



# BeHere: a VR/SAR remote collaboration system based on virtual replicas sharing gesture and avatar in a procedural task

Peng Wang<sup>1</sup> · Yue Wang<sup>1</sup> · Mark Billingham<sup>2</sup> · Huizhen Yang<sup>3</sup> · Peng Xu<sup>4</sup> · Yanhong Li<sup>5</sup>

Received: 18 March 2022 / Accepted: 4 January 2023 / Published online: 13 January 2023  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

## Abstract

In this paper, we focus on the significance of remote collaboration using virtual replicas, avatar, and gesture on a procedural task in industry; thus, we present a Virtual Reality (VR)/Spatial Augmented Reality (SAR) remote collaboration system, BeHere, based on 3D virtual replicas and sharing gestures and avatar. BeHere enables a remote expert in VR to guide a local worker in real-time to complete a procedural task in the real-world. For the remote VR site, we construct a 3D virtual environment using virtual replicas, and the user can manipulate them by using gestures in an intuitive interaction and see their partners' 3D virtual avatar. For the local site, we use SAR to enable the local worker to see instructions projected onto the real-world based on the shared virtual replicas and gestures. We conducted a formal user study to evaluate the prototype system in terms of performance, social presence, workload, and ranking and user preference. We found that the combination of visual cues of gestures, avatar, and virtual replicas plays a positive role in improving user experience, especially for remote VR users. More significantly, our study provides useful information and important design implications for further research on the use of gesture-, gaze- and avatar-based cues as well as virtual replicas in VR/AR remote collaboration on a procedural task in industry.

**Keywords** Virtual reality · Spatial augmented reality · Virtual replicas or avatar · Remote collaboration · Hand gesture · Human–computer interaction

## 1 Introduction

This research aims to explore the significance and the influence of Virtual Reality (VR)/Augmented Reality (AR) remote collaboration on procedural tasks in industry by

means of virtual replicas of physical–mechanical parts, users' gesture and avatar. Since 2019, the COVID-19 pandemic has had a major impact on our life and work style and accelerated the need for remote collaboration tools, especially in areas such as emergency response, remote education, telemedicine, remote technical services, construction, and training (Wang et al. 2019d, 2021a; Billingham 2021; Calandra et al. 2021; Garbett et al. 2021; Marques et al. 2021). In this paper, we discussed how VR/AR can be used to share non-verbal cues (e.g., gestures, avatar, and virtual replicas) to improve remote collaboration on procedural tasks. These tasks are typical cooperative work in many industrial contexts and construction, for instance, remote maintenance or repair in real-time (Kurillo and Bajcsy 2013; de Belen et al. 2019; Ens et al. 2019; Barroso et al. 2020; Russo 2021; Wang et al. 2021a).

Remote collaboration on physical tasks often involves building new virtual collaborative environments in which geographically dispersed participants can convey actions to each other. However, with traditional collaboration tools (e.g., desktop video conferencing), there can be an artificial

✉ Peng Wang  
ilovemymfandb@163.com

✉ Yue Wang  
wang\_yue@cqupt.edu.cn

<sup>1</sup> School of Advanced Manufacturing Engineering,  
Chongqing University of Posts and Telecommunications,  
Chongqing 400065, China

<sup>2</sup> Auckland Bioengineering Institute, The University  
of Auckland, Auckland, New Zealand

<sup>3</sup> Chongqing Innovation Center of Beijing Institute  
of Technology, Beijing, China

<sup>4</sup> Jiangsu JARI Technology Group Co., Ltd, The 716Th  
Research Institute of CSIC, Lianyungang, Jiang Su, China

<sup>5</sup> Jiuquan Iron and Steel (Group) Co., Ltd,  
Jiayuguan City 735100, China

separation between a view of the task space and the communication space (Wang et al. 2019a, b, c, d, 2021a, b, 2022). Users cannot see the face of the person they are working with a view at the same time seeing a view of the remote workspace. So, there is a separation between user actions and the sense of a shared collaborative environment or between a task operation space and the interaction space of virtual information, sometimes leading to communication breakdown that can render even the easiest of tasks problematic. However, VR and AR technologies provide promising ways of overcoming these problems by offering 3D immersive experiences, virtual-real fusion, and natural human–computer interaction (Bottani and Vignali 2019; de Belen et al. 2019; Wang et al. 2020a, 2021b; Jasche et al. 2021; Marques et al. 2021).

There has been significant previous research exploring the effects of sharing gestures, 3D virtual replicas, and virtual avatars in VR/AR remote collaboration. For example, Kim et al. (2020) investigated how two factors (e.g., the distance to the target, the perspective between users) influence the effects of gesture cues for Mixed Reality (MR) remote collaboration. Kritzler et al. (2016) proposed an AR remote collaborative platform, RemoteBob, using 3D virtual replicas and annotations.

These studies showed that sharing non-verbal cues (e.g., 3D virtual replicas, gesture, and avatar) has a positive impact on remote collaboration in terms of performance time, user experience, cognitive load, and decision making. However, it was found that there has been little research that takes advantage of 3D virtual replicas, gestures, and avatar in VR/AR remote collaboration, particularly in assembly training in manufacturing.

In this paper, we describe a novel system, *BeHere*, which uses VR/AR to share user's gesture and avatar cues to address some issues of telepresence, such as the sense of co-presence and making clear gesture-based instructions. The system uses VR and SAR to support remote collaboration based on 3D virtual replicas for a procedural task. Prior research (Orts-Escolano et al. 2016; Wang et al. 2019a, 2021b) has strongly suggested that gesture- and avatar-based embodiments and 3D virtual replica-based instructions produce more effective ways to fluidly communicate about physical tasks. Nevertheless, little is known about the influence of combining these non-verbal cues in VR/AR remote collaboration in terms of performance time and user preference. Therefore, there is a golden opportunity to explore whether combining 3D virtual replicas and normally used non-verbal cues (e.g., gesture and avatar) could improve VR/AR remote collaboration for assembly training in manufacturing.

Compared with our previous similar studies (Wang et al. 2019a, b, c, 2020a, 2021b), which focus on exploring the

effects of sharing remote users' gesture, gaze, head pointer, and combination cues based on gesture and gaze in VR/AR remote collaborative tasks in manufacturing, however, this research pays more attention to explore the influence of sharing local users' avatar cues and virtual replicas of physical mechanical parts based on gesture interaction in procedural tasks. Thus, we have made significant improvements, and most of these updates are presented at length in Sect. 3 and conducted a formal user study using the proposed prototype system in Sects. 4 and 5.

This research was mainly motivated by earlier research (Wang et al. 2019a, 2021b; Yang et al. 2020). Our work builds on these previous studies and extends them. More particularly, we explore how instructions based on virtual replicas, combining gesture and avatar, affect VR/SAR remote collaboration for a procedural task. Compared to previous work, the novelty of our research is threefold:

- (1) We present a novel VR/SAR remote collaborative system, *BeHere*, which uses SAR in the local site and VR in the remote site, to provide instructions based on virtual replicas, while sharing the VR user's gesture and the SAR user's avatar.
- (2) We report on one of the first user studies to explore how sharing the local user's avatar and the remote user's gesture can affect VR/SAR remote collaboration on physical tasks.
- (3) We provide the implementation details and evaluate the system in terms of effectiveness and user experience.

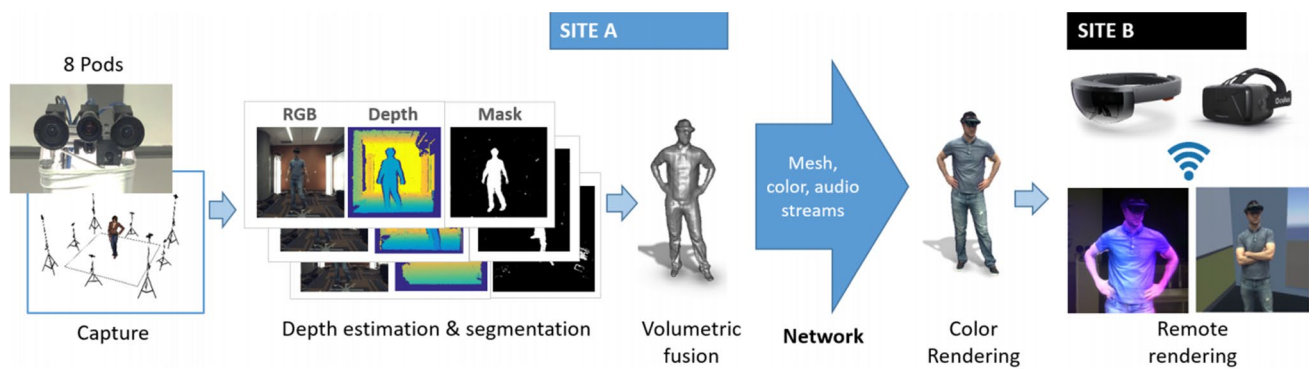
In the following sections, we first present previous related works. Second, we describe the implementation details of our VR/SAR system. Third, we conduct a pilot test and a formal user study to evaluate the prototype system. Then, we discuss the user study results. Next, the limitations and future work are presented. Finally, the conclusion is presented in the last section.

## 2 Related works

Our work builds on previous related research in VR/AR remote collaboration on sharing non-verbal communication cues in the areas of (1) supporting gesture cues, (2) supporting avatar cues, and (3) supporting 3D virtual replicas cues.

### 2.1 Gesture cues

In VR/AR remote collaboration sharing gestures can provide natural communication cues for collaborators to work together remotely in a way that reduces the workload and enhances the feeling of co-presence feeling and user



**Fig. 1** Holoportation framework (Orts-Escolano et al. 2016)

experience (Wang et al. 2019a; Kim et al. 2020). With the rapid development of depth sensors such as Leap Motion and Intel RealSense, gesture-based natural interaction opens a new opportunity for VR/AR human–computer interaction. For example, Huang et al. (2018) developed an MR multi-modal collaborative system, HandsIn3D, that captured and shared the remote expert’s 3D gestures in real-time. Using the system, they explored the effects of sharing gesture and stereoscopic rendering to improve the sense of immersion and co-presence. Wang et al. (2019c) explored the effects of sharing the remote expert’s 2.5D gestures in a SAR-based remote collaboration for an assembly task.

Adding gaze cues to gesture could create even more intuitive and natural interaction. For example, Bai et al. (2020) investigated how sharing gesture and gaze cues from the remote user to the local user could affect MR remote collaboration for searching and picking up Lego bricks in terms of the task performance, the sense of co-presence, mental and physical workload, and social presence. The research found that combining gaze and gesture cues had a positive effect on remote collaborative work, specifically, and provided a stronger sense of co-presence. Higuch et al. (2016) developed a SAR remote collaborative platform based on gaze and gestures, which can be shared with a local user. However, their interface had a problem when the remote user interacted with the mobile display, and his or her gestures were captured by a side sensor; thus, the user maybe unclear about the spatial relationship between their hand and the virtual target in the shared scene. To address this issue, Wang et al. (2019c) created an MR remote collaborative platform based on Gesture and Head Pointing (GHP) to study the influence of combining gesture and head pointing cues in a typical assembly task. They found that the GHP interface significantly improved the remote collaborative experience in terms of empathy and interaction compared to AR annotations. Although there are some AR/MR remote collaborative systems focusing on physical tasks, they paid more attention to the advantages of the naturalness and flexibility of gesture

interaction. Therefore, these works can be improved by multimodal interaction based on other cues.

## 2.2 Avatar cues

With the rapid development of behavior tracking technology, for example, human pose estimation/detection, this can be used in VR/AR remote collaborative systems to create the experience and sense of co-presence where remote collaborators feel like they are face to face. Piumsomboon et al. (2018a, b) presented a novel MR remote collaborative system, Snow Dome, sharing avatar and gesture cues and supporting multi-scale interaction for remote VR users. Based on this study, they improved the Snow Dome system by creating a flying telepresence interface enabling collaborators to work at a larger scale (Piumsomboon et al. 2018b). However, it concentrates on sharing the remote user’s avatar. In our research, we explore the effects of sharing the local user’s avatar in remote collaboration during a typical assembly training. Orts-Escolano et al. (2016) introduced a VR/AR telepresence system, Holoportation (see Fig. 1), allowing users to see, hear, and interact with the shared avatar as well as virtual objects viewed by VR/AR displays. It makes collaborators feel like they were co-located in the same physical space and has a positive influence on remote collaborative works, such as business meetings, family gatherings, and providing dancing instructions. However, the system required lots of high-end hardware, depth cameras, and the configuration is complex. Recently, De Pace et al. (2019) indicated that shared 3D avatar cues can enhance the sense of co-presence in MR remote collaboration on industrial training and repair procedures.

The above-mentioned studies showed that sharing avatar-based cues is very useful for improving AR/MR remote collaborative systems. For assembly training in industry, nevertheless, the influence of combining 3D virtual replicas and these commonly used non-verbal cues (e.g., gestures and avatar) has not been well explored.

