The International Journal of Advanced Manufacturing Technology (2022) 121:7697–7719
https://doi.org/10.1007/s00170-022-09654-7

ORIGINAL ARTICLE

Real-time 3D video-based MR remote collaboration using gesture cues and virtual replicas

Xiangyu Zhang[1,2] · Xiaoliang Bai[1,2] · Shusheng Zhang[1] · Weiping He[1,2] · Peng Wang[1,2] · Zhuo Wang[3] · Yuxiang Yan[1,2] · Quan Yu[1,2]

Received: 19 January 2022 / Accepted: 2 July 2022 / Published online: 2 August 2022

Abstract
With the rapid development of mixed reality (MR) technology, many compact, lightweight, and powerful devices suitable for remote collaboration, such as MR headsets, hand trackers, and 3D cameras, become readily available, providing hardware and software support for remote collaboration. Consequently, exploring MR technologies for remote collaboration on physical industry tasks is becoming increasingly worthwhile. In many complex production scenarios, such as assembly tasks, significant gains can be achieved by having remote experts assist local workers to manipulate objects in local workspaces. However, it can be challenging for a remote expert to carry out effective spatial reference and action demonstration in a local scene. Sharing 3D stereoscopic scenes can provide depth perception and support remote experts to move and explore a local user's environment freely. Previous studies have demonstrated that gesture-based interaction is natural and intuitive, and interaction based on virtual replicas can provide clear guidance, especially for industrial physical tasks. In this study, we develop an MR remote collaboration system that shares the stereoscopic scene of the local workspace by using real-time 3D video. This system combines gesture cues and virtual replicas in a complementary manner to support the remote expert to create augmented reality (AR) guidance for the local worker naturally and intuitively in the virtual reality immersive space. A formal user study was performed to explore the effects of two different modalities interface in industrial assembly tasks: our novel method of using the combination of virtual replicas and gesture cues in the 3D video (VG3DV), and a method similar to the popular method currently of using gesture cues in the 3D video (G3DV). We found that using the VG3DV can significantly improve the performance and user experience of MR remote collaboration in industrial assembly tasks. Finally, some conclusions and future research directions were given.

Keywords Remote collaboration · Mixed reality · Gesture cues · Virtual replicas · MR assembly · 3D video

1 Introduction

In the study, we develop a novel mixed reality (MR) remote collaboration system that combines virtual reality (VR) and augmented reality (AR) to provide remote guidance for industrial assembly tasks. There are many complex production scenarios where a local worker may need a remote expert's assistance to perform physical tasks, such as training, equipment maintenance, and assembly/disassembly [1–5].

Traditional remote guidance methods based on verbal cues or 2D video stream cannot be readily used for operations that require accurate express 3D spatial references and action demonstrations. Spoken language is often ambiguous when describing the locations and operations in 3D spaces [6], resulting in confusion and errors [7]. In addition, 2D video stream has some disadvantages, such as limited viewing perspective and insufficient depth perception. To overcome these limitations and improve efficiency, the remote expert could immerse himself in the 3D stereoscopic scene of the local workspace for 3D spatial referencing, and manipulate the objects naturally to demonstrate action for guiding the local worker.

✉ Xiaoliang Bai
bxl@nwpu.edu.cn

✉ Shusheng Zhang
zssnwpu@163.com

[1] Cyber-Physical Interaction Lab, Northwestern Polytechnical University, Xi'an 710072, China

[2] Cyber-Reality Innovation Centre (China), Nanjing, China

[3] School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China

As the performance of AR and VR head-mounted displays (HMDs) (such as HoloLens[1] and HTC Vive[2]) improves, they can form an immersive space with a wider field of view and support to transfer assistance from a skilled specialist to a novice by using direct and natural interaction. Using VR technology, a 3D stereoscopic scene of the real world can be shared from depth sensors and/or photogrammetry. This technology allows the remote expert to move and explore freely in a local user's scene for 3D spatial referencing [8–13]. Moreover, many previous studies have focused on sharing nonverbal communication cues (e.g., annotation [11], 3D models [10, 12], gesture cues [14, 15], virtual avatar [16], head/eye gaze, and awareness cues [8]) to demonstrate action in the 3D stereoscopic worker scene.

However, there are still some limitations. On the one hand, the quality of the shared 3D reconstruction scene is directly proportional to the bandwidth required for transmission. Thus, it is difficult to share high-quality updates for the shared 3D reconstructed scene in real time [9]. On the other hand, even as existing studies have developed many approaches to enable a remote expert to present non-verbal communication cues on physical objects, it can be challenging or even impossible for a remote expert to manipulate an object from the local user's scene [2]. To overcome the existing limitations of the above research, in this study, we have proposed visualization and interaction techniques for industry physical tasks that require more than simple annotation or gesture cues. Inspired by the research of Ohan et al. [2] and our past work [3], we developed one approach that supports the remote expert to create virtual replicas for their physical counterparts in the 3D stereoscopic scene of the local workspace reconstructed by 3D video stream. Then, the remote expert can use gestures to manipulate the virtual replica to demonstrate to the local worker how to operate the physical counterpart.

As model-based definition (MBD) is widely used in industry, most manufactured parts have corresponding 3D CAD models stored in the repository, which are the high-precision representation of their physical counterparts [17, 18]. Thus, using CAD models as virtual replicas of physical objects does not require extra cost in industrial applications. Besides, many studies have laid the foundation for the use of virtual replicas in MR remote collaboration [2–4, 19, 20]. However, to our best knowledge, there is little research on MR remote collaboration which combines virtual replicas and gesture cues in the local user's 3D stereoscopic scene for industry assembly tasks. Therefore, it is worthwhile to explore whether using the combination of virtual replicas and commonly used non-verbal communication cues (such as gesture cues) in the shared real-time 3D stereoscopic scene could improve MR remote collaboration for industry assembly guidance.

The study provides some innovative and significant contributions:

1. Implementing a novel MR remote collaboration prototype system for industry physical tasks, especially for industrial assembly tasks. The system combines gesture cues and virtual replicas of physical counterparts synergistically in the shared real-time 3D stereoscopic scene of local workspace reconstructed by 3D video. To the best of our knowledge, this system is one of the first collaborative systems supporting the remote expert to create and manipulate virtual replicas of physical counterparts by gestures to intuitively and naturally demonstrate the assembly operation for the local worker in the shared real-time 3D stereoscopic scene.
2. Conducting the first user study comparing gesture cues and the combination of gesture cues and virtual replicas of physical counterparts in real-time 3D video-based MR remote collaboration in industrial assembly tasks.
3. Conducting the first user study exploring the benefits and implications of combining gesture cues and virtual replicas of physical counterparts in real-time 3D video-based MR remote collaboration in industrial assembly tasks.

The paper is organized as follows: firstly, we review prior related work in Sect. 2 and then describe our prototype system framework and technical details in Sect. 3. Thirdly, in Sects. 4 and 5, we report a pilot test and a formal user study to evaluate our prototype. Next, the results are presented in Sect. 6, and then we discuss the results and limitations in Sects. 7 and 8. Finally, some conclusions and future research directions are given in Sect. 9.

2 Related works

In this section, we first review the previous research on MR and AI-enhanced assembly assistance and compare it with remote collaboration. And then, we review prior research on remote collaboration from three closely related areas: collaboration using 3D stereoscopic scenes, non-verbal communication cues in remote collaborations, and collaborations using virtual replicas and virtual proxies.

2.1 MR and AI-enhanced assembly assistance

Currently, many industrial physical tasks, especially assembly tasks, are still manual, which can frequently result in possible operation errors and low efficiency due to the

[1] https://www.microsoft.com/zh-cn/hololens

[2] https://www.vive.com/cn/product/

different proficiency and experience of workers. To minimize these errors and improve efficiency, there has been a lot of research on AR/MR assembly assistance systems to support the manual operation of the worker by using AR/MR technology to provide visual or other forms of assistance [21–25]. And AR/MR assembly assistance has proven to be superior to traditional (paper-based or screen-based) instruction delivery in terms of accuracy and time [24, 25].

However, there are still some limitations. First of all, general MR assembly assistance systems need to author AR/MR content for assembly steps in advance, which is tedious and time-consuming. To overcome this limitation, there has been some research on rapid and easy AR instruction authoring [26–29]. Neb et al. [28] proposed a method to create AR assembly instructions quickly and intuitively with the help of virtual guidance. Bhattacharya et al. [29] developed an automatic AR work instructions generation system based on expert demonstration, named AREDA, which allows the novice author the ability to register the content intuitively without having to understand complicated AR concepts. However, it is not economical to create AR instruction for little batch customized assembly objects and the preset AR content may not conform to the actual assembly scenario, which leads to confusion.

AI-enhanced AR/MR assembly assistance systems support the perception of workers' action and assembly scenarios using AI-based algorithms, such as confirming the completeness of an assembly step or alerting the worker manipulating wrong objects, to automatically activate the AR instructions [30–33]. Su et al. [30] proposed an assembly state recognition method based on CNN to enhance the AR assembly assistance system, which provides step-by-step guidance by detecting the current state of an object composed of several parts. Chang et al. [31] developed an interactive AR disassembly sequence planning system (ARDIS) based on a multi-objective genetic algorithm, which allows the capability of considering user needs and automatically generating AR disassembly instructions. Dimitropoulos et al. [32, 33] focused on using the artificial intelligent algorithm, AI-enhanced wearable devices, and AR/MR technology to improve the human–robot collaborative system in terms of human–robot interaction and ergonomics, and proved that the proposed system can achieve simpler interaction and higher collaboration efficiency through industrial cases. However, on one hand, AI-enhanced assembly assistance systems need to train the AI algorithm in advance according to the workers' needs, preferences, and process states involved in the assembly task, which needs to collect a large amount of relevant data for training and is time-consuming. On the other hand, because the actual working scene is changeable and unexpected situations may occur, resulting in the instability of the AI-enhanced assembly assistance systems sometimes.

The rapid development of MR technology and increasing internet connectivity, and the powerful MR-related devices have provided available collaborative tools for remote collaboration. In this context, MR remote collaboration can help overcome the challenge of distance, make full use of the remote expert knowledge, and flexibly generate real-time AR instructions according to the actual situation of the site (including human activity and the state of parts), without authoring the AR content or training intelligent algorithm beforehand. Next, we will review in detail three areas closely related to the remote collaboration system proposed in the paper.

2.2 Collaboration using 3D stereoscopic scenes

Compared with the traditional 2D video-based remote collaboration, shared 3D stereoscopic scene-based remote collaboration can provide the depth perception of the local workspace, and make the observation perspective of the remote user free from the control of the local user. The local user's surrounding scene is reconstructed either live [15, 34–38] or beforehand [9, 10, 13, 39] and send the 3D reconstructed scene model to the remote side. In recent years, many studies have focused on evaluating the potential of 3D reconstruction in remote collaboration.

Using real-time 3D scene reconstruction in remote collaboration is a very active research field. Izadi et al. [36] developed a real-time 3D scene reconstruction technology that allows users to use an RGBD camera to capture and reconstruct the surrounding scene in real time, and allow users to explore and interact with virtual objects freely in the reconstructed scene. However, the texture data was left out, so the reconstructed scene by this method was limited to mesh level. Similarly, more recent work, BundleFusion [35] is a real-time and high-quality 3D reconstruction method using an RGBD camera, which supports large-scale scene reconstruction and texture information capture. However, this system does not support communication between remote users and local users.

Yang et al. [15] presented an AR remote collaboration system, which uses a Kinect RGBD camera to reconstruct people and objects, and shares them with other users by AR HMD, but it does not provide the ability to reconstruct the scene. Zillner et al. [12] proposed an AR remote collaboration system that uses a high-fidelity dense scene reconstruction technology to realize accurate and intuitive remote instructions. However, similar to InfiniTAM [38], the update speed of this method with respect to environmental changes is relatively slow. Nuernberger et al. [39] proposed an image-based 3D reconstruction method that uses a set of pre-captured images to reconstruct the scene. In this system, users can get a sense of immersion by observing pictures from a specific perspective. However, this system

did not support real-time image updating, so its application in collaborative tasks is limited.

In another case, Anton et al. [11] created a remote collaboration prototype system that uses a zSpace 3D display to show the local user's workspace by 3D video and provides some virtual tools for creating remote guidance at the expert side. However, this system did not provide the ability to depict objects in the local scene with high fidelity and can only support simple annotation.

In summary, although real-time 3D reconstruction has been used in remote collaboration in many cases, a more accurate depiction of the 3D scene and faster update speed is still open for further development. Since the virtual replicas can be represented by readily available offline high-precision CAD models, combining virtual replicas and real-time 3D video in a complementary manner is a good way to increase the model precision without additional bandwidth or update time. However, the effect of combining virtual replicas and real-time 3D video in MR remote collaboration has not been well explored.

2.3 Non-verbal communication cues in MR remote collaboration

Many past studies have shown that nonverbal communication cues have a greater effect on user performance than verbal communication cues in remote collaboration. Thus, many remote collaboration systems, such as Meta-AR-App [1], 3DGAM [3], RemoteBob [19], 2.5DHANDS [40], Vishnu [41], and other studies [42–45], focus on sharing non-verbal communication cues (e.g., text, 2D video, 3D models, gaze cues, gesture cues, virtual avatar). With respect to shared 3D stereoscopic scene-based remote collaboration, there also has been some related research that combined non-verbal communication cues with shared 3D stereoscopic scenes.

