *: significant pairwise difference, $p_{adjustment}$ < .017). (a) Physical/Cognitive demand, (b) Temporal demand, (c) Performance, (d) Effort, (e) Frustration.

which interface is most effective for task completion during VR-AR communication (RQ1). Also, we collected several questionnaires (GCE, NASA-TLX, SUS, and SSQ) to assess the quality of collaboration depending on the interface type (RQ2). Finally, we acquired preference rankings to investigate whether there is a significant difference in user preferences between the interfaces (RQ3).

The result showed a significant difference in completion time based on the interface type, with the virtual replicas interface resulting in the fastest completion time compared to the hand gestures and 3D drawing conditions. Regarding the subjective measures, conflicting results were found depending on the analysis. While the Friedman test showed participants reported feeling better performance, lower effort,

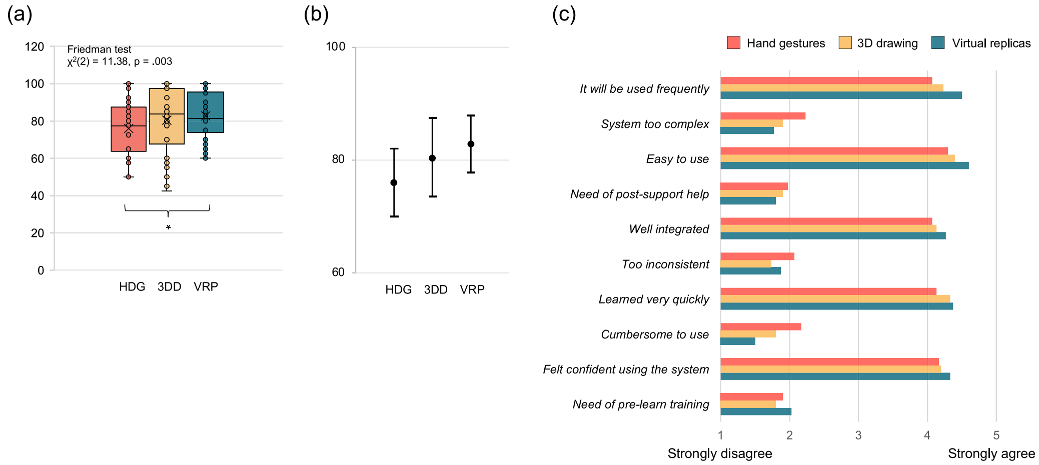


Fig. 11. (a) A box graph of SUS scores (×: mean value of each interface, *: significant pairwise difference, $p_{adjustment} < .017$), (b) CIs of SUS scores, (c) Mean score distributions of each SUS question according to the interaction interfaces.

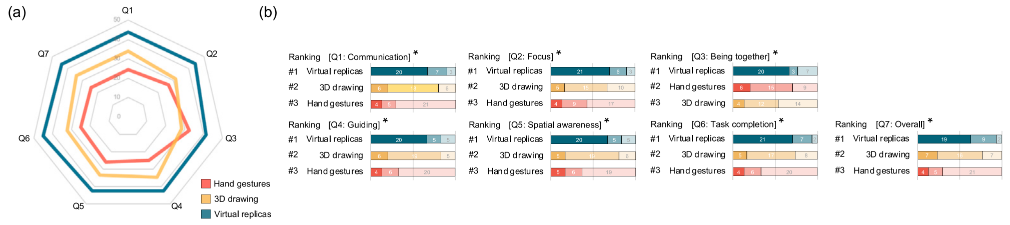


Fig. 12. (a) A radar graph of user preference. A higher number represents a higher rank in each preference question (b) Rank distributions of each preference question depending on the interfaces. The number inside each bar represents the sample size. (*: significant differences across interfaces).

and higher usability scores when experiencing the virtual replicas, these differences were not observed in the CI analysis. In terms of preference, the virtual replicas interface was ranked as the top choice across all aspects.

Based on these findings, the following sections discuss the implications of the results. We consider the interview reports and previous studies to interpret our results. After that, we discuss the limitations of this research and future directions for the next study.

6.1. Efficient VR-AR communication during the assembly task

Consistent with the findings by Zhang et al. (2022) and Tian et al. (2023), our study yielded similar results. Participants achieved the fastest completion times and exhibited a clear preference for the virtual replicas condition. The Friedman test indicated the shortest task completion time within the virtual replicas condition. Notably, confidence intervals for the virtual replicas and 3D drawing conditions did not overlap and indicated a large effect size ($d = 0.93$), whereas the CIs between virtual replicas and hand gestures slightly overlapped (Fig. 9 (a)). This result suggests that the difference between VRP and 3DD is more robust than the other pair, implying the advantage of using virtual replicas instead of drawing annotations in terms of time-saving.

This trend extended to the number and duration of instructions (Fig. 9(b) and (c)). By separating the collaborative scenario into the instruction and assembly phases, we aimed to identify the workflow most impacted by interface type. Since our assembly procedures consisted of four steps, at least four instructional sentences are required. The average number of instructions used was 5.0 in the virtual replicas interface, while the other two interfaces needed more than six instructional turns (6.6 and 6.8 times for hand gestures and 3D drawing, respectively) and the difference between the interfaces was significant. The results from CI analysis also confirmed that the three interfaces did not overlap, and CI of the virtual replicas condition indicated a narrow confidence interval, suggesting a precise estimate of the effect. Considering the large effect sizes between the pairs ($d > 0.8$), participants using virtual replicas requested fewer instructions, consequently reducing instruction time.