### 2.3 Virtual replicas cues

There are 3D CAD models of many parts stored in the repository in industry (Huang et al. 2015), particularly for design, manufacturing, and assembly processes. They can be explored and reused in remote collaboration to provide cues assembly training. For example, Oda et al. (2015) and Elvezio et al. (2017) developed two kinds of 3D interaction and visualization methods (e.g., POINT3D and DEMO3D) using 3D CAD models or virtual replicas in VR/AR remote collaboration to provide effective spatial referencing and demonstration. A user study showed that the instructions based on virtual replicas could reduce time-consuming and error-prone operations in an aircraft engine assembly task. For routine repair tasks, Kritzler et al. (2016) created a telepresence remote expert system, RemoteBob, that could allow remote users to share instructions using virtual replicas with on-site workers and found that virtual replicas cues could avoid miscommunication, provide clear visual cues, and reduce operations errors. However, there is a fractured ecology of the local or remote interface which easily makes a distraction for users.

Recently, Pringle et al. (2018) proposed an AR guidance system based on virtual replicas for real-world maintenance, Yaw Motor Servicing. Although this study does not pay more attention to remote collaborative tasks, it demonstrates the advantages of AR instructions using virtual replicas in industry. Similarly, Elvezio et al. (2015) and Sukan et al. (2016) showed a novel AR-based remote collaborative system, HANDLES, which can provide real-time instructions for 3D rotation operations, and they also indicated that using virtual replicas could support clear visible guidance. However, these researches could be improved by using the combination of gesture cues and avatar cues.

### 2.4 Summary

From the related work on VR/AR/MR remote collaboration discussed above, we can reach four conclusions. (1) There is a lot of research supporting gesture-based instructions in remote collaboration on physical tasks, and multimodal interaction based on gesture and gaze is attracting more and more attention from researchers. Overall, this research has found that sharing gesture and gaze-based cues could significantly enhance remote collaborative works in terms of performance and user experience. (2) Although some research has explored the effects of sharing avatar cues in remote collaboration, there is relatively little research investigating how avatar-based cues would affect remote collaboration for procedural tasks in industry. (3) Researchers have successfully demonstrated the advantages of sharing virtual replicas in industry. More critically, these studies will open up new methods for improving VR/AR/MR remote collaborative

works in an industrial setting. (4) The combination of virtual replicas and non-verbal cues (e.g., avatar, gesture, and annotations) for VR/AR/MR remote collaboration on real-world tasks for procedural tasks in industry has not been fully explored.

To address these problems, we propose a new VR/SAR remote collaborative platform, *BeHere*, where provides clear instructions using avatar-, gesture-, and virtual-replica-based cues to explore the effects of these non-verbal cues in remote collaboration on procedural tasks. It is worth mentioning that especially why the proposed system uses VR and SAR, instead of VR and VR, SAR and VR or others, and the combination of gestures, avatar with virtual replicas. For this problem, the local user must manipulate physical objects during remote collaboration; thus, SAR is a good choice considering the advantages of keeping local workers' hands free and without needing to wear and operate any devices. More importantly, the research (Hietanen et al. 2020) has demonstrated that SAR has many merits (e.g., safety, competence, and ergonomics) compared with wearable AR. Using VR, the prototype could provide remote users with a 3D virtual collaborative environment which can support the rich and free expression by natural interaction based on gesture. Moreover, previous studies (Oda et al. 2015; Kritzler et al. 2016; Elvezio et al. 2017; Wang et al. 2021b) have demonstrated that sharing virtual replicas can provide on-site workers with clear instructions.

## 3 Prototype Implementation

### 3.1 System Architecture

Our prototype system is developed based on the proposed framework (see Fig. 2) and is a multi-user remote collaborative system. It has four key elements: (1) the local site based on SAR to provide instructions and using a depth sensor (Kinect) to detect human pose, (2) the remote site using the HTC Vive VR HMD to provide a virtual 3D immerse collaborative environment and using a gesture sensor (Leap Motion) to detect the VR user's hand gestures, (3) the server responsible for transmitting data, (4) the expansion interface for other remote collaborative VR/AR clients.

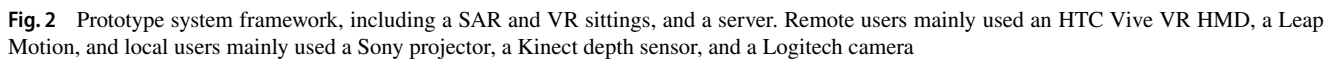
The prototype system, running on the Windows10 Operating System, was developed using the Unity 3D<sup>1</sup> game engine (version 2020.3.18), the MixedRealityToolkit (MRT)<sup>2</sup>, Wampserver,<sup>3</sup> and OpenCvSharp.<sup>4</sup> In our current studies, the system has two clients: a remote VR client and a local

<sup>1</sup> <https://unity.com>.

<sup>2</sup> <https://github.com/microsoft/MixedRealityToolkit-Unity>.

<sup>3</sup> <http://www.wampserver.com>.

<sup>4</sup> <https://github.com/shimat/opencvsharp>.



For the local SAR settings, the client was implemented on an HP ENVY laptop (Intel Core i9-10885H, 32G DDR4 2993 MHz, RAM 16 GB, SSD 2 TB, and NVIDIA GeForce RTX2060 MQ) with a Sony projector, a Kinect depth sensor, and a Logitech camera. For the remote VR settings, the client was running on a desktop computer (Intel(R) Core(TM) i7-10700 K, CPU @ 3.8 GHz, RAM 16 GB, and NVIDIA GeForce RTX 2070 SUPER) with a Leap Motion and an HTC Vive Eye Pro VR HMD. For other interfaces, the system could be extended by adding other VR/AR devices/toolkits (e.g., HoloLens, Oculus Rift, Magic Leap, Meta2, and mobile phone) for supporting multi-user remote collaboration.

For our AR/SAR remote collaborative system, each client can support loading virtual replicas from the server. There are two key preparations before the BeHere system is running.

<sup>6</sup> <https://docs.unity3d.com/Manual/AssetBundles-Workflow.html>.

 Springer

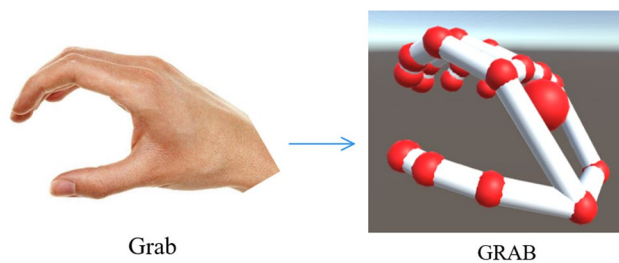


Fig. 3 “GRAB” gesture for interaction (Wang et al. 2021b)

The second step is to remotely access scene resources, which can be carried over the network then shared to local and remote clients by means of WampServer. Local and remote clients can browse the prefabs and asset-bundles by Explorer. When the prototype system is running, each client is able to load resources into the Unity 3D scene using MRT, and the system provides the local site with the same virtual collaborative scenario as the remote site.

### 3.3 Gesture interaction

Prior research (Kim et al. 2019; Piumsomboon et al. 2019; Teo et al. 2019; Wang et al. 2019d, 2021a, b) has demonstrated that gesture-based interaction is commonly used in VR/AR-based studies. Following the work of Wang et al. (2021b), we used the dynamic “grab” gesture (shown in Fig. 3) which can be recognized by the Leap Motion sensor. The “grab” gesture is consistent with our way of grabbing objects in daily life.

The algorithm framework of the gesture-based interaction is shown in Fig. 4. It has four key steps: (1) collision detection between the dynamic gesture recognized by the Leap Motion and the virtual replicas/objects (VO), (2) recognizing the “GRAB” gesture as shown in Fig. 4, (3) calculating the distance between the users’ gesture and the nearest virtual object, and (4) updating the position mapping from gesture to VO. For the third step, firstly we obtained the pinched midpoint (PM) between which the thumb(T) and index(I) finger. Thus, we get:

$$PM = [T(x, y, z) + I(x, y, z)]/2 \quad (1)$$

where PVO is the position of VO, and PMVO is the distance between PM and PVO. So, we get.

$$PMVO = \sqrt{[(T_x - I_x)^2 + (T_y - I_y)^2 + (T_z - I_z)^2]} \quad (2)$$

When the value of PMVO is less than  $\gamma$  mm, we assign the position of the gesture “GRAB” to the VO. TPVO is the distance between the target position and VO. Eventually, when the value of TPVO is less than  $\delta$  mm, the virtual replicas will automatically adsorb to the right position.

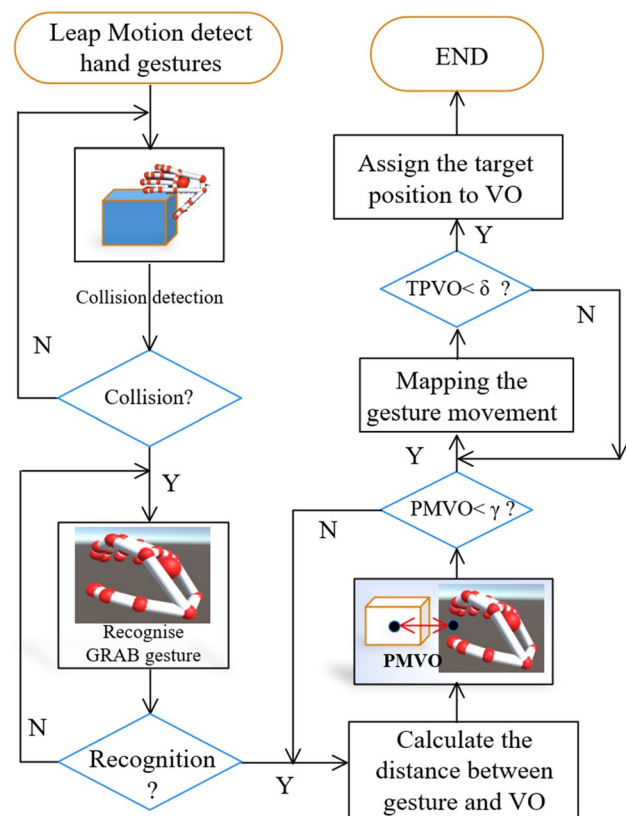


Fig. 4 Interaction algorithm framework

### 3.4 Sharing gesture- and avatar-based cues

Figure 5 illustrates the framework for sharing gestures and gesture-based interaction. About sharing gestures, our method mainly includes five steps as follows:

- (1) Hand tracking. We used the Leap Motion hand-tracking sensor to recognize the gesture. The Leap Motion sensor is commonly used in VR-based interaction and can be perfectly connected with HTC Vive HMD.
- (2) Collecting and sharing data. To make the algorithm efficient and concise, we obtained all joint positions of the user’s two hands when the gestures can be detected using the Leap Motion. Otherwise, when the gestures cannot be detected using the Leap Motion, we set the gesture data for the joint positions to zero. For these two situations, we allocated the same memory space. Next, the data were shared by the MRT framework.
- (3) Building a gesture model. We built a virtual gesture model by referring to the Leap Motion’s gesture model for facilitating data processing. In the current study, we used the skeleton gesture model according to user preferences.

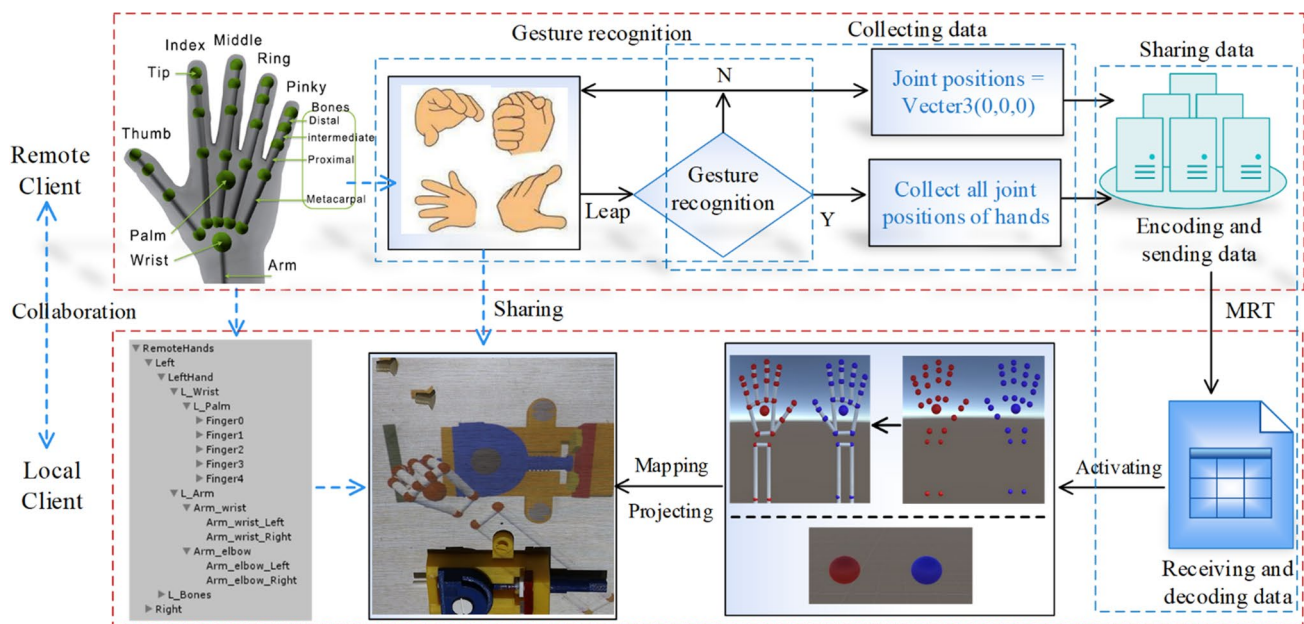


Fig. 5 Systematic overview of sharing gesture and gesture-based interaction

- (4) Activating gesture. The remote VR user's gestures can be consistently mapped to virtual hands on the local site. Thus, the local site's hand model would be activated by the shared gesture data.
- (5) Real-virtual fusion and projecting. On the local site, the prototype needs to calibrate the projector-camera pair to correctly provide the shared instructions in the on-site environment. We performed the calibration process referring to Wang et al. (2019c) as described in Sect. 3.2.2.