For example, the use of augmented feedback, such as annotations [11] on a 3D interface, could allow a remote expert to guide a worker without potentially vague verbal communication. This also applies to AR cues using the simple 3D model to create instructions. For instance, Venerella et al. [10] developed a portable remote collaboration system, which allows the remote expert to place the 3D model, such as pre-built annotation and virtual landmarks, on the 3D scene to guide the local user easily. In addition, gesture cues and gaze indicators enable users to better understand the relationship between objects in 3D scenes and perform tasks quickly. Teo et al. [9] developed a novel MR remote collaboration system which combines 360 video and 3D reconstruction. In this system, they found that sharing gesture cues and gaze indicators improve performance and user experience in search tasks. Moreover, sharing a life-size virtual avatar could improve social co-presence. Piumsomboon et al. [8, 13] presented the CoVAR MR collaboration system using virtual avatars combined with other natural inputs to enhance collaboration (e.g., gaze cues and gesture cues on a shared 3D reconstructed scene).

The research mentioned above showed that sharing non-verbal communication cues are very effective for improving MR remote collaboration. However, although these existing studies can allow a remote expert to guide local workers through non-verbal communication cues in 3D stereoscopic scenes, they cannot support a remote expert to pick up any physical objects in the local worker's scene for action demonstration. To overcome this limitation, in this study, we proposed an MR remote collaborative system, which enables the remote expert to use gestures to create virtual replicas of physical counterparts, just like picking up real physical objects.

2.4 Collaboration using virtual replicas and virtual proxies

There has been some research on using virtual replicas or virtual proxies in MR remote collaboration. Adcock et al. [46] developed a space remote collaboration system, which supports a remote expert to manipulate 2D proxies of physical counterpart structured beforehand on a multi-touch display to use gestures to create 2D translations and rotations guidance to project onto the local user's workspace. However, this approach is difficult to deal with complex surfaces which cannot be projected easily. To overcome such limitations, Tait et al. [47] extend the research of Adcock et al. [46] so that a remote expert is allowed to place virtual replicas corresponding to physical objects in a 3D scene model by a 2D monitor. However, in contrast, remote experts are more flexible when viewing the local user's workspace in immersive VR space. Moreover, Yang et al. [15] presented an AR remote collaboration system, which shares virtual replicas of real objects by physical virtualization technology, and they found that sharing virtual replicas helps to improve efficiency and provides a smoother experience. However, the research only focused on simple physical tasks.

Inspired by Voodoo-doll [48], Elvezio et al. [2, 20] introduced an MR remote assistance system for the 6DOF alignment task in industry. In this system, the remote expert could view the virtual proxies of important tracked objects from the local user's environment and create and operate virtual replicas of physical counterparts to demonstrate operation for the local user in a virtual workspace. However, in this system, due to the differences between the virtual scene and the real scene, it is difficult for remote experts to see accurately and comprehensively what is happening in the real local worker's scene. In addition, Wang et al. [3, 49] described a novel MR remote collaboration prototype system, which combines gesture cues with CAD models of real

parts to create instructions for the local worker. They found this combination can improve collaboration efficiency and user experience. However, this is a fractured ecology where the remote expert needs to shift their attention frequently between tasks and the 2D video of the local scene.

These studies mentioned above presented that sharing virtual replicas can help improve the efficiency and user experience of MR remote collaboration. However, in shared real-time 3D video-based MR remote collaboration, the effect of combining virtual replicas with gesture cues in industry assembly tasks has not been well investigated.

2.5 Summary

From the review of the above research, we can draw the following three conclusions. First, although real-time 3D reconstruction has been used in remote collaboration in many research cases, more accurate and faster 3D spatial description methods are still worth further development. Second, non-verbal communication cues have gained more and more attention to improve MR remote collaboration. However, these cues are difficult to represent the reference and operation of physical objects in the sharing 3D stereoscopic scene of the local user. Third, the effect of combining virtual replicas with nonverbal cues, such as gesture cues, has not been well investigated in MR remote collaboration based on the 3D video in physical tasks, especially in industrial assembly tasks.

Therefore, we develop a novel MR remote collaboration prototype system. Firstly, the system combines high-precision CAD models as virtual replicas and shares the stereoscopic scene of the local worker by the 3D video which increases scene description precision for effective spatial reference without additional bandwidth and update time. Secondly, the system combines gesture cues and virtual replicas in a complementary manner to support the remote expert to create and manipulate virtual replicas of physical counterparts by gesture to demonstrate actions for the local worker naturally and intuitively. Thirdly, based on this prototype, we performed a formal use study to let the local participant as the worker assemble a real vise with the guidance from the remote participant as the expert to explore the advantages of combining virtual replicas with gesture cues in 3D video-based MR remote collaboration in industrial assembly tasks compared with using gesture cues only.

3 Prototype system

Our goal was to design an MR remote collaboration system which supports a remote expert to easily demonstrate assembly action to a local worker in the 3D stereoscopic scene of the local workspace. In this case, the remote expert can interact with the shared real-time 3D video of the local workspace by creating and manipulating 3D virtual replicas of physical counterparts using gestures to demonstrate action in the VR environment, and the local worker can imitate the remote expert demonstration to accomplish physical tasks in the AR environment. We describe our prototype system and implementation details in three aspects: (1) system framework, (2) sharing the real-time 3D stereoscopic scene of the local workspace, and (3) interaction techniques.

3.1 System framework

Our MR remote collaboration system connects a local AR worker side with a remote VR expert side for real-time remote collaboration. Figure 1 presents a schematic diagram of the core elements in our prototype system. It includes three main modules: (a) a server used for assembly process resources management and synchronization, (b) a remote VR site presenting the VR collaborative scene for the remote expert based on the shared real-time 3D video, (c) a local AR site presenting AR instructions for the local worker based on the combination of virtual replicas and gesture cues.

Our prototype was developed on the Windows 10 operating system using Unity3D 2017.4.17f1 game engine with 64-bit WampServer,[3] Microsoft's MixedRealityToolkit[4] (MRTK), Intel's RealSense SDK 2.0,[5] and Point Cloud Library[6] (PCL). An Intel® NUC is used as a server to connect each VR/AR client through a wireless network. Similar to previous research [2, 3, 42, 43, 50–52], to simplify the system to the core elements aligned with the research focus, we used the co-located collaboration to simulate remote collaboration. That is, the VR/AR site users in a room were separated with a physical gap, and they can still talk and hear even if they could not see each other.

For the remote expert site, as shown in Fig. 2a, the client was relying on an OMEN laptop with an HTC VIVE Pro 2 VR display and a Leap Motion[7] hand tracker. The real-time 3D video from the local worker site was rendered in Unity3D after coordinate unification. The remote expert can navigate and observe freely around the 3D stereoscopic scene by wearing VR HMD. To support interaction with the 3D reconstructed scene, the client tracks hands in real-time

[3] https://www.wampserver.com/

[4] https://github.com/topics/mixedrealitytoolkit-unity

[5] https://www.intelrealsense.com/developers/

[6] https://pointclouds.org/

[7] https://www.ultraleap.com/product/leap-motion-controller/

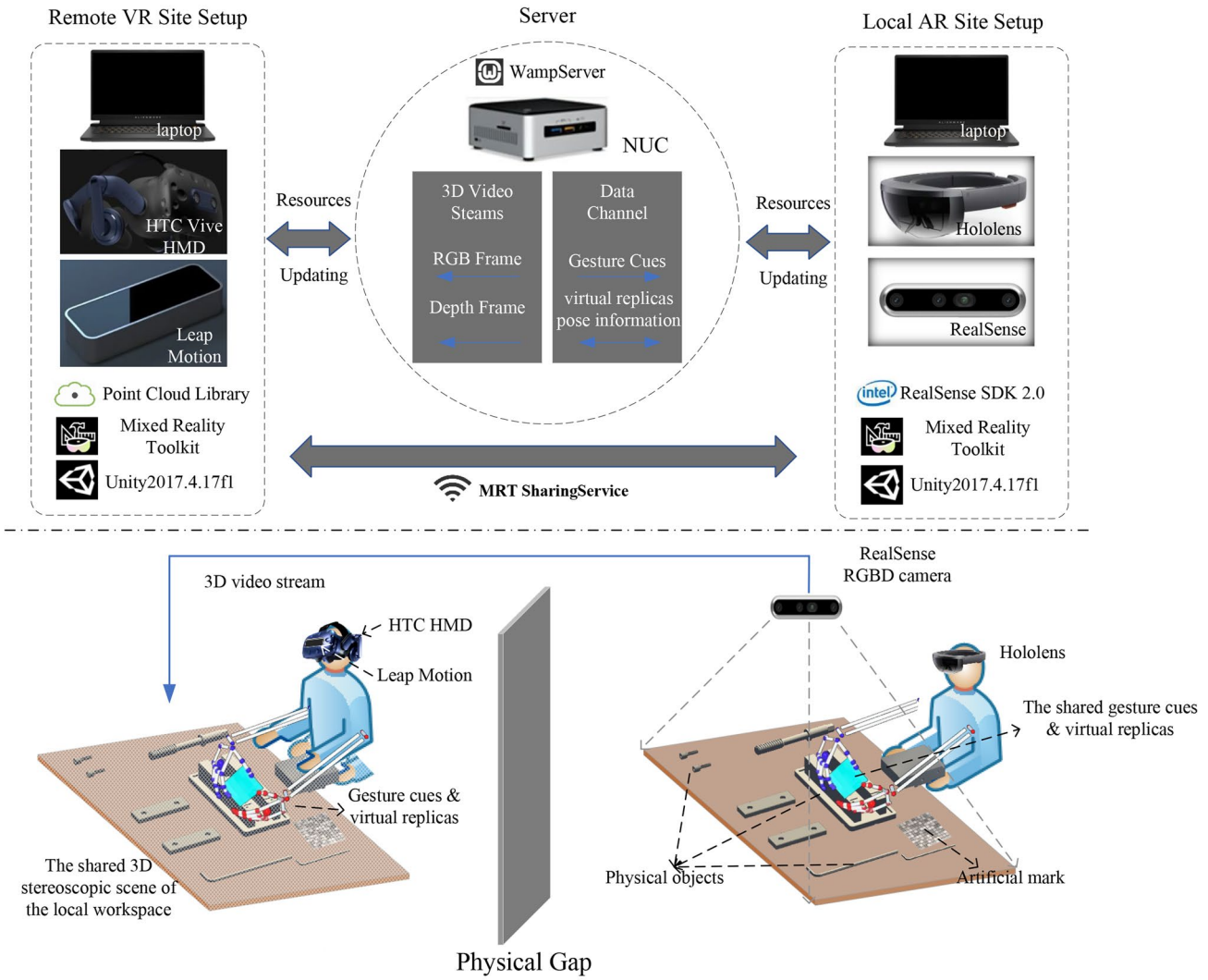


Fig. 1 The framework of the MR remote collaboration prototype mainly includes a server, a remote VR site, and a local AR site

with the Leap Motion hand tracker fixed on the VR HMD, so that the remote expert can create and manipulate virtual replicas of real counterparts by using gestures (e.g., grasp, translate, and rotate). And then, the action demonstration by gesture and virtual replicas will be transmitted to the local site to guide the local worker.

For the local worker site, as shown in Fig. 2b, the client was relying on an ALIENWARE laptop with a HoloLens

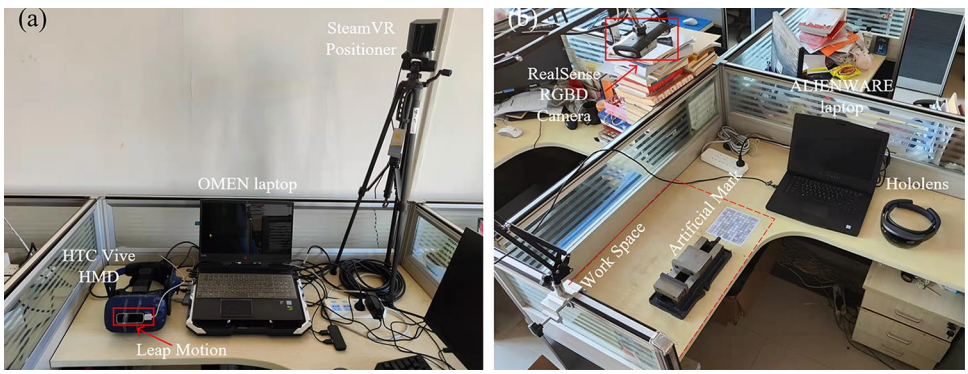


Fig. 2 **a** Remote VR expert side setup. **b** Local AR worker side setup

AR HMD and an Intel® RealSense™ D435i Depth Camera.[8] The RealSense D435i supports high RGB frame resolution (1920 × 1080 pixels) with 30 fps and high depth frame resolution (1280 × 720 pixels) with 90 fps. It was fixed on the top of the head to capture the depth and RGB texture frames of the local workspace. Then, 3D video composed of the depth and RGB texture frames was transferred to the remote expert site to support the remote expert to observe and interact with. An artificial mark was fixed on the workspace as the unified benchmark between the virtual scene coordinate system and the real scene coordinate system, so as to superimpose the virtual instructions from the remote site on the real scene and guide the local worker to perform the assembly task through AR headset.