We posit that the instruction phase encompasses both the remote expert's communication and the local worker's comprehension. Therefore, successful grasping of instructions should translate to fewer instructions and shorter instruction durations. Aligned with this assumption, in-depth interviews revealed a preference for virtual replicas due to their intuitive presentation of task information, leading to a more comfortable and effortless assembly experience. Participants using virtual replicas simply followed the expert's virtual demonstration, minimizing the effort required to interpret instructions. This streamlined approach ultimately resulted in the shortest overall task completion time.

The 3D drawing interface ranked as the second most preferred option among participants, following the virtual replicas interface. According to participant feedback, the 3D drawing condition also aided them in comprehending instructions by providing visually rich information. For instance, one participant remarked that the drawings emphasized the task's focal points (P20: *"I personally preferred the drawing interface because it makes it easier to remember the remote expert's instructions."*). However, some reported that they had difficulty understanding the 3D spatial information of the drawings depending on the angle at which they viewed the drawings. As pointed out by Tian et al. (2023), it appears that the 3D drawing interface may produce mixed results according to the participants' drawing skills. The annotation drawing was just lines which made it difficult for the local user to identify its depth.

Among the three interfaces, hand gestures were rated as the least preferred. The results of completion time suggest that the hand gestures condition required additional interpretation during the interaction compared to the other two conditions. Participants spent more time precisely discerning the remote expert's pointing gestures and ensuring the proper orientation of the waffle blocks through the hand gestures interface. However, some participants appreciated the more detailed explanation of the assembly process, despite acknowledging that the hand gestures condition was less intuitive than the virtual replicas condition (P15: *"While the virtual replicas were helpful to grasp the final product of assembly, hand gestures were useful in understanding each assembly step in detail."*). For hand gestures to serve as a more effective interaction tool, there should be an adequate consensus between the collaborators regarding the meaning attributed to each gesture.

Despite initial expectations for distinct differences, the data showed mixed results in subjective user experience depending on the interface type (Figs. 10 and 11). Though a few measures (NASA-TLX subscales and SUS) indicated a significant difference under the Friedman test, the confidence intervals for the three interfaces overlapped substantially, suggesting the observed difference could be due to chance. One possible explanation can be found in the SUS score of each interface. The average SUS scores for hand gestures, 3D drawing, and virtual replicas interface were 76.0, 80.4, and 82.8, respectively. According to Bangor et al. (2008), all interfaces fall within the range of acceptable usability, with scores above 70 indicating good products and those from the high 70s to the upper 80s denoting better quality. The closeness in usability scores among the interfaces suggests that each has its strengths and contributes positively to user collaboration and experience. Given the overall similarity in scores, it might have been demanding for users to tell the subtle difference in the subjective experience during the collaboration.

Overall, participants performed better when viewing virtual replicas among the different types of interaction interfaces. Under assembly scenarios, it is crucial for collaborative systems to provide precise spatial references for the correct assembly. The virtual replicas interface was most preferred since it visualizes the remote expert's simulation as it is, allowing the local worker to intuitively understand the instructions without further interpretation. However, some participants reported that the virtual replicas condition limited their autonomy because they felt they were simply copying the expert's demonstration (P13: *"I felt like I had to do what was demonstrated, which limited my autonomy".*). Therefore, providing other visual (or multi-modal) annotations alongside the virtual replicas would enrich communication and improve collaborative performance.

Compared to prior studies on remote collaboration, our research offers the following contributions. By improving the WXR system, we have effectively minimized the hardware complexity often associated with multiple sensors and cameras. This streamlined setup increases the feasibility of deploying the system in field-based collaborations. Moreover, our system provides three distinct interaction interfaces simultaneously, allowing collaborators the flexibility to select and switch between interfaces based on their specific working contexts. Although our results highlighted superior task performance and a marked preference for the virtual replicas interface during intricate assembly tasks, this system's adaptability extends to various collaborative scenarios. Such versatility further deepens our understanding of refined communication methods during remote collaboration.

6.2. Limitations and future work

While we demonstrated the effectiveness of virtual replicas in remote collaboration, it has several limitations. Though we adopted a simpler hardware configuration similar to that of previous studies, several participants reported inconvenience while using the hand-held device (P14, P19, and P22: *"I felt uncomfortable with having to hold the iPad in one hand and follow the instructions with the other".*). Our deliberate selection of the hand-held device was designed to minimize fatigue during prolonged usage (Hughes et al., 2020), while simultaneously facilitating accurate tracking of target objects (e.g., waffle blocks) and rendering augmented graphics in appropriate spatial contexts. Nonetheless, this approach inherently restricts the concurrent use of both hands. If advances in AR glasses improve user eye fatigue or if high-performance video see-through headsets become more widely available, we will consider wearable AR devices.

We intentionally implemented the actor-participant pair design like in the study of Teo et al. (2019) since the present research primarily focuses on the local worker's performance and experience when interacting with different types of communication interfaces. Providing well-trained and consistent instructions run by a remote expert (i.e. an actor), we tried to elucidate the effect of the interaction interface itself rather than a person's capability of explaining the task procedures. However, this approach may have limited the extent of dynamic communication between collaborators. In particular, the lack of using visual communication cues on the remote expert end should be improved for future studies.