For sharing the user's 3D avatar, we used the Kinect depth sensor to detect the local user's human pose. We shared the avatar data to be displayed with a 3D avatar in the remote VR environment. This could increase mutual awareness and improve social presence (Kurillo and Bajcsy 2013; Chen et al. 2021). Overall, the framework for sharing an avatar is similar to sharing gestures. The difference is the fifth step, calibration, where we should calibrate the position of the local user's 3D avatar in the remote VR environment. This can be performed by mapping the relationship between a user and physical objects in real-time.

### 3.5 Sharing stereoscopic scene

Overall, the pipeline overview of our 3D capture module mainly has four steps: (1) the RGBD camera first captures the live depth and RGB frames in the local worker side, (2) after one RGB + D frame is captured, this integrated data

are encoded and streamed to the remote side through MRT, (3) once the RGB + D data stream is received by the remote expert side and decoded, each frame is reconstructed in real-time into textured 3D point cloud to render on the VR HMD, and (4) repeat steps (1) to (3) to real-time update the shared 3D stereoscopic scene of the local workspace. For the system architecture and MRT, more details about sharing a stereoscopic scene through an RGBD camera, our approach is based on Zhang et al. (2022) and Anton et al. (2018).

## 4 Pilot test

### 4.1 Participants and procedures

We performed a pilot test to evaluate the usability of the prototype and decide where the video stream of the remote helper should be rendered in the remote VR environment. We conducted a within-subjects study in the pilot test and recruited twelve volunteers, all of whom were university students. The volunteers' ages ranged from 18 to 26 years [9 males, 3 females, mean 22.58, standard deviation (SD) 2.33, standard error (SE) 0.67]. They all answered a short questionnaire about background information, including age, and experience with VR/AR, remote collaboration, and video conference using QQ and WeChat, as shown in Table 1. Most volunteers were novices in the field of VR/AR and remote collaboration, and we arranged participants with VR or AR experience on the remote or local site, respectively, to

**Table 1** Volunteers' demographic information

|                                     | None (%) | Novice (%) | Frequency (%) |
|-------------------------------------|----------|------------|---------------|
| VR                                  | 25       | 67         | 8             |
| AR                                  | 8        | 75         | 17            |
| Remote collaboration                | 17       | 83         | 0             |
| Video conference using QQ or WeChat | 0        | 58         | 42            |

make the results more convincing. Furthermore, most participants knew one another.

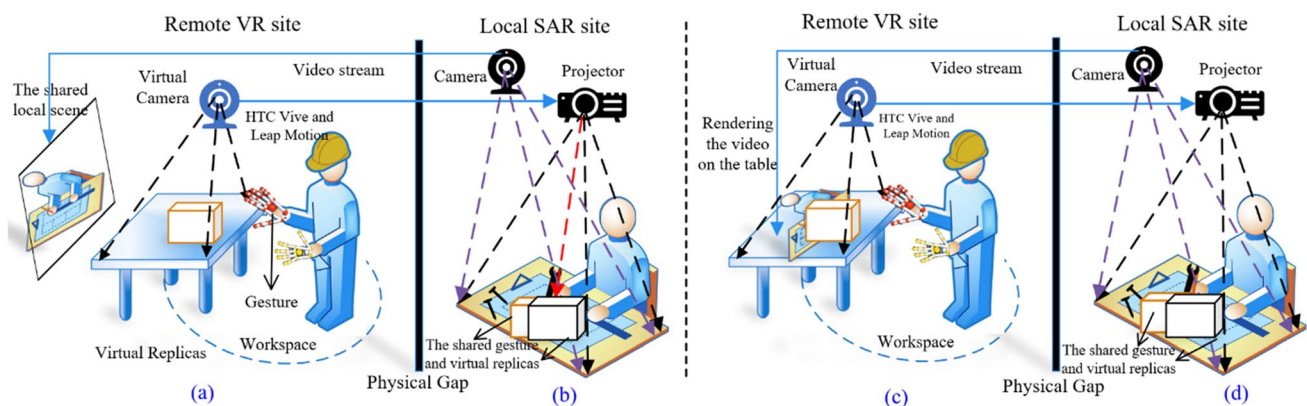
## 4.2 Conditions and procedural task

Figure 6 shows the prototype system used in the pilot test. The VR/SAR remote collaborative system supports sharing virtual replicas to assist with assembly of physical objects. There were two conditions explored in this pilot study: (1)

the *VIRE* condition, in which the shared local live video was rendered on a vertical plane in the VR environment (see Fig. 6a, b), and the *VRA* condition in which the live video was rendered on the table in the VR environment (see Fig. 6c, d). In the literature review, we can conclude that the *VIRE* condition is the existing system for procedural tasks in industry.

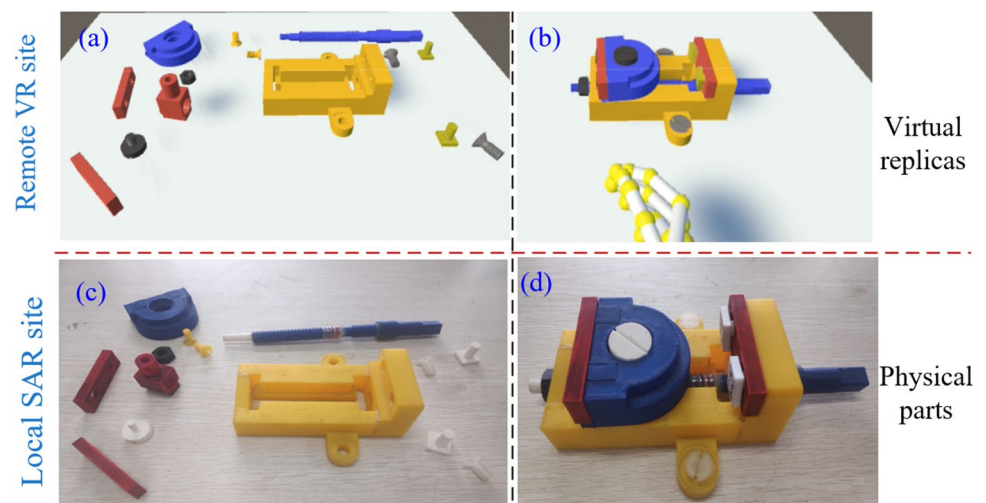
In the current study, we chose a typical procedural task in industry, assembly of a vise, as shown in Fig. 7. The vise includes fourteen 3D printed parts, and the procedural task should follow the assembly process. Figure 7a, b shows the task scene on the remote VR site, and Fig. 7c, d shows the task scene on the local SAR site.

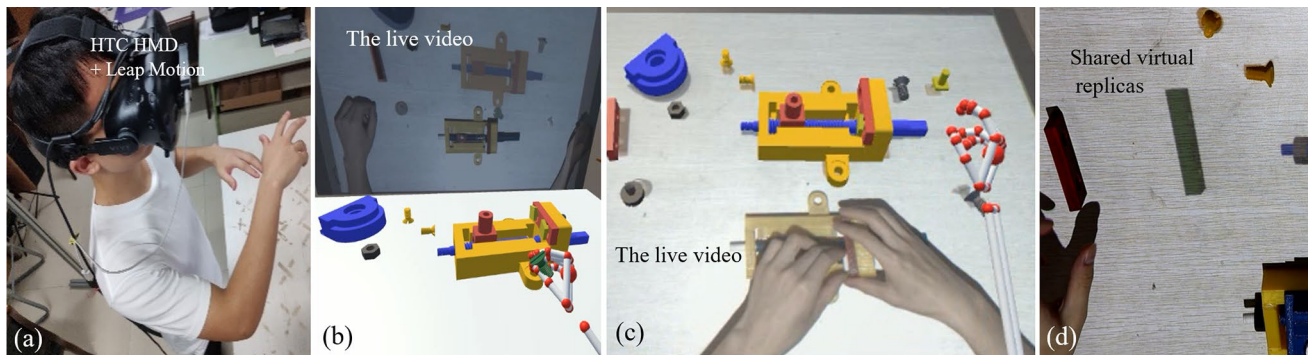
Figure 8 shows the VR/SAR remote collaborative scene for the typical procedural task during the pilot test. To reduce learning effects, we counterbalanced the procedural task for the two different conditions in accordance with a Latin Square Sequence. After each trial, all volunteers were to fill in the System Usability Scale questionnaire [SUS



**Fig. 6** General framework of our VR/SAR remote collaboration system **a, b** the *VIRE* condition, **c, d** the *VRA* condition

**Fig. 7** A typical procedural task using the vise, **a, c** the assembly before, **b, d** the assembly completion





**Fig. 8** VR/SAR remote collaborative scenario. **a** A remote user provides instructions according to the local situation. **b, c** The HTC Vive HMD view on the remote site. **d** The on-site worker scene, **b, d** the VIRE condition (see Fig. 4), **c, d** the VRA condition

(Brooke 1996)]. They had the chance to freely explore the prototype system, and then, we collected the users' feedback.

### 4.3 Results and discussion

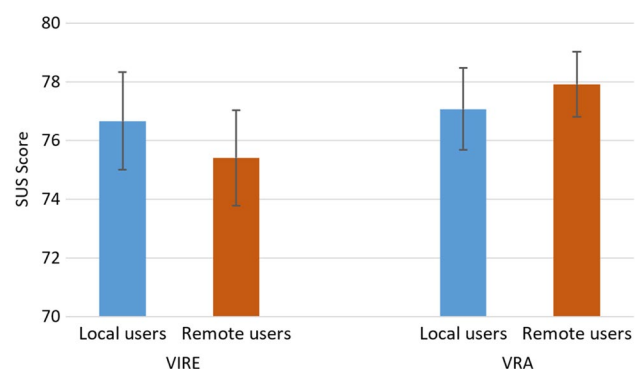
The goal of the pilot study was to compare the VIRE and VRA conditions and see if the different placement of the live video of the local scene in the VR view affected usability and collaboration. Figure 9 shows the evaluation results of the system usability. For more detailed information on interpreting and calculating SUS scores, the reader should refer to the SUS website.<sup>7</sup> For the VIRE condition rendering the shared live video in front of the VR users, the average SUS scores were 76.67 (SD 4.38) and 75.42 (SD 4.31) for local users and remote users, respectively. For the VRA condition rendering the video stream on the virtual table, the average SUS scores were 77.08 (SD 3.68) and 77.92 (SD 2.92) for local users and remote users, respectively. According to Bangor et al. (2009), SUS scores in this range means that the prototype has good usability for both of the tested conditions. Wilcoxon's signed-rank test ( $\alpha=0.05$ ) revealed that there was no significant difference between the VIRE and VRA conditions.

We also collected feedback from the participants on the VIRE and VRA conditions. Most participants (83%, five remote users) liked the VRA condition more than the VIRE condition and felt that it was easier to create a sense of co-presence than the VIRE condition. They generally provided positive feedback such as "It is very interesting to grab virtual replicas by the gesture-based interaction; Easy to learn, the interaction is very similar to our usual way of grabbing physical objects. So I did not need to do a lot of practice before using it; I would like to use it in remote collaborative

work; It is wonderful to see my hand gesture in the VR environment. The VRA interface makes me more focus on the collaborative work. S/he provides me with clear instructions; I like the interface because there is not other interaction and what I should do is to just follow my partner."

However, some participants provided constructive feedback such as "It would be amazing if I can see my partner's hand gesture during remote collaboration; I think that sharing gesture cues could improve the sense of co-presence; I think that sharing gesture could use efficient deictic references; I think that it would be wonderful If I could see my partner's 3D; I think that it will be wonderful If I can see my partner's 3D avatar just like my gestures I see when I immersed in 3D VR environment, which allows us to feel that we are co-located in a shared VR space as face to face communication; I find that it is occasionally awkward when I grabbed small VO."