The key characteristics of our prototype system include the following several points: (1) use co-located collaboration to simulate remote collaboration so that users can hear and speak to each other, (2) the remote expert can navigate and observe the local worker's workspace freely in VR through the real-time 3D stereoscopic scene constructed by 3D video, (3) the remote expert can create and manipulate virtual replicas of physical counterparts by gesture in VR space to demonstrate action, and (4) the local worker can view the AR virtual guidance superimposed on the real workspace by AR HMD.

3.2 Sharing real-time 3D stereoscopic scenes

In many previous studies, the schemes of sharing 3D stereoscopic scenes were to share static 3D reconstructed models or sharing the real-time fused 3D reconstructed scene based on a hand-held RGBD camera or integrated depth sensor AR HMD. These methods allow for the sharing of high-quality, geometrically precise 3D models of local workspace. However, they cannot present the real-time changes that occur on the local workspace and the real-time operations of local workers [53], even if the local worker could scan and reconstruct the scene multiple times. To overcome this limitation, we developed a method to share a stereoscopic scene through an RGBD camera, similar to prior research [11]. The RGB texture and depth frames are captured by the RGBD camera at the same time in each frame. Then, the color and depth frames are transmitted to the remote expert site based on the SharingService application from MRTK to form a 3D stereoscopic scene, supporting real-time updates. In our prototype, we enabled the 3D video streaming to transmit RGB + D data, which the SharingService application does not support by default.

Before our 3D video module can run online, it needs offline calibration and some other preparation steps. These steps were performed once after the camera was fixed and then reused in subsequent online operations. Firstly, the RGBD camera was calibrated by Intel RealSense D400 series dynamic calibration tools[9]. Then, we used this RGBD camera to capture the artificial mark on the workspace by Vuforia[10] to define a relative rigid transform from the camera pose to the mark. Finally, we set the center of the virtual mark which is consistent with the real counterpart's size as the world coordinate center of the VR space, and place the virtual RGBD camera relative to the virtual mark according to the rigid transform just calculated, so that the 3D video stream coordinate system is aligned with the VR site coordinate.

The pipeline overview of our 3D video module has four main 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 is encoded and wirelessly streamed to the remote side through MRTK, (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.

The purpose of our prototype is to demonstrate the proof-of-concept of using the combination of virtual replicas and gesture cues on shared 3D video-based MR remote collaboration. Therefore, the 3D video stream from only one RGBD camera may cause the reconstructed 3D scene to be incomplete from some observation perspective, thus degrading the quality of the 3D stereoscopic scene. Fortunately, the 3D stereoscopic scene from the 3D video can be compensated by using virtual replicas and virtual proxies represented by high-precision CAD models, so as to enhance the observation of remote experts on task-related objects (see detail in Sect. 3.3.1).

3.3 Interaction techniques

We will introduce the interaction techniques in our system in detail from two aspects: (1) creating and loading virtual replicas, and (2) gesture-based interaction.

3.3.1 Creating and loading virtual replicas

In our system, both the remote site and the local site can load virtual replicas from the server and synchronize the location in real time. Before and during our prototype running, it needs some crucial steps.

[8] https://www.intelrealsense.com/depth-camera-d435i/

[9] https://www.intel.com/content/www/us/en/download/645988/intel-realsense-d400-series-dynamic-calibration-tool.html?

[10] https://developer.vuforia.com/

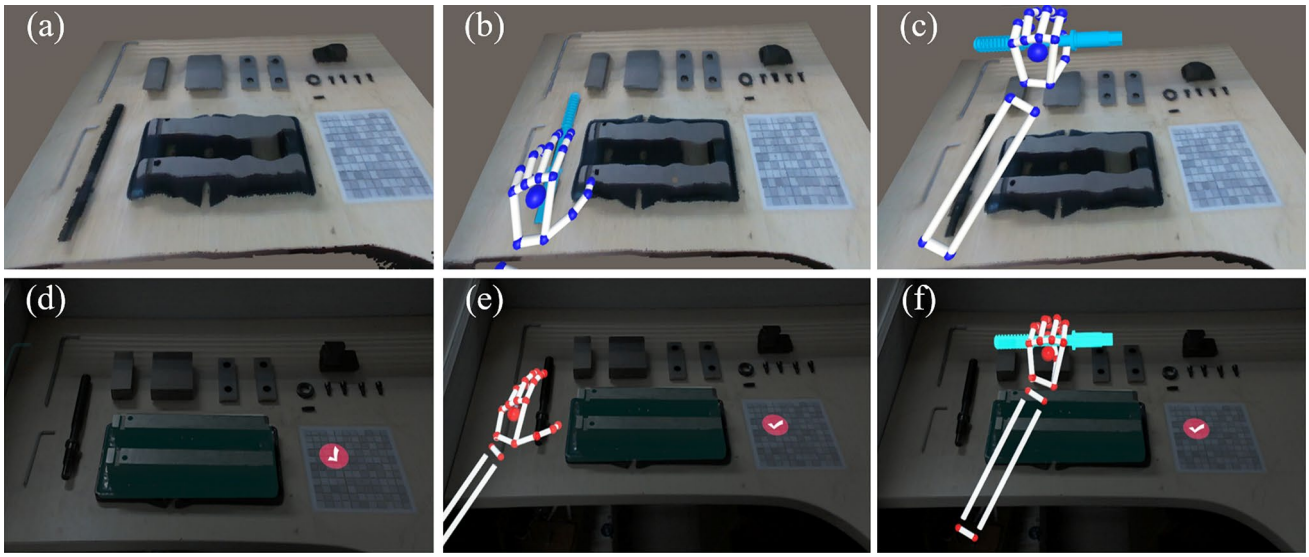


Fig. 3 The visual information of virtual proxies and virtual replicas. **a–c** The remote VR view; **d–f** the local AR view. **a** The remote VR expert does not perform the demonstration, and virtual proxies are completely transparent; **b** remote VR expert's virtual hand gesture close to the reconstructed point cloud, and the virtual proxy becomes opaque; **c** remote VR expert demonstrates assembly operation using virtual replicas and gesture cues; **d** the red checkmark icon superimposed on the artificial mark which indicates that the AR and real coordinate system has been successfully unified; **e** the virtual hand of the remote expert approaching the real part; **f** the AR guidance superimposed on real work scene that the remote VR expert using virtual replicas and gesture cues to demonstrate assembly operation

1. Creating prefab for virtual replicas

   The physical objects related to tasks in local user's workspace which the remote expert interacts with had been modeled with the same geometric size, in the study Lego bricks and parts of a vise, by using SolidWorks 2017. Additionally, physical objects in the local workspace which not related to the task were not modeled, preventing the remote expert from referring to them and causing confusion. Then, these models were made into prefabs and asset-bundles using Unity3D and they can be instantiated into 3D models on demand. More detailed information about prefabs and asset-bundles can be found in the official documentation of Unity3D.

   Our system supports loading prefabs and asset-bundles to the other clients through the network using the WampServer, and each site can instantiate these resources into the Unity3D scene. The instanced virtual replicas with the same ID will synchronize the posture in real time by MRTK between each site.

2. Virtual proxies and virtual replicas

   In our prototype system, the spatial position of virtual proxies in VR spatial coordinate system is consistent with that of the physical counterparts' position in the real workspace coordinate system, and the remote expert cannot change it. In contrast, once virtual replicas are created, the remote expert can manipulate them at will.

   We adopt a method based on the chamfer matching algorithm [54] and color consistency to track the real object in 6DOF coarsely. This method is divided into two stages, (1) an offline preparation stage is used to extract edge templates and color of assembly task-related components CAD model by the virtual camera, and (2) an online execution stage is used to extract edge from RGB stream for matching with edge templates to locate the parts' location. If there are multiple candidate targets, color consistency is used to determine the unique match. The advantage of this method is that the templates directly from the CAD model can be used to track the corresponding real object, therefore, it is more suitable for our research on industrial applications. In order to simplify the system to the core elements aligned with the research focus and improve the system performance, we locate the initial position of the real parts in the local workspace through the first frame from the RGBD camera, and they, in turn, define the position of virtual proxies.

The coordinate system of virtual proxies, 3D video scene, VR space, AR space, and real space were unified by artificial mark. Thus, in VR view, the position and orientation of virtual proxies coincide with reconstructed counterparts in 3D video and in AR view, the AR guidance created by the remote expert can be accurately superimposed on the real local work scene. In Fig. 3, we showed the visual information of virtual proxies and virtual replicas in the remote VR view and the local AR view. In the VR view, when the remote expert does not perform the demonstration, virtual proxies are completely transparent, as shown in Fig. 3a. In the AR view, the red checkmark icon is superimposed on the artificial mark which indicates that the AR and real

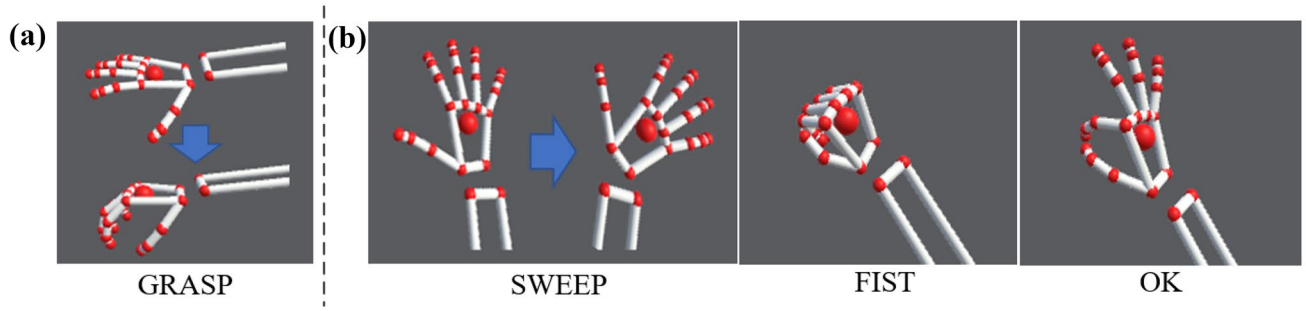


Fig. 4 **a** The chosen gesture for virtual replicas creating. **b** The candidate gestures for virtual replicas clearing

coordinate system has been successfully unified, as shown in Fig. 3d. When the remote expert intends to manipulate the object in the 3D reconstructed scene to demonstrate the operation, as the remote expert's virtual hand moves close to the reconstructed point cloud, the physical object's virtual proxy will become opaque to cover the physical counterpart's reconstructed point cloud, so that the remote expert can observe a high-precision 3D model in VR view, as shown in Fig. 3b. At the same time, in the AR site, local workers can observe the virtual hand gesture of the remote expert approaching the real part, as shown in Fig. 3e. And then, the remote expert can try to grab the corresponding virtual proxy to create a virtual replica of the physical counterpart to demonstrate action, as shown in Fig. 3c. At the same time, in AR view, the local worker can observe the AR guidance superimposed on real work scene that the remote VR expert demonstrating assembly operation using virtual replicas and gesture cues, as shown in Fig. 3f. More details about gesture-based interaction are described in Sect. 3.3.2.

3.3.2 Gesture-based interaction

Our prototype supports intuitive and natural gesture-based interaction, and the basic input of the remote expert is hand operation demonstration. We implement the gesture-based interaction by a Leap Motion hand tracker fixed on the HTC Vive VR HMD, which is commonly used in MR related research.

To find suitable gestures to create/delete virtual replicas, we asked 10 participants three questions as follows: "(1) what gestures do you want to use to create/delete virtual replicas? (2) what gestures do you think are easier to learn to create/delete virtual replicas? (3) what gesture-based interaction do you think is more intuitive and natural to use virtual replicas to demonstrate actions?".

We found that when participants are asked to describe the operation of assembling a part, users tend to use a dynamic gesture to help express ideas more easily. For creating virtual replicas, most users chose the GRASP gesture, as shown in Fig. 4a. This dynamic "GRASP" gesture is consistent with the way we usually grab objects, and it would be equivalent to the remote expert actually picking up a part from the local worker workspace, adjusting its position and posture by gesture to fit it to another part to demonstrate assembly action. In our approach, when the remote expert tries to grab virtual proxies of the physical object in the 3D reconstructed scene, the virtual replica of a physical counterpart would be created automatically. Then, the remote expert can directly manipulate the virtual replica by holding the "GRASP" gesture until the assembly step is completed. The virtual replica will remain in the location where it was released and can be grabbed again by the "GRASP" gesture to manipulate it further. The virtual replica will be automatically absorbed to the correct assembly position when it gets within a preset threshold distance relative to the target position. In the whole demonstration process of the remote expert, the state of gesture and virtual replicas are transmitted to the local AR site and superimposed on the real workspace. And then, the local worker follows these instructions to perform the action.