Our system requires the creation of 3D models of local sites and the AR tracking marker attachment onto target objects as a prerequisite for collaboration, irrespective of the interface type. This preparatory step is intended to facilitate the real-time sharing of local contextual information with remote experts. However, this process presents challenges, as it can be intricate and hassle for local workers. For example, the local worker might require additional knowledge to generate high-quality 3D models and ensure accurate marker placement for optimal tracking quality. To address these challenges and enhance the usability of the system for widespread adoption, future works could explore the integration of advanced 3D reconstruction methodologies (Gu et al., 2023) and tracking techniques aimed at streamlining the preparatory process.

A limited number of hand gestures can restrict the extent of expression. The set of predefined hand gestures was derived from the backbone framework, AFrame (Marcos et al., 2023), and five gestures were determined through a pilot study. While our findings demonstrated the adequacy of these gestures for VR-AR communication, the predefined set may still constrain naturalistic gestural exchanges in real-world contexts. As our experiment was confined to the assembly scenario within a specific collaborative context, it is required to apply diverse use case scenarios with a wider range of participants to clarify the hand gesture effect on the XR experience.

For future studies, we will extend the application of the present system to VR-VR and AR-AR remote collaboration as well. In particular, we will adopt AR glasses to provide better usability for the local worker. Since this device allows a person to move hands freely, the local worker can easily manipulate the assembly parts. Also, we are planning to improve the current hand-tracking function to simulate natural hand gestures using the hand-tracking API of WebXR specification (W3C, 2024). The follow-up study will be a participant-participant pair scenario focusing on more real-world-based collaboration with richer communication, and the tasks will include more complicated assignments such as finding mechanical parts and assembling or disassembling them. During those tasks, we plan to record the participant's physiological signals like heartbeats, eye gaze, and brain waves, which are known to improve co-presence (Jing et al., 2022) and to be related to monitoring any bodily discomfort during collaboration (Chang et al., 2022, 2023a).

7. Conclusion

This research aimed to demonstrate the most effective communication method for remote collaboration in assembly tasks. We compared three interaction interfaces (hand gestures, 3D drawing, and virtual replicas) and collected data on completion time, questionnaires, and preference rankings to assess the quality of VR-AR communication. The results indicated that using virtual replicas as an interaction interface can lead to improved efficiency and a strong preference for assembly collaboration. These findings suggest that virtual replicas can provide intuitive instructions to the local worker, which results in a clearer and faster understanding of the expert's guidelines. In the future, we plan to adopt AR glasses to expand our hardware setup for XR collaborations. We will also conduct user studies under various contexts of use cases to confirm the effect of virtual replicas on VR-AR communication.

CRediT authorship contribution statement

Eunhee Chang: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis. **Yongjae Lee:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Conceptualization. **Mark Billinghurst:** Writing – review & editing. **Byounghyun Yoo:** Writing – review & editing, Supervision, Resources, Funding acquisition, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Industrial Technology Innovation Program (20012462) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea), the National Research Foundation of Korea (NRF) grant (NRF-2021R1A2C2093065 and NRF-2021R1A6A3A14039652) funded by the Korea government (MSIT) and the KIST under the Institutional Program (Grant No. 2E32984).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijhcs.2024.103304.

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. VRCAI '13, Association for Computing Machinery, New York, NY, USA, pp. 235–242. http://dx.doi.org/10.1145/2534329.2534331.

Alem, L., Li, J., 2011. A study of gestures in a video-mediated collaborative assembly task. Adv. Hum.-Comput. Interact. 2011, 987830. http://dx.doi.org/10.1155/2011/987830, 7 pages.

Anton, D., Kurillo, G., Bajcsy, R., 2018. User experience and interaction performance in 2D/3D telecollaboration. Future Gener. Comput. Syst. 82, 77–88. http://dx.doi.org/10.1016/j.future.2017.12.055.

Aschenbrenner, D., Li, M., Dukalski, R., Casper Verlinden, J., Dukalski, R., Verlinden, J., Lukosch, S., 2018. Exploration of different augmented reality visualizations for enhancing situation awareness for remote factory planning assistance. In: Fourth IEEE VR International Workshop on 3D Collaborative Virtual Environments. 3DCVE 2018, pp. 3–7. http://dx.doi.org/10.13140/RG.2.2.14819.66083.

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. CHI '20, Association for Computing Machinery, New York, NY, USA, pp. 1–13. http://dx.doi.org/10.1145/3313831.3376550.

Bangor, A., Kortum, P.T., Miller, J.T., 2008. An empirical evaluation of the system usability scale. Int. J. Hum.-Comput. Interact. 24 (6), 574–594. http://dx.doi.org/10.1080/10447310802205776.

Blender Foundation, 2023. Blender. URL: https://www.blender.org/.

Bottecchia, S., Cieutat, J.-M., Jessel, J.-P., 2010. T.A.C: Augmented reality system for collaborative tele-assistance in the field of maintenance through internet. In: Proceedings of the 1st Augmented Human International Conference. AH '10, Association for Computing Machinery, New York, NY, USA, pp. 1–7. http://dx.doi.org/10.1145/1785455.1785469.