Based on their positive and constructive feedback, first, we improved our VRA interface. More specifically, firstly, we improved the algorithm for gesture-based interaction by adjusting the threshold of  $\gamma$  and  $\delta$  to solve the problem of the difficulty of grasping small VO. Secondly, we enabled



**Fig. 9** SUS results from the pilot test and on the left and right are the results of the VIRE and VRA conditions, respectively

<sup>7</sup> <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>  
<https://measuringu.com/sus>.

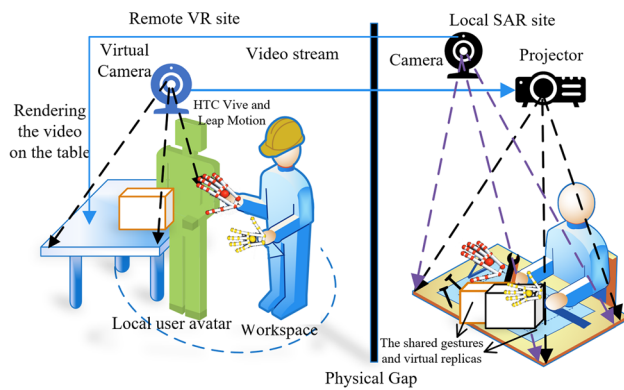


Fig. 10 Prototype system: *BeHere*

the system to support sharing virtual replicas and gestures (VRAG) based on the VRA interface. That is, the VRAG condition becomes the VRA condition when the shared gesture is disabled on the local site. Finally, the prototype system can support sharing local users' avatar cues on basis of the VRAG interface. Therefore, we define it as *BeHere* (see Fig. 10) to distinguish it from other interfaces.

## 5 User study

We improved the VRA system used in the pilot test by sharing gesture- and avatar-based cues. With the final *BeHere* prototype system, we were interested in evaluating the effects of sharing gesture and avatar cues separately and together in the VR/SAR remote collaborative system for a procedural task in VR/SAR remote collaborative tasks. We conducted a user study to explore the usability of these visual cues. In the following section, we describe the user study details and the results.

### 5.1 Experimental details

#### 5.1.1 Conditions

In the current research, our main independent variable was the visual cues (e.g., avatar) which were shared from the local site to the remote user, on the basis of combining gestures and virtual replicas cues which were shared from the remote to the local user. Therefore, our user study has two communication conditions which were consistent with the participants' constructive feedback in the pilot test as follows:

- (1) *VRAG*: the prototype system providing the local SAR user with instructions based on the combination of gestures and virtual replicas.

- (2) *BeHere*: the prototype system supports sharing the local user's 3D avatar in addition to the VRAG condition. In other words, the *BeHere* condition shared gestures and virtual replicas like the VRAG condition, but in addition can show a shared avatar.

#### 5.1.2 Task and hypotheses

For the experimental task, we chose a typical procedural task as illustrated in Fig. 8, assembling a vise. However, we added two constraints for encouraging collaboration. First, the local SAR participants should take only one part at a time from the assembly area. Second, the local participants had to assemble the parts one by one according to the shared instructions from the remote VR site.

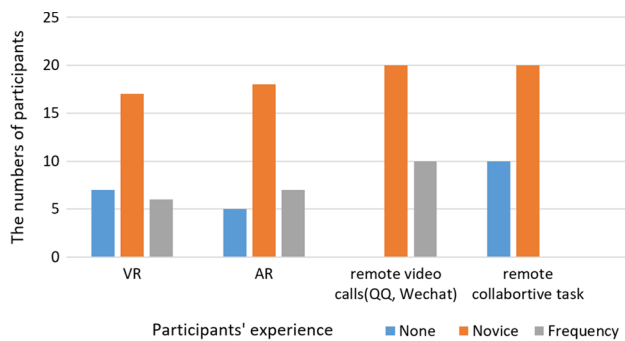
In the user study, the focus of our study is on the evaluation of the impact of sharing gesture and avatar cues for a procedural task. Consequently, we mainly explored the following two research questions: (1) What are the benefits of combining gesture and avatar cues for remote collaboration on a procedural task? (2) How does the sharing of gestures from the remote user or avatar from the local user affect remote collaboration on a procedural task?

In general, the three visual cues (e.g., gestures, avatar, and virtual replicas) have different merits in providing a sense of co-presence and clear instructions. More importantly, these visual cues have been demonstrated to have a positive impact to some extent on remote collaborative tasks in terms of service quality and efficiency. However, no prior study has researched the combination of these visual cues. Thus, our research has the following four hypotheses:

- H1: The *BeHere* condition would improve task performance compared with the VRAG condition.
- H2: The *BeHere* condition is better than the VRAG condition in terms of social presence and user experience for remote participants.
- H3: There is no significant difference in the workload between the *BeHere* condition and the VRAG condition.
- H4: The *BeHere* condition would benefit remote communication compared with the VRAG condition.

#### 5.1.3 Participants

We recruited 30 participants in 15 pairs from our university, aged 18–28 years old (23 males and 7 females, mean 22.43, SD 2.50) by within-subject design. The participants were not the same as before. Their major backgrounds are in various areas such as Intelligent Manufacturing, Mechanical Engineering, and Mechanical design and manufacturing. The participants' background information is shown in Fig. 11. Most users were novices in VR/AR and remote



**Fig. 11** Participants' background

collaboration, which will alleviate the impact of the participants' different backgrounds on the results to some extent.

### 5.1.4 Procedure

The experiment procedure mainly has seven steps as illustrated in Fig. 12. Each trial took almost 40 min, and there is no pause during the experiment. We used a within-subject design where the local and remote participants did not swap for each condition in the user study. We used this design to reduce the experimental time, users' feeling of tiredness and boring, and the impact of the learning effect to a certain degree. After the training session, we found that users will know how to assemble the vise. Therefore, we did not provide the remote VR users with the instructions in the user study. In steps five and six, we collected the objective and subjective data after each condition. More specifically, the objective data were performance time, and the subjective data mainly include workload measured by NASA Task Load Index Questionnaire (Hart and Staveland 1988) and collaborative experience (COEX). We designed the COEX questionnaire (see Table 2) based on the Networked Minds Social Presence Measure (Harms and Biocca 2004), slightly modified to reflect our research objectives.

After finishing both two conditions, we asked all participants to rank the two interfaces based on their user experience according to Table 3 with respect to six items (RC1, interesting; RC2, understanding; RC3, co-presence; RC4, enjoyment; RC5, focus; and RC6, satisfaction).

The remote collaborative scenario for the procedural task with the vise is shown in Fig. 13. Using the VRAG and BeHere interfaces, the remote VR participants can provide the local SAR users with instructions based on gestures and virtual replicas cues, while knowing the on-site situation through the live video, as shown in Fig. 13a, b. They can see the partners' 3D avatar in the VR environment and manipulate the 3D virtual replicas using gestures in natural and intuitive interaction. On the local site, the on-site users can

finish the procedural task following the instructions provided by the shared gestures and virtual replicas (see Fig. 13c).

## 5.2 Results

In this section, we reported the experimental results. They include both objective measures such as performance time and subjective measures such as social presence, workload, and ranking and users' preference.

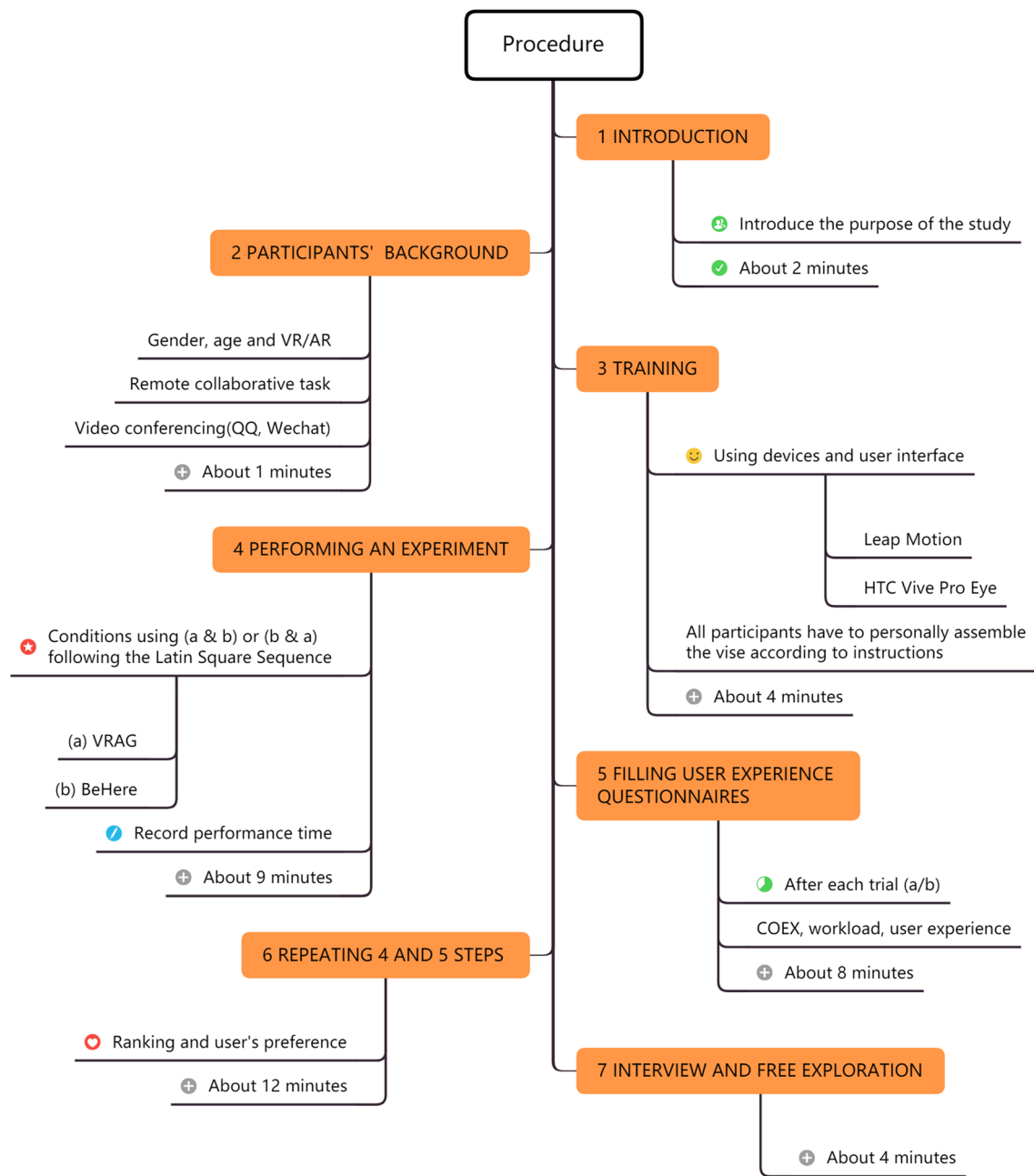
### 5.2.1 Performance

The result of statistical analysis showed that there was no significant difference in the average time needed to complete the procedure task. Before analyzing the collected data, we checked the data of consistency and normality validation using the Shapiro–Wilk test. The result indicated that the data in the VRAG and BeHere condition are no deviation. Then, we conducted a paired t test ( $\alpha=0.05$ ), on the performance time results for each condition, the finding was no significant difference [ $t(14)=0.092$ ,  $p=0.928$ ] between the VRAG interface (mean 129.13 s, SD 8.99) and the BeHere interface (mean 128.93 s, SD 11.38).

### 5.2.2 Social presence

We evaluated user experience by using the COEX questionnaire to study if the different visual cues affected the users' presence and attention. In the current research, we selected four subscales according to our study: Co-presence (CP), Attention Allocation (AA), Perceived Message Understanding (PMU), and Perceived Behavioral Interdependence (PBI) (Harms and Biocca 2004), as illustrated in Table 2. The results of COEX were reported by all participants as shown in Figs. 14 and 15. We conducted a Wilcoxon Signed-Rank test ( $\alpha=0.05$ ) to investigate whether there were significant differences between the BeHere interface and the VRAG interface.

Figure 14 illustrates the remote VR users' COEX, and there were significant differences with respect to CP ( $Z=-2.121$ ,  $p=0.034$ ), AA ( $Z=-2.226$ ,  $p=0.026$ ), PUM ( $Z=-2.425$ ,  $p=0.015$ ), and PBI ( $Z=-2.668$ ,  $p=0.008$ ). Figure 15 illustrates the local SAR users' COEX. Testing the collected data with the Wilcoxon Signed Rank, we found that there were significant differences with respect to CP ( $Z=-2.71$ ,  $p=0.023$ ) and PBI ( $Z=-2.714$ ,  $p=0.007$ ), except for AA ( $Z=-1.300$ ,  $p=0.194$ ) and PUM ( $Z=-0.649$ ,  $p=0.516$ ). Moreover, as shown in Figs. 14 and 15, we found a generally high rating of both conditions on four subscales (mean  $>5.1$  and  $\geq 5$  out of 7 for remote and local users, respectively). This shows that the users in general had a social presence experience with the shared



**Fig. 12** Experiment procedure

avatar- and gesture-based cues, especially, for the remote VR users using the BeHere interface.