For clearing virtual replicas, most users chose the dynamic gesture "SWEEP," the static gesture "FIST," or "OK," as shown in Fig. 4b. However, sometimes, the Leap Motion might incorrectly identify the "GRASP" gesture as the "FIST" gesture, resulting in the inability of remote experts to create virtual replicas. Additionally, since the "OK" gesture usually indicates the meaning of correct, using it as the gesture to delete virtual replicas may confuse local workers. Therefore, we define the dynamic gesture "SWEEP" to delete virtual replicas and set the static gesture "OK" to prompt the local workers that the current assembly step has been completed correctly. Moreover, the properly assembled virtual replica will become light blue and transparent when the remote expert shows the "OK" gesture, and can no longer be operated on to avoid interfering with the remote expert.

To sum up, in Fig. 5, we present visual cues combined with gesture cues and virtual replicas for an assembly step in our remote collaboration system.

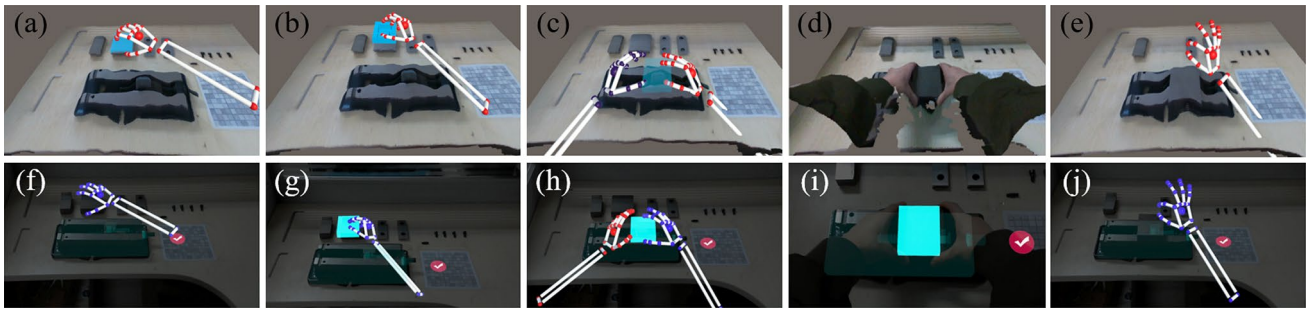


Fig. 5 Visual cues for an assembly step in our remote collaboration system. **a**–**e** The VR view for the remote expert; **f**–**j** the AR view for the local worker. **a**, **f** The remote expert's virtual hand gesture approaches the part, and its virtual proxy become opaque; **b**, **g** the remote expert creates the virtual replica of the part by using dynamic GRASP gesture and operates it to demonstrate action and the local worker can observe the AR guidance superimposed on real work scene; **c**, **h** the remote expert demonstrates how to assemble the part; **d**, **i** the local worker performs the assembly according to the expert demonstration and the remote expert can observe the action of the local worker through real-time 3D video; **e**, **j** the remote expert using OK gesture to indicate that the local worker has assembled the part correctly, at the same time, the virtual replica of the correctly assembled part becomes light blue and transparent in both VR and AR view

4 Pilot test

Before the formal user research, an informal pilot study was conducted with ten compensated participants (five pairs). They are all college students and majored in Mechanical Engineering and Manufacturing. All of the participants have used video-based conferences such as WeChat or Skype and some experience of VR/AR games. The remote participant wore an HTC Vive VR HMD with Leap Motion hand tracker to guide the local participant wearing a HoloLens AR HMD to perform the Lego brick assembly task in the local workspace according to our interactive method, in which arranged the RGBD camera for capturing real-time 3D video. To prevent the influence of remote experts' proficiency on the assembly task, on the VR site, a virtual panel was set in front of the expert to display the prompt information of the current assembly step. When the remote expert uses the OK gesture to remind the local worker that the assembly is completed correctly, the prompt information will automatically switch to the next assembly step. In Fig. 6, we present the remote collaboration scenario for the Lego brick task in the pilot test phase. Each participant would complete a System Usability Scale questionnaire (the specific details of SUS [55]) after completing each trial as illustrated in Fig. 7. They were also given some time to explore our system freely, and we collected the subjective feedback of the participants afterward.

The result of the SUS questionnaires is presented in Fig. 8. For local participants, the average SUS score was 84 (standard error (SE) = 3.1) and for remote participants, the average SUS score was 90.5 (SE = 3.75); therefore, all belong to the good usability [55].

The subjective feeling feedback of participants was generally positive, such as "I can easily observe my partner's environment; This is a beginner-friendly system, I don't need to take much practice to operate it; This interaction approach gives me a very natural and intuitive feeling; It's

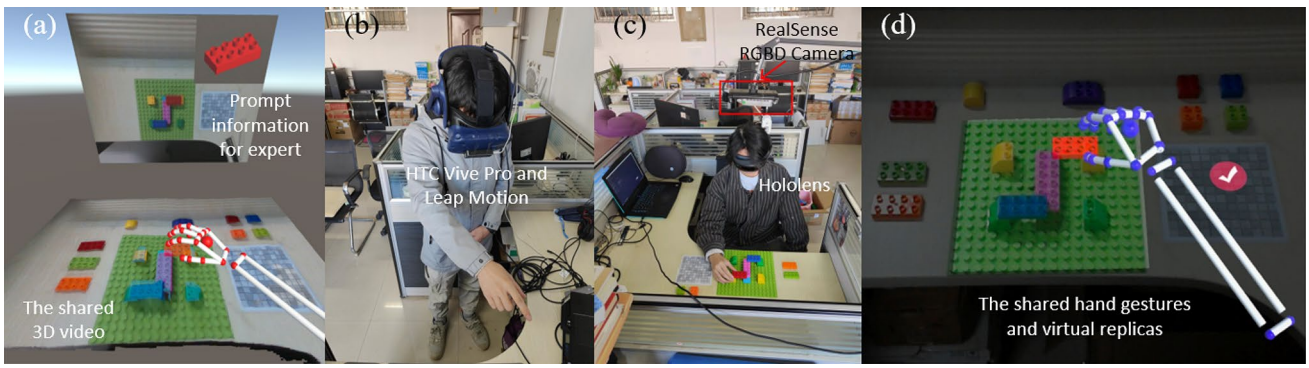


Fig. 6 The remote VR expert guide local AR workers to assemble the Lego brick model using hand gestures and virtual replicas. **a** The HTC Vive view; **b** the remote expert is creating guidance wearing the VR HMD with Leap Motion hand tracker; **c** the local worker is performing assembly of the Lego brick model following guidance shown in the AR HMD; **d** the HoloLens view

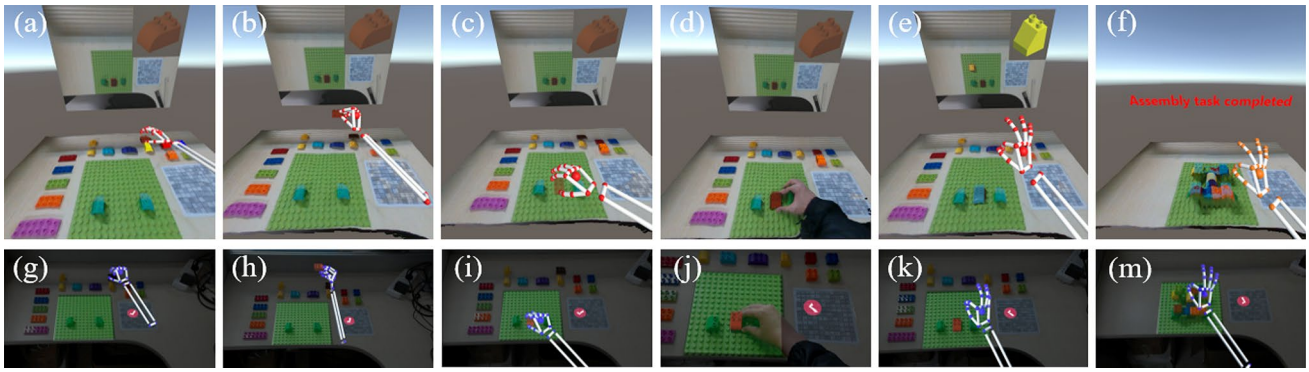


Fig. 7 Typical steps of MR collaborative assembly of Lego brick model in the pilot test. **a**–**f** The VR HMD view; **g**–**m** the AR HMD view. **a**, **g** Expert's gestures approach the brown part, and its virtual proxy become opaque; **b**, **h** expert creates the virtual replica of the brown part through GRASP gesture and operates it; **c**, **i** expert assembles the brown part; **d**, **j** worker performs the assembly according to the expert demonstration; **e**, **k** the brown block is correctly assembled; **f**, **m** assembly task completed

an interesting experience to use it; It's amazing to pick up a digital copy of an object from my partner's environment; I can easily understand my partner's instructions; The existence of virtual proxies makes it easier for me to recognize objects in the 3D video." However, some VR participants mentioned that the prototype system is not very sensitive when trying to grasp the small object to create the virtual replica. They also said that the simultaneous existence of virtual replicas of multiple objects may confuse local workers. Importantly, some participants suggested that it would be more convincing to compare two conditions in the formal user study: one only shares gesture cues just like the state-of-the-art approaches, and the other shares the combination of gesture cues and virtual replicas, to perform a relatively complex industry assembly task.

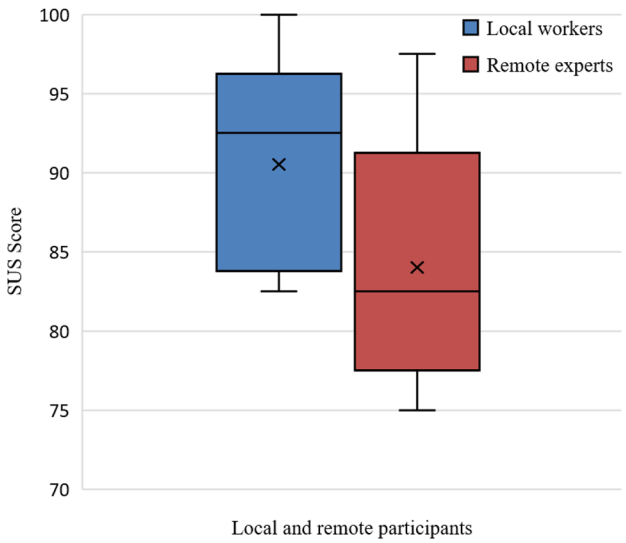


Fig. 8 The SUS questionnaire results of local and remote participants

According to the feedback of participants, we improved the research in the aspects of gesture-based interaction, the form of virtual replicas, task content, and comparison conditions. First, we increased the collision volume of smaller parts to make them easier to grasp by leap motion. Second, our prototype system now only allows one virtual replica to exist at the same time, in other words, when a new virtual replica is created by the expert, the old one will be deleted automatically. In addition, the properly assembled virtual replica will become completely transparent in the VR site when the remote expert shows the "OK" gesture. Third, we conducted a formal user study to compare two conditions under a more complex task in industry (see Sect. 5 for more details).

5 Formal user study

In the study, in order to evaluate the influence of using the combination of virtual replicas with gesture cues for industry assembly tasks in shared real-time 3D video-based MR remote collaboration, a formal user study on our prototype system was conducted. We will elaborate our user study in detail in the following three aspects: (1) conditions, (2) task content and experimental hypotheses, and (3) participants and procedure.

5.1 Conditions

Two scenarios were set with an independent condition, (1) using gesture cues in 3D video (G3DV) and (2) using the combination of virtual replicas with gesture cues in 3D video (VG3DV).

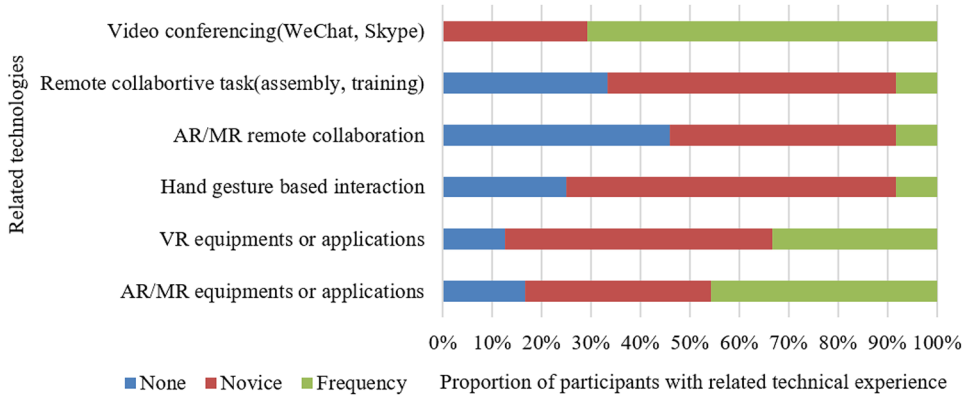


Fig. 9 Participants' experience with related technologies

5.1.1 Condition1

G3DV In this condition, the remote VR expert can use gesture cues to create instructions in the shared 3D video of the local workspace, and the local worker can see the gesture cues superimposed on the workspace on the AR site, which is similar to the popular method of combining gesture cues with 3D stereoscopic scenes [8, 9, 13].

5.1.2 Condition 2

VG3DV In this condition, the remote VR expert can create and manipulate the virtual replicas of physical counterparts by gesture in the shared 3D video of the local workspace, and the local worker can observe the AR guidance that the remote expert manipulates the virtual replicas by using gestures superimposed on the workspace.