Brooke, J., 1996. SUS: A 'quick and dirty' usability scale. In: Usability Evaluation in Industry, first ed. CRC Press, London, pp. 189–194. http://dx.doi.org/10.1201/9781498710411, chapter 21.

Chang, E., Billinghurst, M., Yoo, B., 2023a. Brain activity during cybersickness: A scoping review. Virtual Real. 27 (3), 2073–2097. http://dx.doi.org/10.1007/s10055-023-00795-y.

Chang, E., Kim, H.T., Yoo, B., 2022. Identifying physiological correlates of cybersickness using heartbeat-evoked potential analysis. Virtual Real. 26 (3), 1193–1205. http://dx.doi.org/10.1007/s10055-021-00622-2.

Chang, E., Lee, Y., Yoo, B., 2023b. A user study on the comparison of view interfaces for VR-ar communication in XR remote collaboration. Int. J. Hum.-Comput. Interact. http://dx.doi.org/10.1080/10447318.2023.2241294.

Cho, H., Park, S., Park, C., Jung, S.-U., 2021. Efficient mapping technique under various spatial changes for SLAM-based AR services. In: Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology. VRST '21, Association for Computing Machinery, New York, NY, USA, pp. 1–3. http://dx.doi.org/10.1145/3489849.3489916.

Choi, J., Yoon, B., Jung, C., Woo, W., 2017. ArClassNote: Augmented reality based remote education solution with tag recognition and shared hand-written note. In: 2017 IEEE International Symposium on Mixed and Augmented Reality. ISMAR-Adjunct, IEEE, Nantes, France, pp. 303–309. http://dx.doi.org/10.1109/ISMAR-Adjunct.2017.94.

D'Angelo, S., Begel, A., 2017. Improving communication between pair programmers using shared gaze awareness. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, Association for Computing Machinery, New York, NY, USA, pp. 6245–6290. http://dx.doi.org/10.1145/3025453.3025573.

Dey, A., Piumsomboon, T., Lee, Y., Billinghurst, M., 2017. Effects of sharing physiological states of players in a collaborative virtual reality gameplay. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, Association for Computing Machinery, New York, NY, USA, pp. 4045–4056. http://dx.doi.org/10.1145/3025453.3026028.

El Ammari, K., Hammad, A., 2019. Remote interactive collaboration in facilities management using BIM-based mixed reality. Autom. Constr. 107, 102940. http://dx.doi.org/10.1016/j.autcon.2019.102940, 19 pages.

Elvezio, C., Sukan, M., Oda, O., Feiner, S., Tversky, B., 2017. Remote collaboration in AR and VR using virtual replicas. In: ACM SIGGRAPH 2017 VR Village. SIGGRAPH '17, Association for Computing Machinery, New York, NY, USA, pp. 1–2. http://dx.doi.org/10.1145/3089269.3089281.

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., Billinghurst, M., 2019. Revisiting collaboration through mixed reality: The evolution of groupware. Int. J. Hum. Comput. Stud. 131, 81–98. http://dx.doi.org/10.1016/j.ijhcs.2019.05.011.

Fakourfar, O., Ta, K., Tang, R., Bateman, S., Tang, A., 2016. Stabilized annotations for mobile remote assistance. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, New York, NY, USA, pp. 1548–1560. http://dx.doi.org/10.1145/2858036.2858171.

Fang, W., Chen, L., Zhang, T., Chen, C., Teng, Z., Wang, L., 2023. Head-mounted display augmented reality in manufacturing: A systematic review. Robot. Comput.-Integr. Manuf. 83, 102567. http://dx.doi.org/10.1016/j.rcim.2023.102567, 27 pages.

Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E., Kramer, A.D.I., 2004. Gestures over video streams to support remote collaboration on physical tasks. Hum. Comput. Interact. 19 (3), 273–309. http://dx.doi.org/10.1207/s15327051hci1903_3.

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. OzCHI '20, Association for Computing Machinery, New York, NY, USA, pp. 629–638. http://dx.doi.org/10.1145/3441000.3441038.

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. SA '16, Association for Computing Machinery, New York, NY, USA, pp. 1–4. http://dx.doi.org/10.1145/2999508.2999531.

Gao, L., Bai, H., Lindeman, R., Billinghurst, M., 2017. Static local environment capturing and sharing for MR remote collaboration. In: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications. SA '17, Association for Computing Machinery, New York, NY, USA, pp. 1–6. http://dx.doi.org/10.1145/3132787.3139204.

Gasques, D., Johnson, J.G., Sharkey, T., Feng, Y., Wang, R., Xu, Z.R., Zavala, E., Zhang, Y., Xie, W., Zhang, X., Davis, K., Yip, M., Weibel, N., 2021. ARTEMIS: A collaborative mixed-reality system for immersive surgical telementoring. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21, Association for Computing Machinery, New York, NY, USA, pp. 1–14. http://dx.doi.org/10.1145/3411764.3445576.

Gauglitz, S., Lee, C., Turk, M., Höllerer, T., 2012. Integrating the physical environment into mobile remote collaboration. In: Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services. MobileHCI '12, Association for Computing Machinery, New York, NY, USA, pp. 241–250. http://dx.doi.org/10.1145/2371574.2371610.

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. UIST '14, Association for Computing Machinery, New York, NY, USA, pp. 449–459. http://dx.doi.org/10.1145/2642918.2647372.