From the interview, some participants in the remote VR site said: “It is very interesting to see my partner’ 3D avatar; The BeHere interface provides me with a stronger feeling of co-presence; I find that the virtual objects on the table will occasionally occlude the rendered live video; The interaction is natural and intuitive, especially, the BeHere interface

makes me feel like that I’m working together with my partner in the same place; It will be wonderful if the BeHere interface can present my partner’s facial expression. The system will be perfect if it can support haptic feedback; I feel like my partner next to me as if we were co-located.” For local SAR participants, the BeHere interface can provide a significantly stronger feeling of CP and PBI compared to the VRAG interface, as illustrated in Fig. 15.

**Table 2** COEX questionnaire: 1 is strongly disagree and 7 is strongly agree, the higher the better

| C#  | Questions (7-point scale rating)                                  | Item                                       |
|-----|-------------------------------------------------------------------|--------------------------------------------|
| C1  | I noticed my partner                                              | Co-presence (CP)                           |
| C2  | My Partner's presence was obvious to me                           |                                            |
| C3  | I caught my partner's attention                                   |                                            |
| C4  | I remained focused on my partner throughout our interaction       | Attentional allocation                     |
| C5  | My partner did not receive my full attention                      |                                            |
| C6  | My partner's thoughts were clear to me                            | Perceived message understanding (PMU)      |
| C7  | It was easy to understand my partner                              |                                            |
| C8  | My partner had difficulty understanding me                        | Perceived behavioral interdependence (PBI) |
| C9  | My behavior was often in direct response to my partner's behavior |                                            |
| C10 | I reciprocated my partner's actions                               |                                            |
| C11 | My partner's behavior was closely tied to my behavior             |                                            |

**Table 3** Ranking criteria: 1 is best and 2 is worst, the lower the better

| RC# | Which condition was best...                                |
|-----|------------------------------------------------------------|
| RC1 | At helping you feel more <i>interesting</i> ?              |
| RC2 | At making you feel easy to <i>understand</i> your partner? |
| RC3 | At providing the <i>co-presence</i> sense                  |
| RC4 | At helping you to <i>enjoy</i> the procedural task?        |
| RC5 | At making you <i>focus</i> on the task?                    |
| RC6 | At making you <i>satisfied</i> with the performance?       |

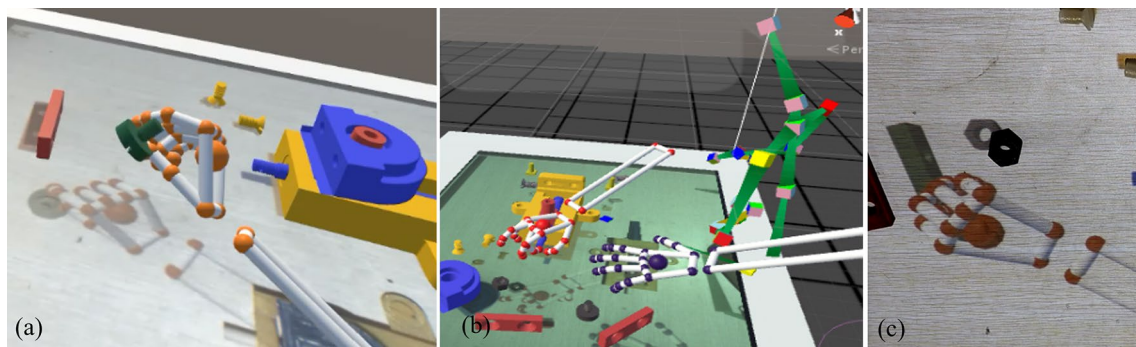
### 5.2.3 Workload

We compared the participants' physical and mental workload by using the NASA-TLX survey, similar to many prior related researches (Bai et al. 2020; Wang et al. 2020a; Yang et al. 2020). The value is lower, and the result is better. Figure 16 shows the average workload assessment between two conditions for TLX, and we focused on items for the remote VR participants because the local SAR participants have the same conditions in the remote collaborative procedural task. Almost generally, the remote participants using the BeHere

interface had a lower workload rating than using the VRAG interface in Temporal Demand, Mental Demand, Performance, and Frustration Level, except for Physical Demand. A Wilcoxon Signed-Rank test ( $\alpha=0.05$ ) showed there was no significant difference in the remote VR participants' workload ( $p>0.2$ ).

### 5.2.4 Ranking and preference

Figure 17 shows the average ranking results (1 = the best, 2 = the worst). For local and remote users, in most cases, the BeHere condition was ranked better than the VRAG condition. To investigate if users ranked two interfaces significantly different, we conducted chi-square test ( $\alpha=0.05$ ). For remote VR participants, there was a significant difference in terms of RC1(interesting,  $\chi^2(1)=10.44$ ,  $p=0.001$ ), RC3 (co-presence,  $\chi^2(1)=15.60$ ,  $p<0.001$ ), RC4 (enjoyment,  $\chi^2(1)=10.44$ ,  $p=0.001$ ) except item RC2 (understanding,  $\chi^2(1)=3.22$ ,  $p=0.073$ ), RC5 (focus,  $\chi^2(1)=0.129$ ,  $p=0.720$ ), and RC6 (satisfaction,  $\chi^2(1)=1.16$ ,  $p=0.281$ ). For local SAR participants, there was no significant difference in all items ( $p>0.075$ ) between the two conditions.



**Fig. 13** VR-SAR remote collaborative scenario. **a** The HTC Vive HMD view on the remote site using the VRAG interface. **b** The HTC Vive HMD view on the remote site using the BeHere interface. **c** The on-site worker scene

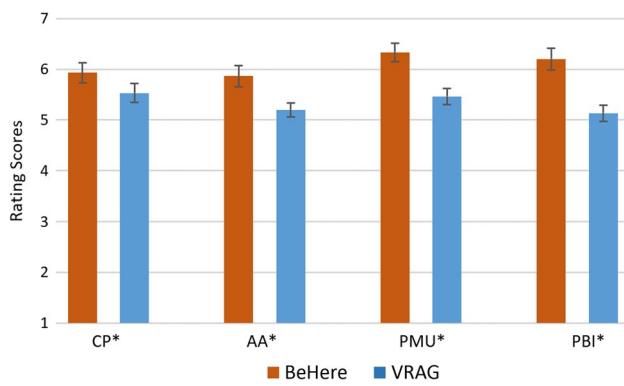


Fig. 14 Remote user' COEX (\*statistically significant)

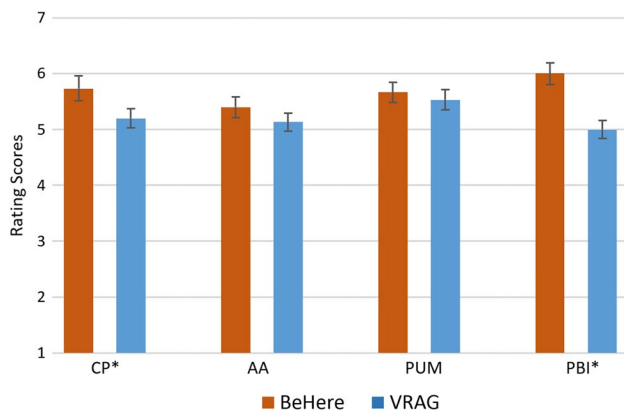


Fig. 15 Local user' COEX (\*statistically significant)

For the participants' preferences, all users picked out the interface they most preferred. Almost 73% (13 remote + 9 local users) of them chose the BeHere condition, 25% (2 remote + 6 local users) of them chose the VRAG condition

as shown in Fig. 18. More specifically, almost generally, most participants liked the BeHere condition. From the interview, some local participants said: "My partner using the BeHere interface has a more positive effect on me than the VRAG interface; The instructions based on gesture and virtual replicas are very useful and intuitive; I think the system is acceptable for this simple procedural task but it maybe has some problems for some tasks in the real-world such as the 3D operations and occlusion; I hope to see my partner's avatar as the remote site; It will be more interesting if the system can provide instructions using HoloLens; The shared gestures could improve my sense of co-presence."

## 6 Further exploration

Based on Sect. 5 and the participants' feedback, the results showed our research is engaging. Therefore, we further explored the 3D shared VR environment comparing the conditions with the local user avatar and without it.

### 6.1 Experimental details

#### 6.1.1 Conditions

In this user study, the main independent variable was the local participants' avatar based on the 3D shared virtual environment as shown in Fig. 19. Therefore, the user study has two communication conditions as follows:

- (1) *3DVRAG*: the prototype system providing the local SAR user with instructions based on the combination of gestures and virtual replicas in the 3D shared virtual environment.

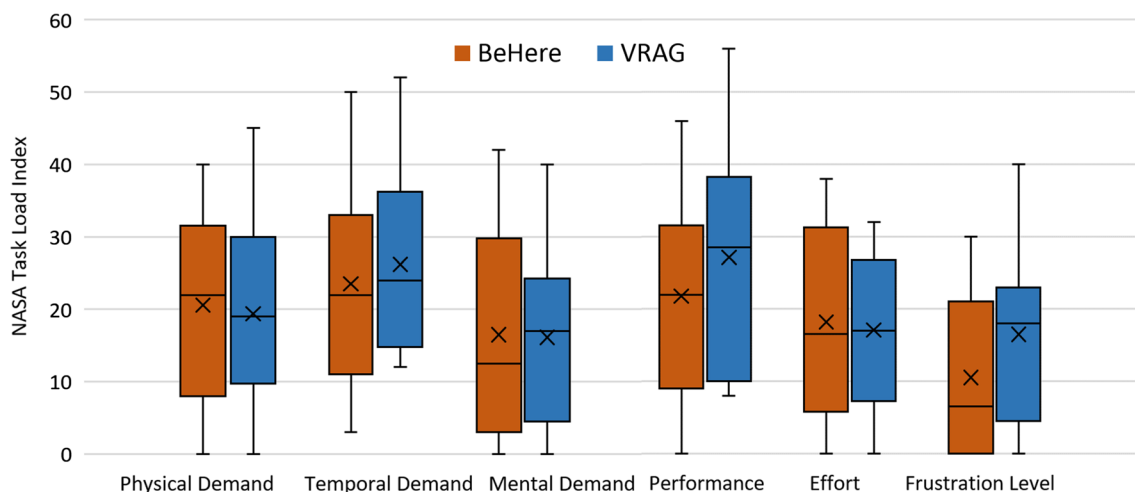
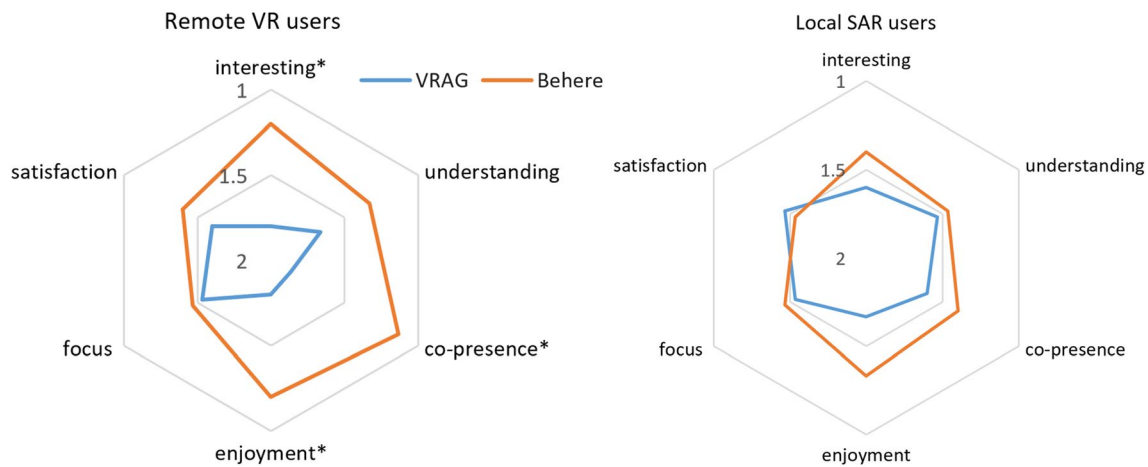
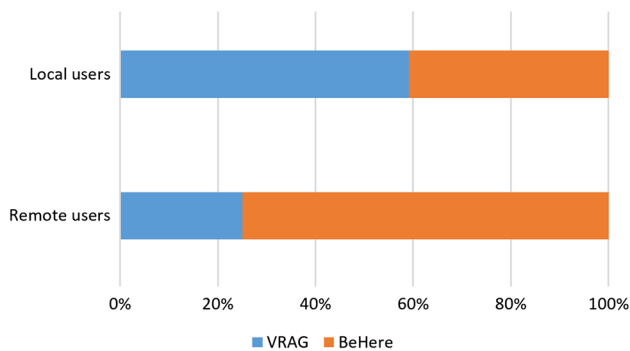


Fig. 16 NASA RTLX of perceived workload for remote VR users under the two conditions



**Fig. 17** Overall average ranking results (\*statistically significant)



**Fig. 18** Users' preference and most preferred the BeHere interface

- (2) *3DBeHere*: the prototype system supports sharing the local user's 3D avatar in the 3D shared virtual environment based on the 3DVRAg condition.