5.2 Task content and experimental hypotheses

According to the feedback of pilot tests, we accepted the suggestions of participants to conduct more complex industry assembly tasks for the formal user study. The main task of the experiment is that the local participant as a worker assembles a vise model with the guidance from the remote participant as the expert.

Our research is focused on the factors of performance and user experience from MR remote collaboration based on our interactive approach (that is, VG3DV condition) in industrial assembly tasks. Therefore, we hypothesized that the VG3DV compared with the G3DV could improve more collaboration efficiency, fewer operation errors, better user experience, lower workload, and higher preference. Thus, we make the following five hypotheses:

H1: Performance. Compared to the G3DV, the VG3DV interface will provide a significantly faster performance on industry assembly tasks.
H2: Error. The VG3DV condition will significantly reduce operating errors.
H3: Workload. The VG3DV condition will present a lower workload.
H4: User experience. Compared to the G3DV, the VG3DV condition will provide a significantly better user experience for all roles of participants in terms of confidence, enjoyment, focus, etc.
H5: User preferences. Most participants will prefer the VG3DV condition to the G3DV condition.

5.3 Participants and procedure

In the formal user study, we invited 24 college students (12 pairs) from our university, including 22 males and 2 females, aged from 22 to 34 (M = 28 years, SD = 0.7), and were all right-handed. Their major backgrounds were diverse such as Mechanical Engineering, Mechanical Electronics, and Industrial Design. In Fig. 9, we present more detail of participants' experience using related technologies. All participants volunteered to participate in the study and everyone received a souvenir worth 30 CNY (about $ 4.6).

We conducted a within-subject user study, and each participant pair performs two rounds (e.g., VG3DV and G3DV) of an experiment. Each participant was randomly assigned as a remote expert or a local worker, and they did not change roles between the two sites during the experiment.

The user study procedure mainly includes followed the six steps: (1) introduction, (2) filling out background questionnaire, (3) training and explaining details, (4) completing the assembly task in one condition (VG3DV or G3DV), (5) filling user experience questionnaires(see Table 1), (6) repeating steps (4) and (5) with exchanging the condition

Table 1 Likert scale rating items of the collaborative experience questionnaire

G#	Questions: 1 (I strongly disagree) to 7 (I strongly agree)
G1–G8 for all participants	
G1	The interface was **natural and intuitive**
G2	I was able to stay **focused** on the task actively
G3	I felt very **confident** that we completed the task correctly
G4	I **enjoyed** the collaborative experience
G5	I was **satisfied** with my task performance
G6	I caught my partner's **attention**
G7	I **reciprocated** (my partner's) actions well
G8	It was easy to **collaborate** together for assembly tasks
G9 and G10 for participants on the remote site	
G9	It was easy to **provide** clear **instructions** in real-time
G10	I can help my partner when he(she) needed **assistance**
G11 and G12 for participants on the local site	
G11	It was easy to **understand** my partner's instructions
G12	The instructions from my partner were **helpful**

between the VG3DV and the G3DV following a Latin Square Sequence to reduce learning effects, and ranking the two conditions according to preference (see Table 2), (7) allowing participants explore the prototype freely and collecting their subjective feedback. The whole experiment process for each pair of participants took about 45 min. In Fig. 10, we presented more details about the experimental procedure.

Prior to the start of the experiment, participants on the remote VR site were trained to assemble the vise for providing guidance to local participants during the experiment. After the task started, the remote expert creates guidance for the local worker to assemble the vise step by step with the appropriate tool. We recorded the task completion time and operation errors for local and remote participants (e.g., WPA is the number of wrong parts assembled, and IGP is the number of incorrect guidance provided). For errors without obvious causes, we interviewed the participants on the error after the trial.

Figure 11 shows the scenario of our MR remote collaboration prototype system for vise assembly task guidance in our formal user study. The main process of assembling the vise is shown in Fig. 12.

Table 2 Ranking criteria

R#	Which interface do you think was best…
R1	at helping you keep **focused** on the task actively?
R2	at making you feel more **confident**?
R3	at making you feel **satisfied** with the task performance?
R4	at helping you **collaborate** more easily with your partner?
R5	at making you **enjoy** the task process?
R6	at making you feel more **passion**?

Our evaluation for the user experience was through the NASA Task Load Index (NASA-TLX) questionnaire and the collaborative experience questionnaire. NASA-TLX is a multidimensional, validated, reliable, and standardized subjective test which can quantitatively evaluate the task workload [56]. The collaborative experience questionnaire (see Table 1) refers to networked minds measure of social presence questionnaire [57] and a few past MR remote collaboration research [5, 9, 43, 58].

6 Results

In this section, firstly, we present the analysis results of the objective data, including performance and error evaluation. Next, the analysis results of the subjective data are presented, including NASA-TLX, collaborative experience questionnaire, and ranking. Before analyzing the results, we checked the normality validation and consistency of the collected data, and no deviation was found. To check whether there were significant differences between conditions, we performed the statistical analysis at the $p = 0.05$ significance level.

6.1 Performance

As shown in Fig. 13, we present the time needed to complete the remote collaborative assembly task across two different conditions. To analyze the performance, a paired t-test ($\alpha = 0.05$) was performed to compare the results, and it showed significant differences ($t(11) = 7.126$, $p < 0.001$) between the G3DV condition and VG3DV condition on the time spent. Moreover, descriptive statistics showed that participants took less time using our novel VG3DV interface

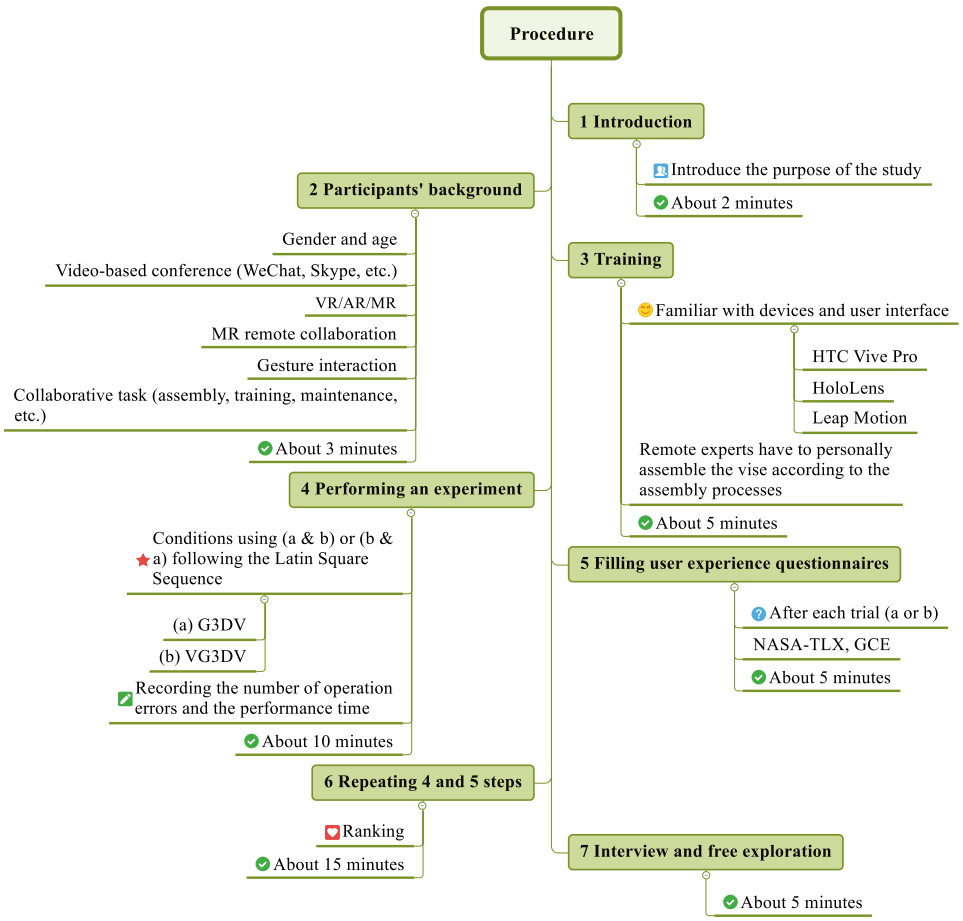


Fig. 10 The procedure of the formal user research

(M = 412.58 s, SE = 18.35) than the traditional G3DV interface (M = 479.91 s, SE = 16.31) in vise assembly task.

6.2 Error evaluation

Figure 14 presents the average operation errors for each categories. The results of the Wilcoxon signed rank test ($\alpha = 0.05$) show significantly differences in IGP ($Z = -2.000$, $p = 0.046$) and WPA ($Z = -2.754$, $p = 0.006$) between the G3DV and the VG3DV interface in vise assembly task. Compared with using G3DV interface (IGP: M = 0.5, SE = 0.19; WPA: M = 1.5, SE = 0.32), remote and local participants using our VG3DV interface (IGP: M = 0.17, SE = 0.11; WPA: M = 0.42, SE = 0.14) made fewer errors.

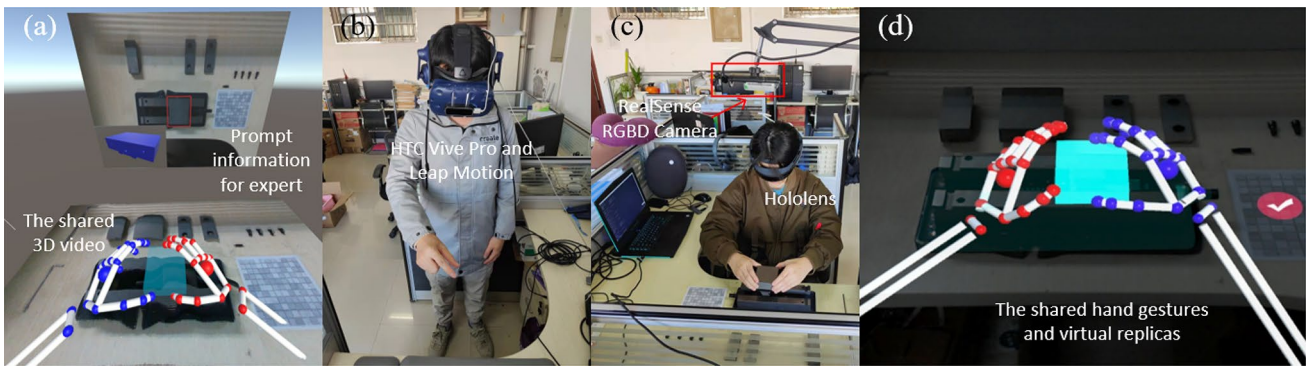


Fig. 11 The remote VR expert guide AR local workers to assemble the vise model using hand gesture and virtual replica. **a** The HTC Vive view; **b** the remote expert is creating guidance wearing the VR HMD with Leap Motion hand tracker; **c** the local worker is performing assembly of the vice model following guidance shown in the AR HMD; **d** the HoloLens view

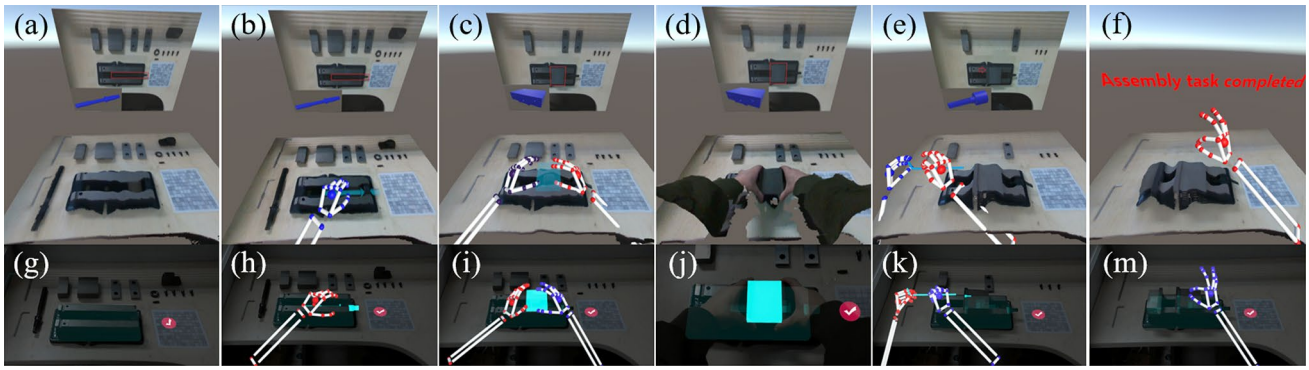


Fig. 12 Typical steps of MR collaborative assembly of the vise model in the formal user study. **a**–**f** The VR HMD view; **g**–**m** the AR HMD view. **a**, **g** Initial assembly scenario; **b**, **h** experts guide the assembly of the long shaft part with the left hand; **c**, **i** experts guide the assembly of the large cover part with both hands; **d**, **j** worker performs the assembly according to the expert demonstration; **e**, **k** experts guide the assembly of the large nut using the large hex wrench; **f**, **m** assembly task completed

6.3 NASA-TLX

In Fig. 15, we summarized the workload assessment collected by the NASA-TLX questionnaire, and the results show that the VG3DV interface requires a lower degree of workload for both remote experts to provide instructions and local workers to perform the vice assembly task. The paired t-test ($\alpha = 0.05$) showed significant differences for both remote experts ($t(11) = 2.939$, $p = 0.013$) and local workers ($t(11) = 2.450$, $p = 0.032$) between each interface condition.