Genest, A.M., Gutwin, C., Tang, A., Kalyn, M., Ivkovic, Z., 2013. KinectArms: A toolkit for capturing and displaying arm embodiments in distributed tabletop groupware. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work. CSCW '13, Association for Computing Machinery, New York, NY, USA, pp. 157–166. http://dx.doi.org/10.1145/2441776.2441796.

Gu, J., Trevithick, A., Lin, K.-E., Susskind, J.M., Theobalt, C., Liu, L., Ramamoorthi, R., 2023. NerfDiff: Single-image view synthesis with nerf-guided distillation from 3D-aware diffusion. In: Proc. 40th Int. Conf. Mach. Learn.. Vol. 202, PMLR, pp. 11808–11826.

Gunn, C., Adcock, M., 2011. Using sticky light technology for projected guidance. In: Proceedings of the 23rd Australian Computer-Human Interaction Conference. OzCHI '11, Association for Computing Machinery, New York, NY, USA, pp. 140–143. http://dx.doi.org/10.1145/2071536.2071557.

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. CHI '12, Association for Computing Machinery, New York, NY, USA, pp. 619–622. http://dx.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: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Vol. 52, North-Holland, pp. 139–183. http://dx.doi.org/10.1016/S0166-4115(08)62386-9.

He, Z., Du, R., Perlin, K., 2020. CollaboVR: A reconfigurable framework for creative collaboration in virtual reality. In: 2020 IEEE International Symposium on Mixed and Augmented Reality. ISMAR, IEEE, Porto de Galinhas, Brazil, pp. 542–554. http://dx.doi.org/10.1109/ISMAR50242.2020.00082.

Higuch, K., Yonetani, R., Sato, Y., 2016. Can eye help you? Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, Association for Computing Machinery, New York, NY, USA, pp. 5180–5190. http://dx.doi.org/10.1145/2858036.2858438.

Higuchi, K., Chen, Y., Chou, P.A., Zhang, Z., Liu, Z., 2015. ImmerseBoard: Immersive telepresence experience using a digital whiteboard. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, Association for Computing Machinery, New York, NY, USA, pp. 2383–2392. http://dx.doi.org/10.1145/2702123.2702160.

Huang, W., Alem, L., Tecchia, F., 2013. HandsIn3D: Supporting remote guidance with immersive virtual environments. In: Human-Computer Interaction – INTERACT 2013. Vol. 8117, Springer, pp. 70–77. http://dx.doi.org/10.1007/978-3-642-40483-2_5.

Huang, W., Kim, S., Billinghurst, M., Alem, L., 2019. Sharing hand gesture and sketch cues in remote collaboration. J. Vis. Commun. Image Represent. 58, 428–438. http://dx.doi.org/10.1016/j.jvcir.2018.12.010.

Hughes, C.L., Fidopiastis, C., Stanney, K.M., Bailey, P.S., Ruiz, E., 2020. The psychometrics of cybersickness in augmented reality. Front. Virtual Real. 1, 602954. http://dx.doi.org/10.3389/frvir.2020.602954.

Izadi, S., Agarwal, A., Criminisi, A., Winn, J., Blake, A., Fitzgibbon, A., 2007. C-slate: A multi-touch and object recognition system for remote collaboration using horizontal surfaces. In: Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems. TABLETOP'07, IEEE, Newport, RI, USA, pp. 3–10. http://dx.doi.org/10.1109/TABLETOP.2007.34.

Jing, A., May, K., Matthews, B., Lee, G., Billinghurst, M., 2022. The impact of sharing gaze behaviours in collaborative mixed reality. Proc. ACM Hum.-Comput. Interact. 6 (CSCW2), 463. http://dx.doi.org/10.1145/3555564, 27 pages.

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. In: CHI EA '21, Association for Computing Machinery, New York, NY, USA, pp. 1–7. http://dx.doi.org/10.1145/3411763.3451844.

Jo, H., Hwang, S., 2013. Chili: Viewpoint control and on-video drawing for mobile video calls. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems. In: CHI EA '13, Association for Computing Machinery, New York, NY, USA, pp. 1425–1430. http://dx.doi.org/10.1145/2468356.2468610.

Kasahara, S., Heun, V., Lee, A.S., Ishii, H., 2012. Second surface: Multi-user spatial collaboration system based on augmented reality. In: SIGGRAPH Asia 2012 Emerging Technologies. SA '12, Association for Computing Machinery, New York, NY, USA, pp. 1–4. http://dx.doi.org/10.1145/2407707.2407727.

Katti, J., Bhavsar, P., Gidh, M., Desale, R., Chaudhari, P., Patil, S., 2021. Knowledge transfer to pre-schoolers based on augmented reality. In: 2021 5th International Conference on Intelligent Computing and Control Systems. ICICCS, IEEE, Madurai, India, pp. 1334–1338. http://dx.doi.org/10.1109/ICICCS51141.2021.9432313.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. 3 (3), 203–220. http://dx.doi.org/10.1207/s15327108ijap0303_3.

Kim, S., Lee, G., Billinghurst, M., Huang, W., 2020. The combination of visual communication cues in mixed reality remote collaboration. J. Multimodal User Interfaces 14 (4), 321–335. http://dx.doi.org/10.1007/s12193-020-00335-x.