In both conditions, we also shared the on-site worker scene through the live video as shown in Fig. 19d, e. For the experimental task, we chose a typical procedural task as illustrated in Fig. 19, assembling a vise. We added two constraints for encouraging collaboration as the first user study.

### 6.1.2 Task and hypotheses

In general, the three visual cues (e.g., gestures, avatar and virtual replicas) have different merits in providing a sense of co-presence and clear instructions. Thus, our research has the following three hypotheses:

- (1) FEH1: the 3DBeHere condition would improve task performance compared with the 3DVRAg condition.

- (2) FEH2: the 3DBeHere condition is better than the 3DVRAg condition in terms of social presence and user experience for remote participants.
- (3) FEH3: there is no significant difference in the workload between the BeHere condition and the 3DVRAg condition.

### 6.1.3 Participants

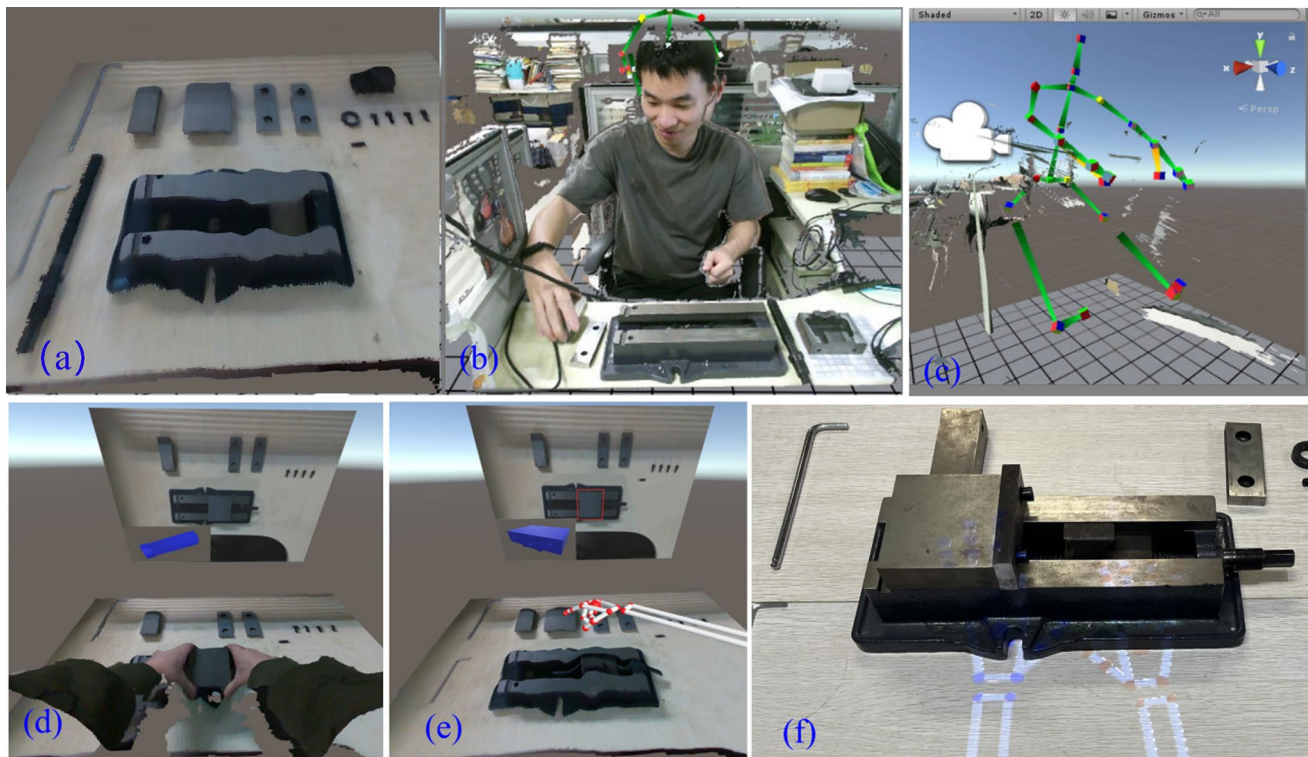
We recruited 22 participants in 10 pairs from our university, aged 18–31 years old (17 males and 5 females, mean 24.77, SD 2.84) by within-subject design. Their major backgrounds are in various areas such as Robot engineering, Mechanical Engineering, and Mechanical design and manufacturing. Most participants were novices in VR/AR-based remote collaboration, which will alleviate the impact of the participants' different backgrounds on the results to some extent.

### 6.1.4 Procedure

The experiment procedure mainly has seven steps as illustrated in Fig. 20. Each trial took almost 51 min. We used a within-subject design where the local and remote participants did not swap for each condition in the user study. We collected the objective and subjective data after each condition as the user study in Sect. 5.

## 6.2 Results

In the following section, we reported the experimental results. They include both objective measures such as performance time and subjective measures such as social presence, workload, and ranking and users' preference.



**Fig. 19** VR-SAR remote collaborative scenario based on 3D shared VR environment. **a–c** The HTC Vive view sharing avatar' cues, **a**, **b** 3D virtual settings of local site and a local participants' avatar, **c** the

local participants' avatar in a new perspective, **d** the HTC Vive view without sharing avatar' cues, **e** instructions based on gesture cues in the HTC Vive view, **f** the on-site worker scene

### 6.2.1 Performance

We conducted a paired  $t$  test ( $\alpha=0.05$ ), on the performance time results for each condition, the finding was no significant difference [ $t(10)=0.874$ ,  $p=0.403$ ] between the 3DVrag interface (mean 421.27 s, SD 31.74) and the 3DBeHere interface (mean 413.55 s, SD 34.07) as shown in Fig. 21.

### 6.2.2 Social presence

We evaluated user experience by using the COEX questionnaire to study if the different visual cues affected the users' presence and attention as user study in Sect. 2.2. The results of COEX were reported by all participants as shown in Figs. 22 and 23. We conducted a Wilcoxon Signed-Rank test ( $\alpha=0.05$ ) to investigate whether there were significant differences between the 3DBeHere interface and the 3DVrag interface.

Figure 22 illustrates the remote VR users' COEX, and there were significant differences with respect to CP ( $Z=-2.070$ ,  $p=0.038$ ) and PBI ( $Z=-2.280$ ,  $p=0.023$ ) except for AA and PUM ( $p>0.3$ ). Figure 23 illustrates the local SAR users' COEX. Testing the collected data with the Wilcoxon Signed Rank, we found that there were no significant differences ( $p>0.09$ ). Moreover, as shown in

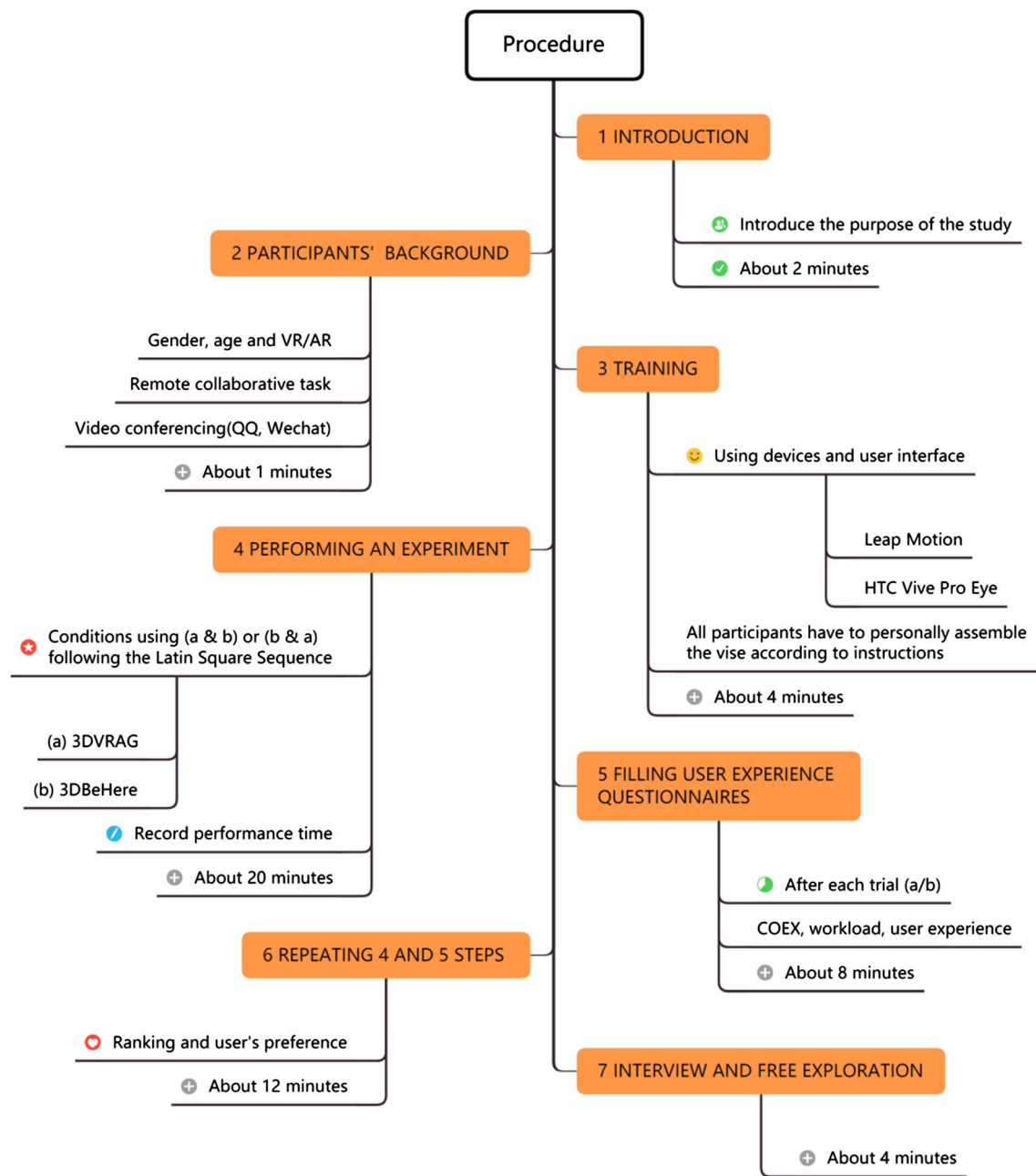
Figs. 22 and 23, we found a generally high rating of both conditions on four subscales. This shows that the users, in general, had a social presence experience with the shared avatar- and gesture-based cues, especially, for the remote VR users using the 3DBeHere interface.

### 6.2.3 Workload

Figure 24 shows the average workload assessment between two conditions for TLX, and we focused on items for the remote VR participants because the local SAR participants have the same conditions in the remote collaborative procedural task. A Wilcoxon Signed-Rank test ( $\alpha=0.05$ ) showed there was no significant difference on the remote VR participants' workload ( $p>0.35$ ) as shown in Table 4.

### 6.2.4 Ranking and preference

Figure 25 shows the average ranking results (1 = the best, 2 = the worst). For local and remote users, in most cases, the BeHere condition was ranked better than the Vrag condition. To investigate if users ranked two interfaces significantly different, we conducted a chi-square test ( $\alpha=0.05$ ). For remote VR participants, there was a significant difference in terms of RC1 (interesting,  $\chi^2(1)=10.46$ ,  $p<0.001$ ),



**Fig. 20** Experiment procedure

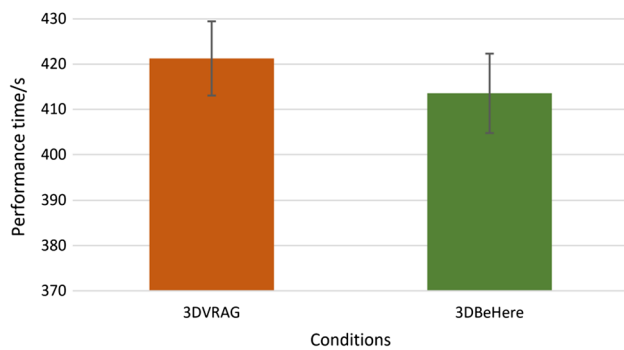
RC3 (co-presence,  $\chi^2(1) = 8.504, p = 0.004$ ) except item RC2 (understanding), RC4 (enjoyment), RC5 (focus), and RC6 (satisfaction) ( $p > 0.2$ ). For local SAR participants, there was no significant difference in all items ( $p > 0.2$ ) between the two conditions.