6.4 User experience

To evaluate the collaborative experience, we analyzed the user feedback to the questions presented in Table 1, with respect to G1 (interaction), G2 (focus), G3 (confidence), G4 (enjoyment), G5 (satisfaction), G6 (attention), G7 (reciprocation), G8 (collaboration), G9 (providing instructions), G10 (assistance), G11 (understanding), and G12 (helpfulness). The Wilcoxon signed rank test ($\alpha = 0.05$) was conducted to check if there were significant differences between the G3DV interface and the VG3DV interface. In Figs. 16 and 17, we showed the results of the user experience questionnaire from remote and local participants respectively.

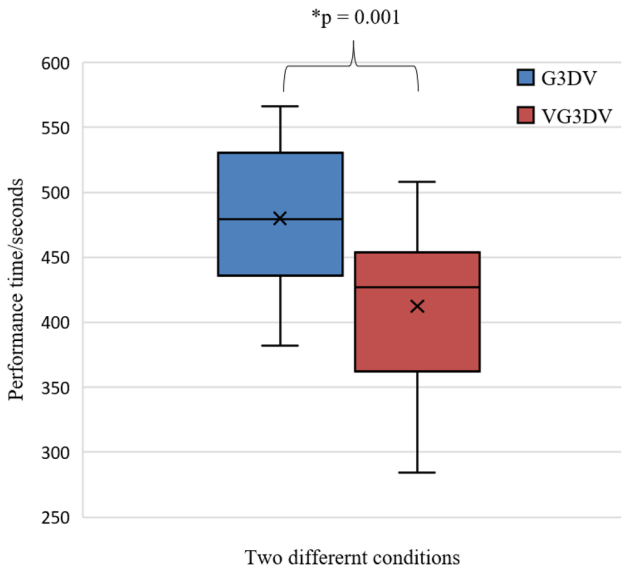


Fig. 13 The completion time

For participants in VR remote site, as showed in Fig. 16, there were significant differences in terms of interaction (G1: $Z = -2.692$, $p = 0.07$), focus (G2: $Z = -2.511$, $p = 0.012$), confidence (G3: $Z = -2.144$, $p = 0.032$), enjoyment (G4: $Z = -2.200$, $p = 0.028$), attention (G6: $Z = -1.997$, $p = 0.046$), reciprocation (G7: -1.980, $p = 0.048$), collaboration (G8: $Z = -2.654$, $p = 0.008$), providing instructions (G9: $Z = -2.821$, $p = 0.005$), and assistance (G10: $Z = -2.503$, $p = 0.011$).

For participants in AR local site, as showed in Fig. 17, there were significant differences in terms of interaction (G1: $Z = -2.503$, $p = 0.012$), focus (G2: $Z = -1.981$, $p = 0.048$), confidence (G3: $Z = -2.448$. $p = 0.014$), enjoyment (G4: $Z = -2.625$, $p = 0.009$), satisfaction (G5: $Z = -2.157$, $p = 0.031$), reciprocation (G7: $Z = -2.555$, $p = 0.011$), collaboration (G8: $Z = -2.311$, $p = 0.021$), understanding (G11: $Z = -2.965$, $p = 0.003$), and helpfulness (G12: $Z = -2.539$, $p = 0.011$).

6.5 User preferences

After finishing the assembly task across both interfaces, participants completed the preference questionnaire (see Table 2) to rank the interfaces with respect to six aspects (R1, focus; R2, confidence; R3, satisfaction; R4, collaboration; R5, enjoyment; R6, passion). Figure 18 shows the result of participants' preferences. Whether the participants

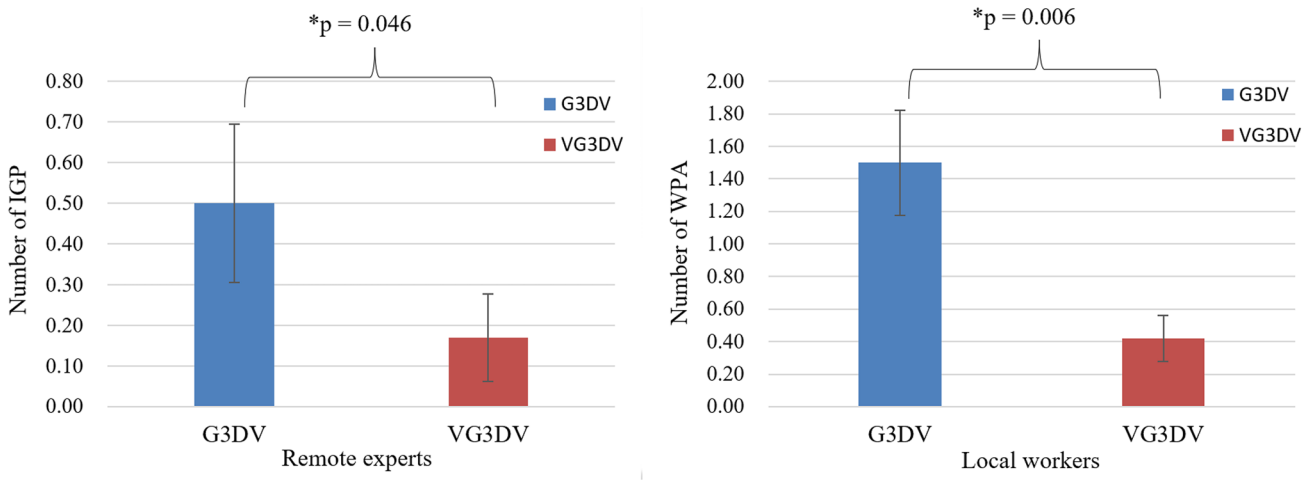


Fig. 14 The IGP and WPA numbers (error bar ± SE)

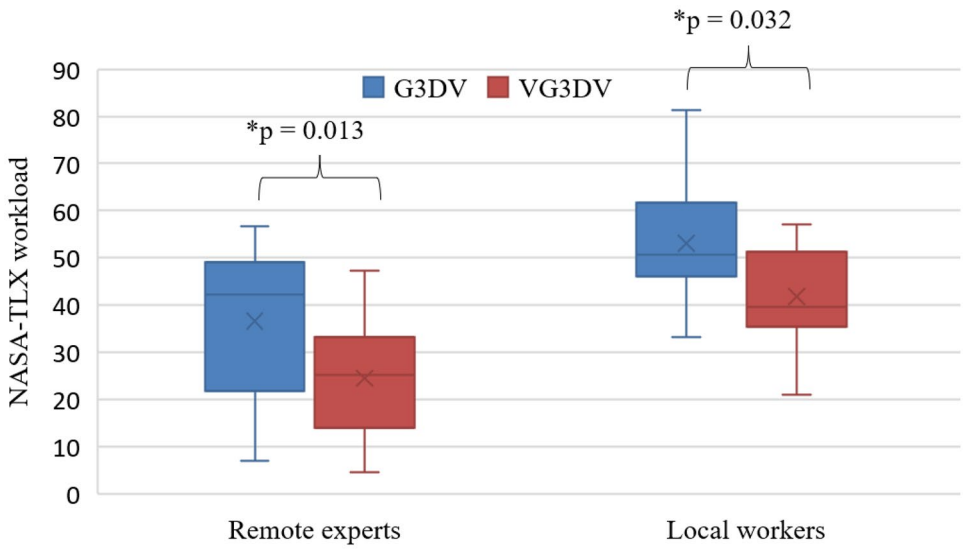
were local workers or remote experts, almost all participants prefer the VG3DV interface to the G3DV interface in all categories.

7 Discussion

In the study, we evaluated the influence of using the combination of gesture cues and virtual replicas in shared real-time 3D video-based MR remote collaboration. We developed two different interfaces (G3DV and VG3DV) and evaluated them through physical industry assembly tasks with respect to five hypotheses, such as performance (H1), error (H2), workload (H3), user experience (H4), and user preferences (H5).

As illustrated in Figs. 13 and 14, as we expected, there were significant differences between G3DV and VG3DV, supported for the hypotheses H1 and H2. The experimental results of the H1 hypothesis and H2 hypothesis confirm each other because when the error operation occurs, remote experts and local workers need more information exchange for correcting it, obviously reducing the efficiency of completing the task. Moreover, specifically, in one assembly step, the remote expert should first find the part for this step and guide the local worker to pick it up, and then instruct the local worker to assemble the parts. However, it is difficult for remote experts to provide clear instructions for these processes using the G3DV interface. As shown in Fig. 19, on one hand, the G3DV interface is limited by the low accuracy 3D reconstructed scene which makes it more difficult for experts to identify parts (see Fig. 19a, b). In contrast, the VG3DV interface can provide virtual proxies which makes it easier for experts to determine the correct part (see Fig. 19c). On the other hand,

Fig. 15 NASA-TLX workload assessment



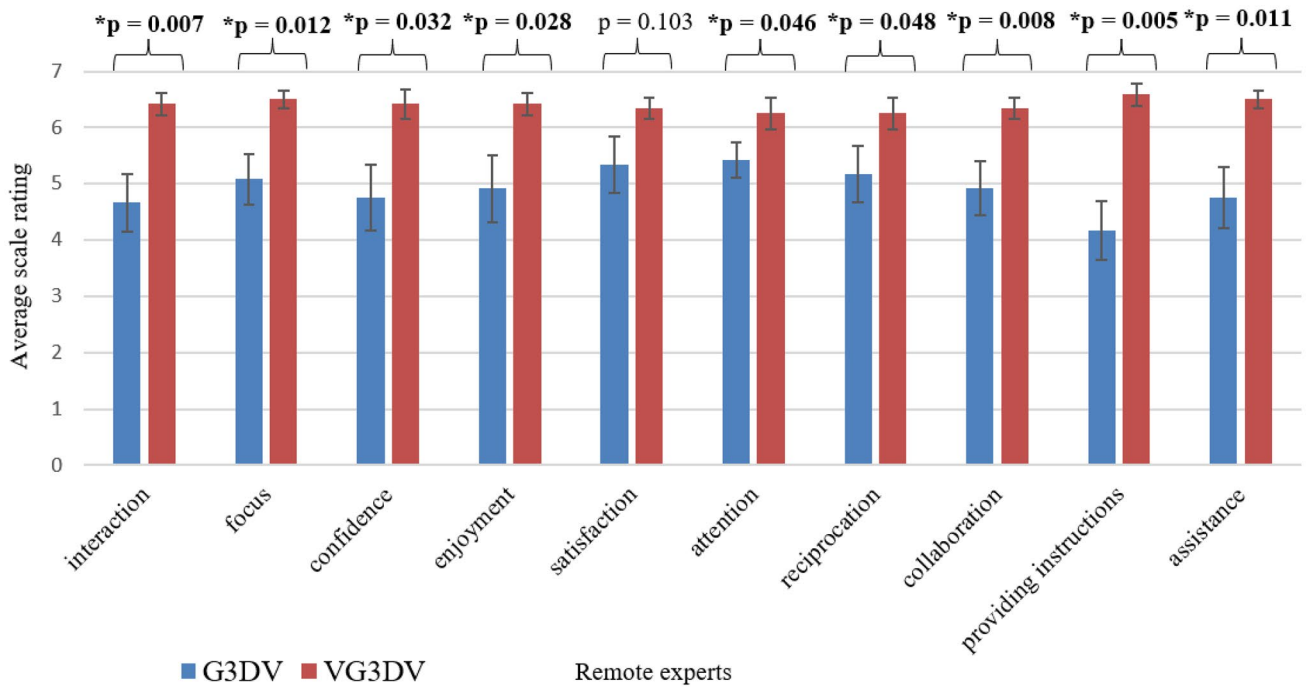


Fig. 16 The results of the average Likert scale rating reported by remote experts on listed items (1: entirely disagree, 7: entirely agree, error bar: ± SE, * represents that there was a significant difference between two conditions)

when parts were clustered densely, remote experts using the G3DV interface was hard to point out the right parts only by gesture (see Fig. 19d, e). In the VG3DV condition, remote experts only need to create the virtual replica of the target part by the GRASP gestures to naturally specify parts (see Fig. 19f).