Kim, S., Lee, G.A., Ha, S., Sakata, N., Billinghurst, M., 2015. Automatically freezing live video for annotation during remote collaboration. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. In: CHI EA '15, Association for Computing Machinery, New York, NY, USA, pp. 1669–1674. http://dx.doi.org/10.1145/2702613.2732838.

Kim, S., Lee, G.A., Sakata, N., 2013. Comparing pointing and drawing for remote collaboration. In: 2013 IEEE International Symposium on Mixed and Augmented Reality. ISMAR, IEEE, Adelaide, SA, Australia, pp. 1–6. http://dx.doi.org/10.1109/ISMAR.2013.6671833.

Kim, S., Lee, G., Sakata, N., Billinghurst, M., 2014. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In: 2014 IEEE International Symposium on Mixed and Augmented Reality. ISMAR, IEEE, Munich, Germany, pp. 83–92. http://dx.doi.org/10.1109/ISMAR.2014.6948412.

Kim, B., Seo, S., 2023. EfficientNetV2-based dynamic gesture recognition using transformed scalogram from triaxial acceleration signal. J. Comput. Des. Eng. 10 (4), 1694–1706. http://dx.doi.org/10.1093/jcde/qwad068.

Laan Labs, 2023. 3D scanner app. URL: https://3dscannerapp.com/.

Le Chénéchal, M., Duval, T., Gouranton, V., Royan, J., Arnaldi, B., 2016. Vishnu: Virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In: 2016 IEEE Third VR International Workshop on Collaborative Virtual Environments. 3DCVE, IEEE, Greenville, SC, USA, pp. 9–12. http://dx.doi.org/10.1109/3DCVE.2016.7563559.

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. J. Comput. Des. Eng. 10 (3), 1047–1059. http://dx.doi.org/10.1093/jcde/qwad035.

Lee, Y., Moon, C., Ko, H., Lee, S.-H., Yoo, B., 2020. Unified representation for XR content and its rendering method. In: The 25th International Conference on 3D Web Technology. In: Web3D '20, Association for Computing Machinery, New York, NY, USA, pp. 1–10. http://dx.doi.org/10.1145/3424616.3424695.

Lee, G.A., Teo, T., Kim, S., Billinghurst, M., 2017. Mixed reality collaboration through sharing a live panorama. In: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications. SA '17, Association for Computing Machinery, New York, NY, USA, pp. 1–4. http://dx.doi.org/10.1145/3132787.3139203.

Lee, G.A., Teo, T., Kim, S., Billinghurst, M., 2018. A user study on MR remote collaboration using live 360 video. In: 2018 IEEE International Symposium on Mixed and Augmented Reality. ISMAR, IEEE, Munich, Germany, pp. 153–164. http://dx.doi.org/10.1109/ISMAR.2018.00051.

Lee, Y., Yoo, B., 2021. XR collaboration beyond virtual reality: Work in the real world. J. Comput. Des. Eng. 8 (2), 756–772. http://dx.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. In: Web3D '21, Association for Computing Machinery, New York, NY, USA, pp. 1–9. http://dx.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. DIS '20, Association for Computing Machinery, New York, NY, USA, pp. 615–625. http://dx.doi.org/10.1145/3357236.3395429.

Lindlbauer, D., Wilson, A.D., 2018. Remixed reality: Manipulating space and time in augmented reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, New York, NY, USA, pp. 1–13. http://dx.doi.org/10.1145/3173574.3173703.

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

Nuernberger, B., Lien, K.-C., Höllerer, T., Turk, M., 2016. Anchoring 2D gesture annotations in augmented reality. In: 2016 IEEE Virtual Reality. VR, IEEE, Greenville, SC, USA, pp. 247–248. http://dx.doi.org/10.1109/VR.2016.7504746.

Oda, O., Elvezio, C., Sukan, M., Feiner, S., Tversky, B., 2015. Virtual replicas for remote assistance in virtual and augmented reality. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology. UIST '15, Association for Computing Machinery, New York, NY, USA, pp. 405–415. http://dx.doi.org/10.1145/2807442.2807497.

Ogawa, T., Kiyokawa, K., Takemura, H., 2005. A hybrid image-based and model-based telepresence system using two-pass video projection onto a 3D scene model. In: Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality. ISMAR'05, IEEE, Vienna, Austria, pp. 202–203. http://dx.doi.org/10.1109/ISMAR.2005.3.

Okajima, Y., Yamamoto, S., Bannai, Y., Okada, K., 2009. An instruction method for displaying trajectory of an object in remote collaborative MR on the basis of changes in relative coordinates. In: 2009 Ninth Annual International Symposium on Applications and the Internet. IEEE, Bellevue, WA, USA, pp. 43–49. http://dx.doi.org/10.1109/SAINT.2009.16.

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P.L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P.A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S.B., Kohli, P., Lutchyn, Y., Keskin, C., Izadi, S., 2016. Holoportation: Virtual 3D teleportation in real-time. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology. UIST '16, Association for Computing Machinery, New York, NY, USA, pp. 741–754. http://dx.doi.org/10.1145/2984511.2984517.

Palmer, D., Adcock, M., Smith, J., Hutchins, M., Gunn, C., Stevenson, D., Taylor, K., 2007. Annotating with light for remote guidance. In: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces. OZCHI '07, Association for Computing Machinery, New York, NY, USA, pp. 103–110. http://dx.doi.org/10.1145/1324892.1324911.