For the participants' preferences, all users picked out the interface they most preferred. Almost 63% (8 remote + 6 local users) of them chose the 3DBeHere condition, 24% (5 remote + 5 local users) of them chose the 3DVRAG

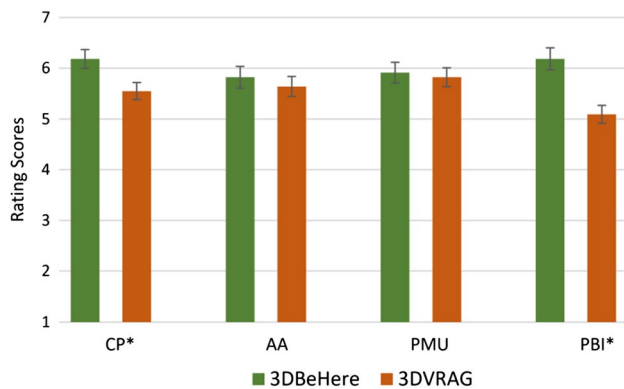
condition. More specifically, almost generally, most participants liked the 3DBeHere condition.

## 7 Discussion

In this section, we discussed the study results, some observations, and the possible reasons for the experimental results at length. The formal user study showed how the visual cues



**Fig. 21** Performance time

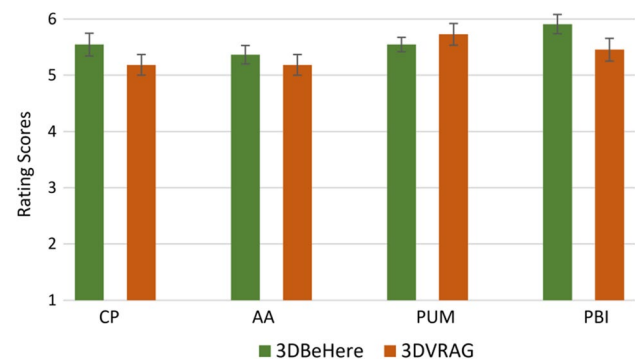


**Fig. 22** Remote user's COEX (\*statistically significant)

(e.g., virtual replicas, gestures, and Avatar) affect VR/SAR remote collaboration on a procedural task. In the study, we listed four and three hypotheses for VRAG & BeHere and 3DVRaG & 3DBeHere, respectively, in reference to the research key points. The hypotheses are verified based on the research results in performance, social presence, workload, ranking and preference.

For performance time, although the average time using the BeHere interface is faster than the VRAG interface to complete the procedural task, there was no significant difference between the two conditions. Therefore, hypothesis H1 was rejected. We think that there were two reasons for this. On the one hand, the procedural task is relatively simple and the number of parts is small. On the other hand, although the remote scenario is different and this will influence the local site, the collaborative scenario is almost the same for the local participants under both conditions. For 3D environment conditions, although the average time using the 3DBeHere interface is faster than the 3DVRAG interface to complete the procedural task, there was no significant difference between the two conditions. Therefore, hypothesis FEH1 was rejected.

For social presence, we evaluated it using the COEX questionnaire as shown in Table 2. In most cases, there

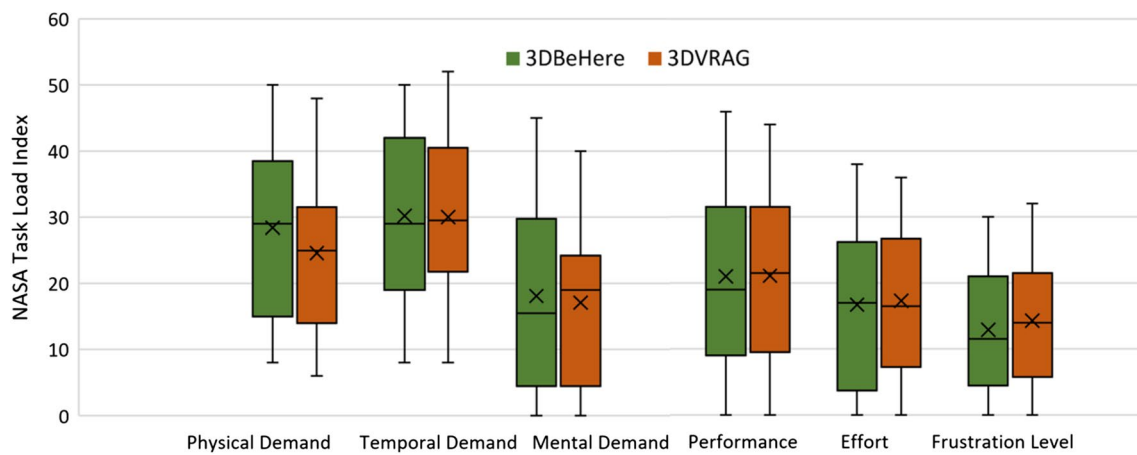


**Fig. 23** Local user's COEX (\*statistically significant)

were significant differences between local and remote participants. Specifically, for remote VR participants, the combination of sharing remote VR users' gestures and local SAR users' avatar was able to provide a significantly stronger feeling of CP, AA, PUM, and PBI compared to the VRAG interface, as illustrated in Fig. 14. Thus, the hypothesis H2 is accepted. More importantly, from the interview, some local participants said: "It is easy to provide instructions for my partners using gesture and virtual replicas, but I hope the prototype system can support making annotations as shown in Fig. 26." For 3D environment conditions, we evaluated it using the COEX questionnaire. As a result, there were significant differences between remote participants. Specifically, for remote VR participants, the 3DBeHere interface was able to provide a significantly stronger feeling of CP and PBI compared to the 3DVRAG interface. Therefore, the hypothesis FEH2 is accepted.

About the workload, we focused on the remote VR site. There was no significant difference in the workload assessment using the NASA-TLX survey as shown in Fig. 16. Therefore, the hypothesis H3 is accepted. A reasonable explanation is that although the remote VR participants could see their partners' avatar, this provides more visual cues creating the sense of co-presence and the feeling of being together with collaborators. In addition, according to the performance time and social presence, we can draw the conclusion that the hypothesis H4 was rejected. For 3D environment conditions, we focused on the remote VR site. There was no significant difference in the workload assessment using the NASA-TLX survey. Therefore, the hypothesis FEH3 is accepted.

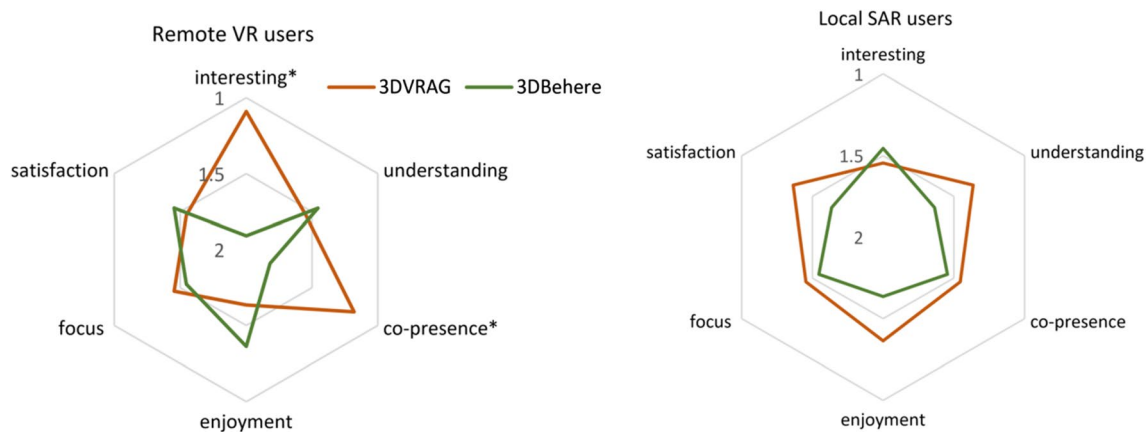
Compared with the VRAG/3DVRAG condition, the BeHere/3DBeHere interface established a sound foundation of co-presence as shown in Figs. 14, 15 and 17. From the results of the ranking and users' preferences, we can conclude that sharing the local users' avatar has a positive effect on improving the sense of co-presence for remote VR users. Moreover, most participants preferred the BeHere/3DBeHere interface, which supports sharing instructions based on



**Fig. 24** NASA RTLX of perceived workload for remote VR users under the two conditions

**Table 4** Workload assessment

|          | Physical demand | Temporal demand | Mental demand | Performance | Effort  | Frustration level |
|----------|-----------------|-----------------|---------------|-------------|---------|-------------------|
| <i>Z</i> | − 0.918         | − 0.421         | − 0.357       | − 0.070     | − 0.178 | − 0.534           |
| <i>p</i> | 0.359           | 0.674           | 0.721         | 0.944       | 0.859   | 0.594             |



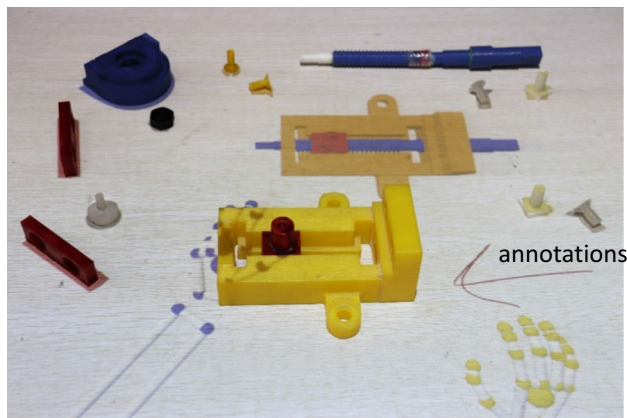
**Fig. 25** Overall average ranking results (\*statistically significant)

gestures and virtual replicas from the remote VR site to the local site and presenting local users' 3D avatar in the VR environment.

## 8 Limitations and future works

Based on the participants' feedback, although our proposed VR/SAR remote collaborative system generally works well for a procedural task in the controlled environment, there are also some disadvantages to our research. First, in the current research, we just recruited student volunteers and the procedural task is relatively simple. We think that the

study will obtain more convincing results by using more complex and typical real-world tasks with more participants from different fields such as workers, designers, and engineers. Remote collaborative tasks must be difficult and long enough to encourage interaction between collaborators and for the AR-based solution being used to provide enough contribution (Marques et al. 2022). In general, tasks can benefit from deliberate drawbacks and conditions, i.e., incorrect, vague, or missing information, to force more complicated tasks and elicit remote collaboration (Marques et al. 2022). For example, suggest using an object which does not exist in the environment of the other collaborator or removing a red cable which is green in the other collaborator's context.



**Fig. 26** Making annotations by gesture

Such situations help introduce different levels of complexity, which go beyond the standard approaches used, and elicit more realistic real-life situations. Now, we are trying to use building sets using a pump assembly as a task to improve the prototype. There could be other foreseeable further applications in the future.

Second, in the research, we only shared the local users' avatar and used a skeleton to represent local collaborators. These solutions might impact the system usability to some degree. Therefore, the system could be improved by applying a more human-like 3D avatar and providing mutual sharing to augment the embodiment and remote collaboration in further investigation. More importantly, we are interested in enabling the system to support AR annotations by gestures as shown in Fig. 26.

Third, Wang et al. (2020b) presented a VR-SAR remote collaborative system supporting passive haptic feedback, which could enhance the VR user experience in terms of passion for collaboration, controllability of sketching and pointing, satisfaction and enjoyment. Thus, in the near future, we will improve the prototype system by supporting passive haptic feedback on the remote VR site, and explore how it will impact remote collaborative work.

Besides, the user study did not capture as much data as it could have. For example, in the future, it would be good to record people's conversations and perform conversational analysis. We could also have used other subjective surveys, and a number of other measures such as errors.

Finally, the system provides remote users with real-time information by the live video, which will lose some depth of information that is very important for some tasks (Teo et al. 2020). Additionally, our prototype system did not support the real-time update of the 3D workspace scene without any delay. More research focuses on supporting 3D capture, gaze cues, and spatial auditory to improve remote collaboration (Bai et al. 2020; Yang et al. 2020; Wang et al. 2021a). Consequently, we would like to reconstruct and share the

3D local scene in real-time by using a Kinect and to support the multimodal interaction based on gesture and gaze cues in the future. Moreover, we want to improve the prototype system using point-cloud avatars, as described in Gamelin et al. (2021).

## 9 Conclusion

In the paper, we proposed a new VR/SAR remote collaborative system, BeHere, which supports sharing of gestures, avatar cues, and 3D virtual replicas for remote assistance on a real-world procedural task. Our research can be divided into two parts: a pilot test and a formal user study. First, we performed a pilot test to evaluate and improve the prototype system based on participants' comments. The results showed that users felt that the system usability was good. Then, we conducted a full user study that investigates the effects of sharing gestures and avatar cues based on 3D virtual replicas. In addition, we presented the system architecture and key technical challenges and reported the results and discussion from a pilot test and a formal user study with six prototype systems (e.g., VIRE, VRA, VRAG, BeHere, 3DVIRAG, and 3DBeHere). The VRAG/BeHere and 3DVIRAG/3DBeHere system extends the earlier VRA system by taking into account participants' feedback from the pilot test, and the VIRE system is the existing method usually used in the study.