For local workers, instructions provided by the G3DV interface were sometimes not clear enough. Local workers

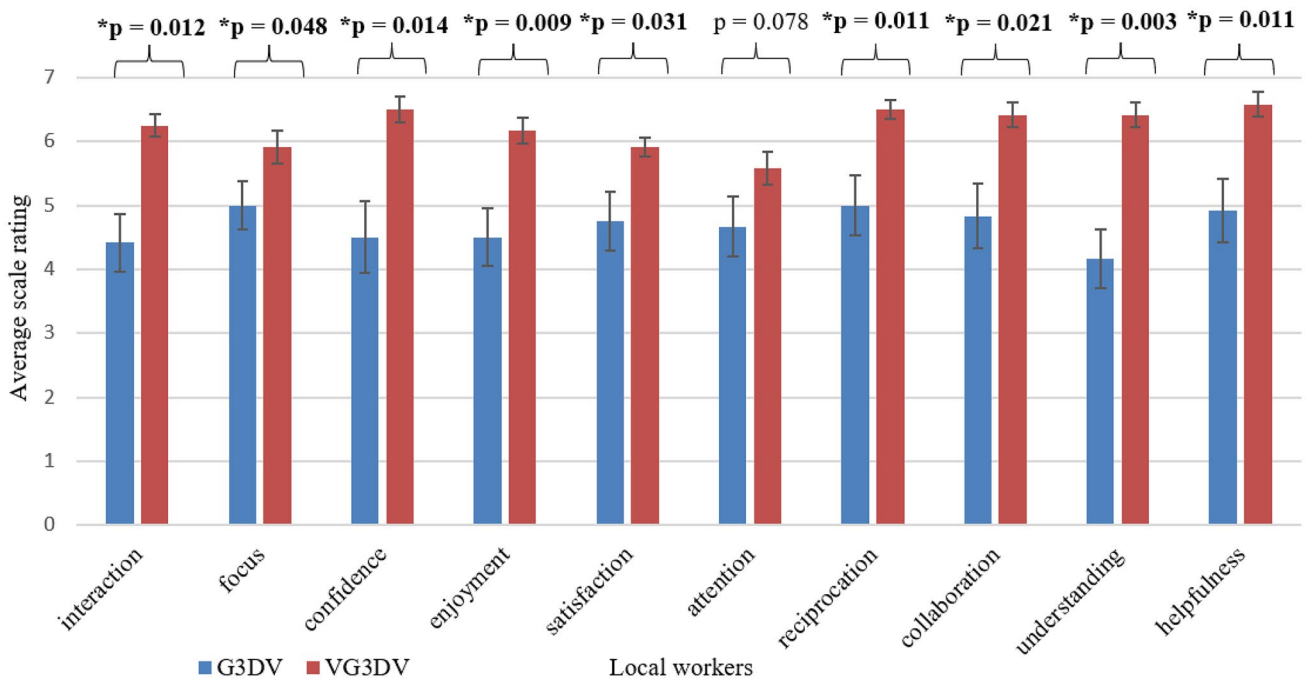


Fig. 17 The results of the average Likert scale rating reported by local workers on listed items (1: entirely disagree, 7: entirely agree, error bar: ± SE, * represents that there was a significant difference between two conditions)

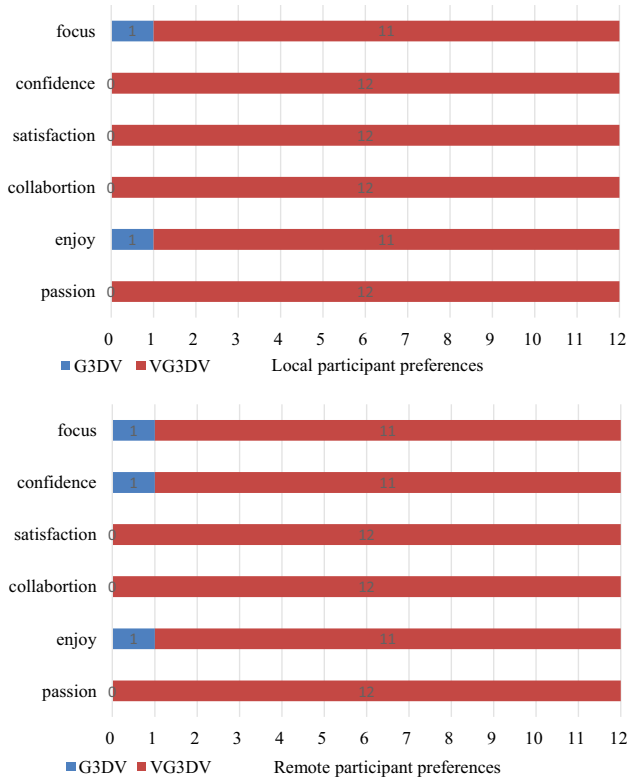


Fig. 18 The users' preference-based ranking results

might need several trial assemblies and many verbal cues exchanges with remote experts to determine the correct assembly method (see Fig. 20). For example, when assembling the long shaft part of the vice, there would be one correct method and the other three wrong assembly methods (see Fig. 20a–d). In such a case, gesture cues from the G3DV interface are hard to reduce the scope of possible assembly methods and local workers needed extra verbal communication with remote experts to confirm (see Fig. 20e, f). On the contrary, the virtual replica from the VG3DV interface can represent the assembly process clearly and intuitively (see Fig. 20g, h). Moreover, in another condition, when assembling parts such as a thread slider that can change the assembly position (see Fig. 21a, b), gesture-based instructions are difficult to indicate accurately where parts should be assembled (see Fig. 21c, e). In contrast, the VG3DV interface provided the virtual replica overlapped on the real parts, so that local workers only need to assemble the real part to overlap the virtual replica (see Fig. 21d, f). Therefore, the VG3DV interface which combines gesture cues and virtual replicas in real-time 3D video-based MR remote collaboration is more natural and easier to create more clear instructions. To some extent, this can explain why the VG3DV interface could significantly improve performance and reduce errors.

The report of the NASA-TLX survey showed significant differences in the workload assessment between the two conditions for both local workers and remote experts which supported hypothesis H3 (see Fig. 15). As we analyzed above, on the VR expert side, the high-precision virtual replicas and natural interactive method of the VG3DV interface reduce the workload of remote experts; and at the same time, on the AR worker side, intuitive and accurate guidance provided by the VG3DV interface reduce the workload of local workers.

We evaluated the user experience by the collaborative experience questionnaire, and the results presented significant differences for all participants in both sites with respect to the feeling of natural and intuitive for interaction (G1), improving the users' focus (G2) and confidence (G3), the sense of enjoyment (G4), the capability of reciprocation between users (G7), and easier collaboration (G8), as illustrated in Figs. 16 and 17. Besides, for the remote participants, using the VG3DV interface could significantly improve the attention to the partner (G6), and the capacity to provide clear instructions (G9) and real-time assistance (G10). Participants as remote experts gave many positive and valuable feedback on the system such as "It is amazing to grasp virtual parts from the 3D stereoscopic scene of the local workspace and I can observe the structure and assembly features of the part to perform more accurate guidance; I think using virtual replicas can naturally and intuitively represent where and how parts should be assembled; Using hand to create and manipulate the virtual replicas allows me to use less descriptive verbal to explain the assembly method; I like this 3D immersive environment for assembly guidance, however, the real-time 3D video is somewhat distorted which makes it difficult for me to distinguish parts, fortunately, the existence of virtual replicas and virtual proxies makes up for this; I can see my partner's real-time action and judge whether he has completed the assembly operation correctly, just as we are in the same space; It is sometimes difficult to grasp small parts with gestures, which requires some practice and skills."

For the local participants, using the VG3DV interface could significantly improve the sense of satisfaction (G5), provide better understand (G11), and more helpful (G12) instructions. Participants as local workers also gave many positive and valuable feedback on the system such as "The shared virtual replicas were created and operated by remote experts make me feel like the expert is right next to me, picking up parts from my workspace and guiding me how to assemble it; I believe that the VG3DV interface can make me understand instructions more easily and reduces my mental workload, because what need I do is only to carry out the assembly operation according to the assembly process demonstrated by the virtual replicas and gestures, rather than trying to assemble multiple times only according to the expert's gesture cues from the G3DV interface and confirming with experts whether the installation is correct; Virtual replicas could provide me more helpful instructions; The

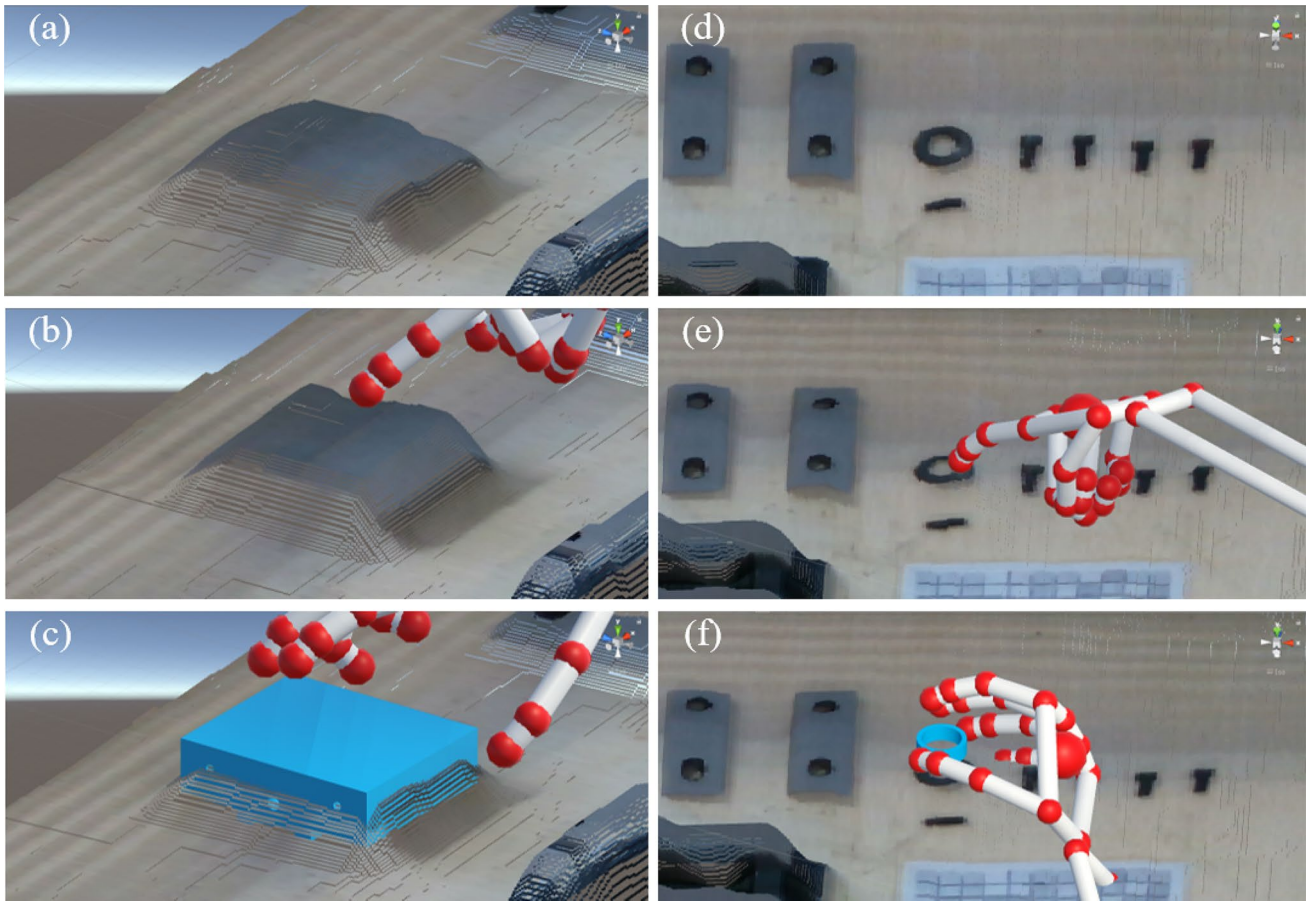


Fig. 19 The remote VR expert creates guidance based on the G3DV or the VG3DV. **a**, **d** The collaborative scene on the remote VR site. **b**, **e** The remote VR expert guiding with G3DV. **c**, **f** The remote VR expert guiding with VG3DV

virtual replicas superimposed on the real part can let me not worry about missing the gesture cues of the remote expert, However, it might block the real parts sometimes, so in this case I need to adjust the observation direction." Therefore, the results of the collaborative experience questionnaire supported hypothesis H4 to a great extent.

For the user preferences, as we expected, most participants preferred the VG3DV interface with respect to the feeling of focus, enjoyment, self-confidence, the degree of passion and satisfaction, and the sense of collaboration (see Fig. 18).

The results also demonstrate that the VG3DV interface has the potential to change the mode of interaction in the shared 3D stereoscopic scene. One of the most interesting findings was in the analysis of the verbal cues observed in conversation records between remote experts and local workers. Augmented feedback with virtual replicas by the VG3DV allowed for less ambiguous and more concise communication as the remote experts were able to direct the local worker without having to describe the spatial position and assembly method. When using the G3DV interface, verbal cues are often very specific such as "pick up the round nut; assemble from left; put the shaft through that hole; tighten until the part reaches the position pointed to by my finger." In contrast, when using the VG3DV interface, remote experts more frequently used terms such as "pick up this; assemble it here."

In terms of gesture cues, the VG3DV interface allowed for more natural and intuition gesture-based instructions as the remote experts were able to manipulate the virtual part as if it was a real part. When using the G3DV interface, experts often used the pointing gesture, more specifically, the static pointing gesture is used to indicate the target part and the continuously pointing gesture is used to indicate where it should be assembled. In contrast, when using the VG3DV interface, the remote experts more frequently utilized the "Grasp" gesture, that is, grabbing a part or a tool to create the virtual replicas to indicate the target part and keeping the "Grasp" gesture to manipulate it until the part reaches the desired position. Besides, some participants in the expert site considered that using the dynamic gesture "SWEEP" to remove virtual replicas is natural and simple, although it can create small distractions at times.