Park, S., Cho, H., Park, C., Yoon, Y.-S., Jung, S.-U., 2020. AR room: Real-time framework of camera location and interaction for augmented reality services. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. VRW, IEEE, Atlanta, GA, USA, pp. 736–737. http://dx.doi.org/10.1109/VRW50115.2020.00219.

Pauchet, A., Coldefy, F., Lefebvre, L., Picard, S.L.D., Bouguet, A., Perron, L., Guerin, J., Corvaisier, D., Collobert, M., 2007. Mutual awareness in collocated and distant collaborative tasks using shared interfaces. In: Baranauskas, C., Palanque, P., Abascal, J., Barbosa, S.D.J. (Eds.), Human-Computer Interaction – INTERACT 2007. Springer, Berlin, Heidelberg, pp. 59–73. http://dx.doi.org/10.1007/978-3-540-74796-3_8.

Piumsomboon, T., Dey, A., Ens, B., Lee, G., Billinghurst, M., 2019. The effects of sharing awareness cues in collaborative mixed reality. Front. Robot. AI 6, 5. http://dx.doi.org/10.3389/frobt.2019.00005, 18 pages.

Piumsomboon, T., Lee, G.A., Billinghurst, M., 2018a. Snow dome: A multi-scale interaction in mixed reality remote collaboration. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. In: CHI EA '18, Association for Computing Machinery, New York, NY, USA, pp. 1–4. http://dx.doi.org/10.1145/3170427.3186495.

Piumsomboon, T., Lee, G.A., Hart, J.D., Ens, B., Lindeman, R.W., Thomas, B.H., Billinghurst, M., 2018b. Mini-me: An adaptive avatar for mixed reality remote collaboration. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, New York, NY, USA, pp. 1–13. http://dx.doi.org/10.1145/3173574.3173620.

Radu, I., Joy, T., Schneider, B., 2021. Virtual makerspaces: Merging AR/VR/MR to enable remote collaborations in physical maker activities. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1–5. http://dx.doi.org/10.1145/3411763.3451561.

Ready Player Me Inc., 2023. Ready player me. URL: https://readyplayer.me/.

Roo, J.S., Hachet, M., 2017. Towards a hybrid space combining spatial augmented reality and virtual reality. In: 2017 IEEE Symposium on 3D User Interfaces. 3DUI, IEEE, Los Angeles, CA, USA, pp. 195–198. http://dx.doi.org/10.1109/3DUI.2017.7893339.

Sakong, K., Nam, T.-j., 2006. Supporting telepresence by visual and physical cues in distributed 3D collaborative design environments. In: CHI '06 Extended Abstracts on Human Factors in Computing Systems. In: CHI EA '06, Association for Computing Machinery, New York, NY, USA, pp. 1283–1288. http://dx.doi.org/10.1145/1125451.1125690.

Sasikumar, P., Gao, L., Bai, H., Billinghurst, M., 2019. Wearable RemoteFusion: A mixed reality remote collaboration system with local eye gaze and remote hand gesture sharing. In: 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct. ISMAR-Adjunct, IEEE, Beijing, China, pp. 393–394. http://dx.doi.org/10.1109/ISMAR-Adjunct.2019.000-3.

Schäfer, A., Reis, G., Stricker, D., 2022. A survey on synchronous augmented, virtual, and mixed reality remote collaboration systems. ACM Comput. Surv. 55 (6), 116. http://dx.doi.org/10.1145/3533376, 27 pages.

Seo, J.H., Lee, I.D., Yoo, B., 2021. Effectiveness of rough initial scan for high-precision automatic 3D scanning. J. Comput. Des. Eng. 8 (5), 1332–1354. http://dx.doi.org/10.1093/jcde/qwab049.

Sodhi, R.S., Jones, B.R., Forsyth, D., Bailey, B.P., Maciocci, G., 2013. Bethere: 3D mobile collaboration with spatial input. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '13, Association for Computing Machinery, New York, NY, USA, pp. 179–188. http://dx.doi.org/10.1145/2470654.2470679.

Speicher, M., Cao, J., Yu, A., Zhang, H., Nebeling, M., 2018. 360Anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. Proc. ACM Hum.-Comput. Interact. 2 (EICS), 9. http://dx.doi.org/10.1145/3229091, 20 pages.

Tait, M., Billinghurst, M., 2015. The effect of view independence in a collaborative AR system. Comput. Support. Coop. Work 24 (6), 563–589. http://dx.doi.org/10.1007/s10606-015-9231-8.

Tang, J.C., Minneman, S.L., 1990. VideoDraw: A video interface for collaborative drawing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '90, Association for Computing Machinery, New York, NY, USA, pp. 313–320. http://dx.doi.org/10.1145/97243.97302.

Teo, T., Lawrence, L., Lee, G.A., Billinghurst, M., Adcock, M., 2019. Mixed reality remote collaboration combining 360 video and 3D reconstruction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, New York, NY, USA, pp. 1–14. http://dx.doi.org/10.1145/3290605.3300431.