In our study, we combined SAR and VR to share users' gestures and avatar in remote collaboration on a procedural task. Recently, Hietanen et al. (2020) proved that using SAR has lots of advantages (e.g., safety, competence, and ergonomics) compared with wearable AR. According to this, on the local site, we used a projector-based AR which did not need users to wear any technology or operate any devices during the task. Additionally, we concentrated on a vise assembly in industry as an illustrative case in a pilot test and a user study. To the best of our knowledge, our work is one of the first remote collaborative systems providing gestures and virtual replicas information in VR from the remote site to the local site and meanwhile sharing avatar cues in the on-site settings from the local site to the remote VR site. In the user study, we explored four hypotheses in terms of performance (H1, FEH1), social presence (H2, FEH2), workload (H3, FEH3), and communication (H4). The results showed that there were significant differences in performance and communication, except for social presence, and workload. The study revealed that sharing gestures, avatar, and virtual replicas have an active role in remote collaboration, especially, for remote VR participants with respect to the social presence (e.g., interesting, co-presence, and enjoyment). Furthermore, most participants liked the BeHere/3BeHere interface compared with the VRAG/3DVIRAG interface

according to the results of users' ranking and preference. These visual cues played a key role in providing situational awareness and facilitating conversational grounding in VR/SAR remote collaboration on a procedural task. More importantly, our work provides a foundation way for future studies in the combination of gesture-based interaction using virtual replicas supporting haptic feedback in remote collaboration.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10055-023-00748-5>.

**Acknowledgements** We would like to appreciate the anonymous reviewers for their constructive suggestions for enhancing this paper. Particularly, we would like to appreciate Mark Billingham for checking the English of an early version. We would also like to thank members of Chongqing University of Posts and Telecommunications and Northwestern Polytechnical University for their participation in the experiment.

**Author contributions** Peng Wang initiated the manuscript and idea for the article. Yue Wang and Billingham Mark improved the experiment and paper, particularly, Mark Billingham checked the English of an early version, and his constructive comments are gratefully acknowledged which have helped the author to improve the paper. Moreover, HuiZhen Yang and Peng Xu developed the VR-SAR prototype system. Yanhong Li collected and analyzed the experimental data.

**Funding** This research was sponsored by the national key research and development program (2020YFB1710300), Aeronautical science fund of China (2019ZE105001), and General Project of Chongqing Natural Science Foundation (cstc2019jcyj-msxmX0530, cstb2022nscq-msx1153), the special project of the intelligent factory and network collaborative manufacturing of the Ministry of Science and Technology: the development and application demonstration of ship manufacturing process control platform (2020YFB1712600), and the research on key technologies for 5G applications of ship assembly and construction sub-workshop of the Ministry of Industry and Information Technology (CJ04N20).

**Data availability** Not applicable.

**Code availability** Not applicable.

## References

- Anton D, Kurillo G, Bajcsy R (2018) User experience and interaction performance in 2D/3D telecollaboration. *Fut Gener Comput Syst* 82:77–88
- Bai H, Sasikumar P, Yang J, Billingham M (2020) A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In: *Proceedings of conference on human factors in computing systems*, pp 1–13
- Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: adding an adjective rating scale. *J Usab Stud* 4:117–123
- Barroso J, Fonseca L, Marques B, et al (2020) Remote collaboration using mixed reality: exploring a shared model approach through different interaction methods. In: *European virtual conference on computer-supported cooperative work*, pp 1–4
- Billingham M (2021) Grand challenges for augmented reality. *Front Virt Real*. <https://doi.org/10.3389/frvir.2021.578080>
- Bottani E, Vignali G (2019) Augmented reality technology in the manufacturing industry: a review of the last decade. *IISE Trans* 51(3):284–310. <https://doi.org/10.1080/24725854.2018.1493244>
- Brooke J (1996) SUS—a quick and dirty usability scale. *Usab Eval Ind* 189(194):4–7
- Calandra D, Cannavò A, Lamberti F (2021) Improving AR-powered remote assistance: a new approach aimed to foster operator's autonomy and optimize the use of skilled resources. *Int J Adv Manuf Technol* 114:3147. <https://doi.org/10.1007/s00170-021-06871-4>
- Chen L, Liu Y, Li Y, Yu L, Gao B, Caon M, Yue Y, Liang H (2021) Effect of visual cues on pointing tasks in co-located augmented reality collaboration. In: *Proceedings—SUI 2021: ACM symposium on spatial user interaction*, November 09–10, 2021. ACM, New York <https://doi.org/10.1145/3485279.3485297>
- De Belen RAJ, Nguyen H, Filonik D, Favero DD, Bednarz T (2019) A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. *AIMS Electron Electr Eng* 3(2):181–223
- De Pace F, Manuri F, Sanna A, Zappia D (2019) A comparison between two different approaches for a collaborative mixed-virtual environment in industrial maintenance. *Front Robot AI* 6:18
- Elvezio C, Feiner S, Tversky B (2015) [POSTER ] Interactive visualizations for monoscopic eyewear to assist in manually orienting objects in 3D. *IEEE Int Symp Mixed Augment Real* 2015:180–181. <https://doi.org/10.1109/TSMAR.2015.54>
- Elvezio C, Sukan M, Oda O, et al (2017) Remote collaboration in AR and VR using virtual replicas. In: *Association for computing machinery (ACM)*, pp 1–2
- Ens B, Lanir J, Tang A et al (2019) Revisiting collaboration through mixed reality: the evolution of groupware. *J Hum Comput Stud* 131:81–98. <https://doi.org/10.1016/j.ijhcs.2019.05.011>
- Gamelin G, Chellali A, Cheikh S et al (2021) Point-cloud avatars to improve spatial communication in immersive collaborative virtual environments. *Pers Ubiquit Comput* 25:467–484. <https://doi.org/10.1007/s00779-020-01431-1>
- Garbett J, Hartley T, Heesom D (2021) A multi-user collaborative BIM-AR system to support design and construction. *Autom Constr* 122(103487):1–11. <https://doi.org/10.1016/j.autcon.2020.103487>
- Harms C, Biocca F (2004) Internal consistency and reliability of the networked minds measure of social presence. In: *7th annual international workshop: presence*
- Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv Psychol* 52(6):139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hietanen A, Pieters R, Lanz M et al (2020) AR-based interaction for human-robot collaborative manufacturing. *Robot Comput-Integr Manuf* 63:101891. <https://doi.org/10.1016/j.rcim.2019.101891>
- Higuch K, Yonetani R, Sato YBT-CHIC on HF in CS (2016) Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks, pp 5180–5190
- Huang R, Zhang S, Bai X et al (2015) An effective subpart retrieval approach of 3D CAD models for manufacturing process reuse. *Comput Ind* 67:38–53
- Huang W, Alem L, Tecchia F, Duh BL (2018) Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration. *J Multimodal User Interfaces* 12:77–89
- Jasche F, Kirchhübel J, Ludwig T, Tolmie P (2021) BeamLite: diminishing ecological fractures of remote collaboration through mixed reality environments. In: *ACM international conference proceeding series*, June 20–25, 2021, Seattle, WA, USA
- Kim S, Lee G, Huang W, et al (2019) Evaluating the combination of visual communication cues for HMD-based mixed reality remote

- collaboration. In: Proceedings of the 2019 CHI conference on human factors in computing systems. ACM, p 173
- Kim S, Jing A, Park H et al (2020) Hand-in-air (HiA) and hand-on-target (HoT) style gesture cues for mixed reality collaboration. *IEEE Access* 8(99):1–1. <https://doi.org/10.1109/ACCESS.2020.3043783>
- Kritzler M, Murr M, Michahelles FBT-IC on the I of T (2016) Remote-Bob: support of on-site workers via a telepresence remote expert system. In: International conference on the internet of things, pp 7–14
- Kurillo G, Bajcsy R (2013) 3D teleimmersion for collaboration and interaction of geographically distributed users. *Virt Real* 17:29–43. <https://doi.org/10.1007/s10055-012-0217-2>
- Marques B, Silva S, Alves J, Araujo T, Dias P, Santos BS (2021) A conceptual model and taxonomy for collaborative augmented reality. *IEEE Trans vis Comput Graph* 14(8):1–18. <https://doi.org/10.1109/TVCG.2021.3101545>
- Marques B, Silva S, Teixeira A et al (2022) A vision for contextualized evaluation of remote collaboration supported by AR. *Comput Graph (pergamon)* 102(2022):413–425. <https://doi.org/10.1016/j.cag.2021.10.009>
- Oda O, Elvezio C, Sukan M, et al (2015) Virtual replicas for remote assistance in virtual and augmented reality. In: UIST 2015—proceedings of the 28th annual ACM symposium on user interface software and technology, pp 405–415
- Orts-Escalano S, Kim D, Cai Q, et al (2016) Holoportation: virtual 3D teleportation in real-time. In: UIST 2016—proceedings of the 29th annual symposium on user interface software and technology, pp 741–754
- Piumsomboon T, Lee GA, Billinghamurst M (2018a) Snow dome: a multi-scale interaction in mixed reality remote collaboration. In: Conference on human factors in computing systems, pp 1–4
- Piumsomboon T, Lee GA, Ens B et al (2018b) Superman vs giant: a study on spatial perception for a multi-scale mixed reality flying telepresence interface. *IEEE Trans vis Comput Graph* 24:2974–2982
- Piumsomboon T, Dey A, Ens B et al (2019) The effects of sharing awareness cues in collaborative mixed reality. *Front Robot AI* 6(5):1–18. <https://doi.org/10.3389/frobt.2019.00005>
- Pringle A, Campbell AG, Hutka S et al (2018) [Poster] Using an industry-ready AR HMD on a real maintenance task: AR benefits performance on certain task steps more than others. *IEEE Int Symp Mixed Augment Real Adjunct (ISMAR)* 2018:236–241. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00075>
- Russo M (2021) AR in the architecture domain: state of the art. *Appl Sci (switzerland)* 11:1–38
- Sukan M, Elvezio C, Feiner S, Tversky B (2016) Providing assistance for orienting 3D objects using monocular eyewear. In: Proceedings of the 2016 symposium on spatial user interaction, pp 89–98
- Teo T, Lawrence L, Lee GA, et al (2019) Mixed reality remote collaboration combining 360 video and 3D reconstruction. In: Proceedings of the 2019 CHI conference on human factors in computing systems. ACM, p 201
- Teo T, Norman M, Lee GA, Billinghamurst M, Adcock M (2020) Exploring interaction techniques for 360 panoramas inside a 3D reconstructed scene for mixed reality remote collaboration. *J Multimodal User Interfaces* 14:373–385. <https://doi.org/10.1007/s12193-020-00343-x>
- Wang P, Zhang S, Bai X et al (2019a) 2.5DHANDS: a gesture-based MR remote collaborative platform. *Int J Adv Manuf Technol* 102:1–15. <https://doi.org/10.1007/s00170-018-03237-1>
- Wang P, Zhang S, Bai X, et al (2019b) Head pointer or eye gaze: which helps more in MR remote collaboration. In: The 26th IEEE conference on virtual reality and 3D user interfaces, pp 1–2
- Wang P, Zhang S, Bai X et al (2019c) A gesture- and head-based multimodal interaction platform for MR remote collaboration. *Int J Adv Manuf Technol* 105:3031–3043. <https://doi.org/10.1007/s00170-019-04434-2>
- Wang P, Zhang S, Billinghamurst M, et al (2019d) A comprehensive survey of AR/MR-based co-design in manufacturing. *Eng Comput* 36:1715–1738
- Wang P, Bai X, Billinghamurst M et al (2020a) Using a head pointer or eye gaze: The effect of gaze on spatial AR remote collaboration for physical tasks. *Interact Comput* 32:1–2. <https://doi.org/10.1093/iwcomp/iwaa012>
- Wang P, Bai X, Billinghamurst M et al (2020b) Haptic feedback helps me? A VR-SAR remote collaborative system with tangible interaction haptic feedback helps me? A VR-SAR remote collaborative system with tangible. *Int J Hum-Comput Interaction* 00:1–16. <https://doi.org/10.1080/10447318.2020.1732140>
- Wang P, Bai X, Billinghamurst M, et al (2021a) AR/MR remote collaboration on physical tasks: a review. *Robot Comput-Integr Manuf*. <https://doi.org/10.1016/j.rcim.2020.102071>
- Wang P, Bai X, Billinghamurst M et al (2021b) 3DGAM: using 3D gesture and CAD models for training on mixed reality remote collaboration. *Multimed Tools Appl* 80:1–26. <https://doi.org/10.1007/s11042-020-09731-7>
- Wang Y, Wang P, Luo Z, Yan Y (2022) A novel AR remote collaborative platform for sharing 2.5D gestures and gaze. *Int J Adv Manuf Technol*. <https://doi.org/10.1007/s00170-022-08747-7>
- Yang J, Sasikumar P, Bai H et al (2020) The effects of spatial auditory and visual cues on mixed reality remote collaboration. *J Multimodal User Interfaces* 14:337–352. <https://doi.org/10.1007/s12193-020-00331-1>
- Zhang X, Bai X, Zhang S et al (2022) Real-time 3D video-based MR remote collaboration using gesture cues and virtual replicas. *Int J Adv Manuf Technol* 121(11):7697–7719

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.