Teo, T., Sakurada, K., Fukuoka, M., Sugimoto, M., 2022. Evaluating techniques to share hand gestures for remote collaboration using top-down projection in a virtual environment. In: Uchiyama, H., Normand, J.-M. (Eds.), International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments. In: ICAT-EGVE 2022, The Eurographics Association, pp. 1–9. http://dx.doi.org/10.2312/egve.20221272.

Tian, H., Lee, G.A., Bai, H., Billinghurst, M., 2023. Using virtual replicas to improve mixed reality remote collaboration. IEEE Trans. Vis. Comput. Graphics 29 (5), 2785–2795. http://dx.doi.org/10.1109/TVCG.2023.3247113.

Unity Technologies, 2024. Unity. URL: https://unity.com/.

Vidal-Balea, A., Blanco-Novoa, O., Fraga-Lamas, P., Vilar-Montesinos, M., Fernández-Caramés, T.M., 2020. Creating collaborative augmented reality experiences for industry 4.0 training and assistance applications: Performance evaluation in the shipyard of the future. Appl. Sci. 10 (24), 9073. http://dx.doi.org/10.3390/app10249073, 23 pages.

W3C, 2022. WebXR device API. URL: https://immersive-web.github.io/webxr/.

W3C, 2023. WebRTC: Real-time communication in browsers. URL: https://www.w3.org/TR/webrtc/.

W3C, 2024. WebXR hand input module. URL: https://www.w3.org/TR/webxr-hand-input-1/.

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

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Zhang, X., Wang, S., He, W., Yan, Y., Ji, H., 2021b. AR/MR remote collaboration on physical tasks: A review. Robot. Comput.-Integr. Manuf. 72, 102071. http://dx.doi.org/10.1016/j.rcim.2020.102071, 32 pages.

Wang, X., Love, P.E.D., Kim, M.J., Wang, W., 2014. Mutual awareness in collaborative design: An augmented reality integrated telepresence system. Comput. Ind. 65 (2), 314–324. http://dx.doi.org/10.1016/j.compind.2013.11.012.

Wang, Z., Zhang, S., Bai, X., 2021c. A mixed reality platform for assembly assistance based on gaze interaction in industry. Int. J. Adv. Manuf. Technol. 116 (9), 3193–3205. http://dx.doi.org/10.1007/s00170-021-07624-z.

Wang, P., Zhang, S., Bai, X., Billinghurst, M., He, W., Sun, M., Chen, Y., Lv, H., Ji, H., 2019. 2.5DHANDS: A gesture-based MR remote collaborative platform. Int. J. Adv. Manuf. Technol. 102 (5), 1339–1353. http://dx.doi.org/10.1007/s00170-018-03237-1.

WHATWG, 2023a. DOM standard. URL: https://dom.spec.whatwg.org/.

WHATWG, 2023b. WebSockets. URL: https://websockets.spec.whatwg.org/.

Xiang, S., Wang, R., Feng, C., 2021. Mobile projective augmented reality for collaborative robots in construction. Autom. Constr. 127, 103704. http://dx.doi.org/10.1016/j.autcon.2021.103704, 17 pages.

Yang, P., Kitahara, I., Ohta, Y., 2015. Remote mixed reality system supporting interactions with virtualized objects. In: 2015 IEEE International Symposium on Mixed and Augmented Reality. IEEE, Fukuoka, Japan, pp. 64–67. http://dx.doi.org/10.1109/ISMAR.2015.22.

Young, J., Langlotz, T., Cook, M., Mills, S., Regenbrecht, H., 2019. Immersive telepresence and remote collaboration using mobile and wearable devices. IEEE Trans. Vis. Comput. Graphics 25 (5), 1908–1918. http://dx.doi.org/10.1109/TVCG.2019.2898737.

Yu, K., Eck, U., Pankratz, F., Lazarovici, M., Wilhelm, D., Navab, N., 2022. Duplicated reality for co-located augmented reality collaboration. IEEE Trans. Vis. Comput. Graphics 28 (5), 2190–2200. http://dx.doi.org/10.1109/TVCG.2022.3150520.

Yu, K., Winkler, A., Pankratz, F., Lazarovici, M., Wilhelm, D., Eck, U., Roth, D., Navab, N., 2021. Magnoramas: Magnifying dioramas for precise annotations in asymmetric 3D teleconsultation. In: 2021 IEEE Virtual Reality and 3D User Interfaces. VR, IEEE, Lisboa, Portugal, pp. 392–401. http://dx.doi.org/10.1109/VR50410.2021.00062.

Zhang, X., Bai, X., Zhang, S., He, W., Wang, P., Wang, Z., Yan, Y., Yu, Q., 2022. Real-time 3D video-based MR remote collaboration using gesture cues and virtual replicas. Int. J. Adv. Manuf. Technol. 121 (11), 7697–7719. http://dx.doi.org/10.1007/s00170-022-09654-7.

Zillner, J., Mendez, E., Wagner, D., 2018. Augmented reality remote collaboration with dense reconstruction. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct. ISMAR-Adjunct, IEEE, Munich, Germany, pp. 38–39. http://dx.doi.org/10.1109/ISMAR-Adjunct.2018.00028.

Zillner, J., Rhemann, C., Izadi, S., Haller, M., 2014. 3D-board: A whole-body remote collaborative whiteboard. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology. UIST '14, Association for Computing Machinery, New York, NY, USA, pp. 471–479. http://dx.doi.org/10.1145/2642918.2647393.