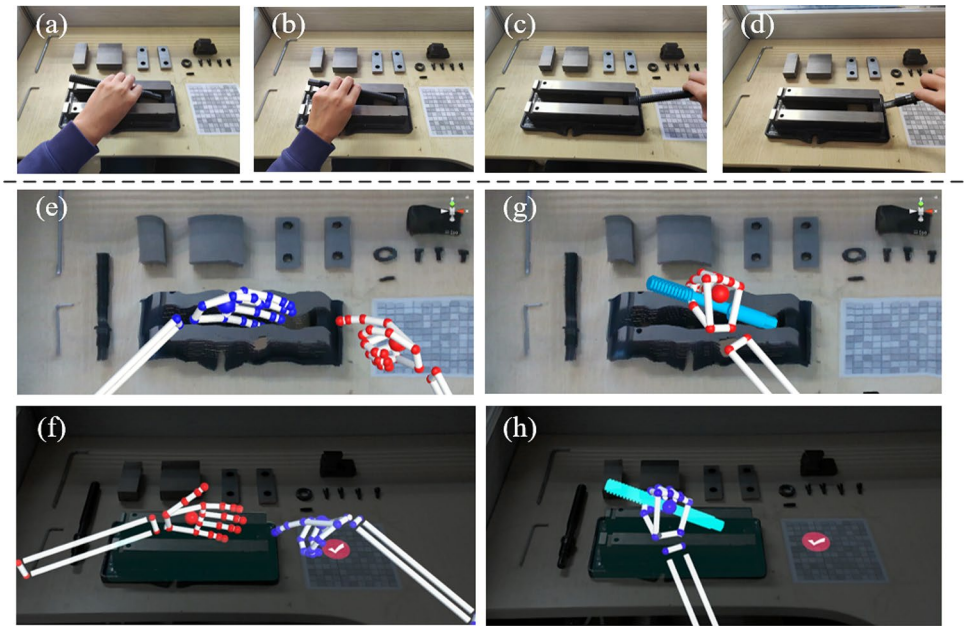


Fig. 20 Guiding the assembly of the long shaft based on the G3DV or the VG3DV. **a** The correct method for the long shaft; **b**–**d** three wrong assembly methods for the long shaft; **e** the VR HMD view based on the G3DV; **f** the AR HMD view based on the G3DV; **g** the VR HMD view based on the VG3DV; **h** the AR HMD view based on the VG3DV

Finally, the research has some implications for the design and development of MR remote collaborative systems based on the shared 3D stereoscopic scene. Combining gesture cues and virtual replicas enables remote experts to interact with objects in the shared 3D stereoscopic scene naturally and intuitively and provide clear instructions, especially in industrial assembly tasks. Therefore, we propose that the MR remote collaboration system for industry tasks should support sharing this combination, and researchers should consider how to design gestures for natural manipulation of virtual replicas as if the manipulation of their physical counterparts.

8 Limitations

Generally, the subjects responded with generally positive and favorable views of the combination of virtual replicas and gesture cues as an innovative interactive method on a shared 3D video MR remote collaboration system and recognized the potential benefits for industrial assembly tasks it can bring. However, there are still certain limitations in our current research which can be further investigated. First, the sample size of statistical data analysis in our present user study is quite modest and the diversity is low (24 participants, 12 pairs, aged from 22 to 34 (M = 28 years, SD = 0.7)). Most of the participants are young college students, so they may be more receptive and favorable to new technologies than actual industry workers (they may older and have only a shorter formal education). We believe that increasing the number of participants with different educational levels and age distribution will help to explore the impact of education and age factors on the performance and user experience of our remote collaboration mode, and will help to translate the research results into practical benefits.

Second, our current research adopts co-located collaboration to simulate remote collaboration. In this case, participants are in the same room and they can speak to each other, which can reduce typical problems of verbal communication in remote

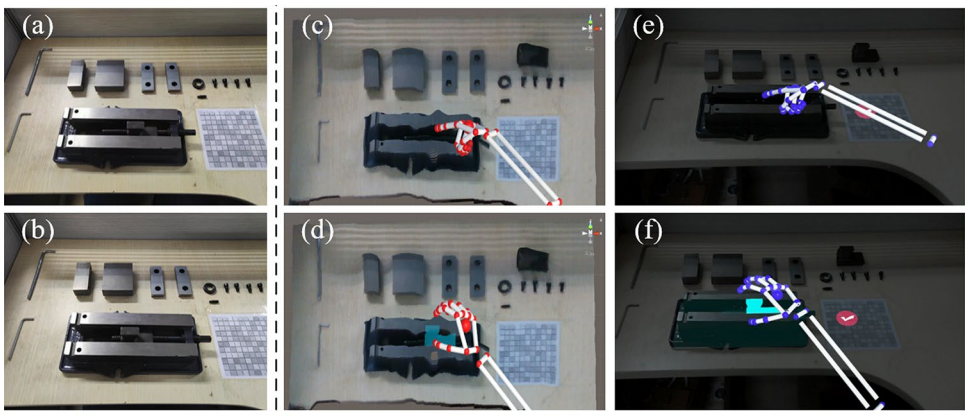


Fig. 21 Guiding the assembly of the thread slider based on the G3DV or the VG3DV. **a** The correct assembly position for the thread slider; **b** a wrong assembly position for the thread slider; **c** the VR HMD view based on the G3DV; **d** the VR HMD view based on the VG3DV; **e** the AR HMD view based on the G3DV; **f** the AR HMD view based on the VG3DV

assistance applications, such as sound distortion, communication delay, and so on. We think that this factor may have an impact on remote collaboration, such as objective results (performance time and error evaluation) or user experience. Therefore, we are trying to improve the prototype based on WebRTC [11] for the geographically separated remote collaboration.

Third, our prototype currently does not support the sharing of other nonverbal cues other than gesture cues and virtual replicas, such as the popular AR annotations. Earlier research [11] reported a 3D + interface which supports remote experts annotating on the 3D reconstructed scene of the local worker side. This study presented that the users seemed to be satisfied with annotations but there was no sign performance improvement in statistical analysis. However, it is worth comparing our method with other existing interactive methods to explore the advantages and limitations of our method.

Fourth, in our current research, the 3D stereoscopic local scene information is captured by one 3D (RGB + D) camera for reliable real-time network data transmission performance. However, even the most advanced 3D (RGB + D) camera lacks the ability to accurately and reliably capture the environment for real-time visualization to fool the eye. Therefore, the 3D video is often affected by noise and artifacts. Furthermore, a single viewpoint is insufficient for obtaining information at the occluding boundaries of a part, specifically, a part which is relatively high (such as a vice base), or a part which has sharp edges (such as such a Lego brick). Fortunately, as we expected, according to the feedback of participants, the virtual replicas and virtual proxies represented by high-precision 3D models can make up for the low accuracy of the 3D reconstructed scene to a certain extent. However, sharing a more accurate real-time 3D reconstruction scene may bring better performance and user experience, which will be our direction of improving remote collaboration.

Finally, in order to make the prototype system smoother, the current research does not continuously track the physical parts on the local site but initialized the location of virtual proxies through the first frame captured by the RGBD camera. A higher-performance real-time object tracking method might contribute to the system stability and help the system adapt to the misoperation of local workers. Therefore, a new real-time object tracking method would be developed to improve our prototype system.

9 Conclusions and future works

In this paper, we presented a study to evaluate the influence of combining 3D gesture cues and virtual replicas for industrial assembly tasks in real-time 3D video-based MR remote collaboration. We developed two different interfaces, our novel method of using the combination of virtual replicas and gesture cues in the 3D video (VG3DV), and a method similar to the popular method currently of using gesture cues in the 3D video (G3DV). First, we conducted a pilot test to evaluate the prototype and then improved it according to the feedback from participants. Next, we performed a formal user study to evaluate the proposed interaction mode with respect to performance time, error, workload, user experience, and user preferences in an industrial assembly task. The results confirm our initial hypothesis that in industrial assembly tasks, using the combination of gesture cues and virtual replicas on shared 3D video-based MR remote collaboration can provide better performance and user experience than the traditional method of using gesture cues only. In VG3DV condition, remote experts were able to more directly and clearly guide the local worker to pick up the correct part and assemble it at the correct positions. Finally, positive feedback from the participants with the VG3DV interface suggests that this interaction mode of MR remote collaboration has potential in industrial assembly tasks.

In future work, we intend to enhance the perception abilities of the MR remote collaborative system to make them more suitable for domain-specific tasks, especially industrial mechanical assembly. Next, we will further enhance our research to compare the method with other existing state-of-the-art methods, such as sharing AR annotations or multimodal interaction. Moreover, it would be interesting to explore the impact of enhancing our prototype by haptic feedback interaction. Finally, we also would like to explore how to improve performance and user experience to take advantage of the full potential of MR remote collaboration based on shared 3D stereoscopic scenes, such as sharing more accurate real-time 3D reconstruction scenes by capturing the local scene with multiple RGBD cameras at the same time.

Acknowledgements We would like to appreciate Professor Shusheng Zhang, Weiping He, and Xiaoliang Bai for their science leadership, and the constructive opinions of Dr. Peng Wang and Dr. Zhuo Wang significantly improved the paper. We also would like to appreciate Yuxiang Yan for donating the vise model used in our research. In addition, we would like to thank Yuxiang Yan and Quan Yu for helping experiment with data collection and preparing some supplementary materials.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Xiangyu Zhang, Yuxiang Yan, and Quan Yu. The first draft of the manuscript was written by Xiangyu Zhang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was partly supported by Defense Industrial Technology Development Program (grant number XXXX2018213A001 and No. XXXX2018205B021), National Key R&D Program of China (grant number 2019YFB1703800, 2021YFB1714900, 2021YFB1716200, and

[11] https://webrtc.org

2020YFB1712503), the Programme of Introducing Talents of Discipline to Universities (111 Project), China (grant number B13044), and the Fundamental Research Funds for the Central Universities, NPU (grant number 3102020gxb003).

Declarations

Competing interests The authors declare no competing interests.

References

1. Villanueva A, Zhu Z, Liu Z, Peppler K, Redick T, Ramani K (2020) Meta-AR-App: an authoring platform for collaborative augmented reality in STEM classrooms. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems 1–14. https://doi.org/10.1145/3313831.3376146
2. Oda O, Elvezio C, Sukan M, Feiner S, Tversky B (2015) Virtual replicas for remote assistance in virtual and augmented reality. Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology 405–415. https://doi.org/10.1145/2807442.2807497
3. Wang P, Bai X, Billinghurst M, Zhang S, Wei S, Xu G, He W, Zhang X, Zhang J (2020) 3DGAM: using 3D gesture and CAD models for training on mixed reality remote collaboration. Multimed Tools Appl 80(20):31059–31084. https://doi.org/10.1007/s11042-020-09731-7
4. Elvezio C, Sukan M, Feiner S, Tversky B (2015) Interactive visualizations for monoscopic eyewear to assist in manually orienting objects in 3D. IEEE Int Symp Mix Augment Real 2015:180–181. https://doi.org/10.1109/ISMAR.2015.54
5. Wang P, Zhang S, Billinghurst M, Bai X, He W, Wang S, Sun M, Zhang S (2019) A comprehensive survey of AR/MR-based co-design in manufacturing. Eng Comp 36:1715–1738. https://doi.org/10.1007/s00366-019-00792-3
6. Talmy L (2000) Semantic conflict and resolution. Toward a cognitive semantics, vol 2. MIT, Cambridge, pp 323–337
7. Heiser J, Tversky B, Silverman M (2004) Sketches for and from collaboration. Vis Spat Reason Des III 3:69–78
8. Piumsomboon T, Dey A, Ens B, Lee G, Billinghurst M (2017) CoVAR: mixed-platform remote collaborative augmented and virtual realities system with shared collaboration cues. IEEE Int Symp Mix Augment Real 2017:218–219. https://doi.org/10.1109/ISMAR-Adjunct.2017.72
9. Teo T, Lawrence L, Lee G, Billinghurst M, Adcock M (2019) Mixed reality remote collaboration combining 360 video and 3D reconstruction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems 1–14. https://doi.org/10.1145/3290605.3300431
10. Venerella J, Franklin T, Sherpa L, Tang H, Zhu Z (2019) Integrating AR and VR for mobile remote collaboration. IEEE Int Symp Mix Augment Real Adjunct 2019:104–108
11. Anton D, Kurillo G, Bajcsy R (2018) User experience and interaction performance in 2D/3D telecollaboration. Future Generation Computer Systems. Futur Gener Comput Syst 82:77–88. https://doi.org/10.1016/j.future.2017.12.055
12. Zillner J, Mendez E, Wagner D (2018) Augmented reality remote collaboration with dense reconstruction. IEEE Int Symp Mix Augment Real Adjunct 2018:18–19. https://doi.org/10.1109/ISMAR-Adjunct.2018.00028
13. Piumsomboon T, Day A, Ens B, Lee Y, Lee G, Billinghurst M (2017) Exploring enhancements for remote mixed reality collaboration. SIGGRAPH Asia 2017 Mobile Graph Interact Appl 1–5. https://doi.org/10.1145/3132787.3139200
14. Huang W, Alem L, Tecchia F, Duh HBL (2018) Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration. J Multimodal User Interfaces 12(2):77–89. https://doi.org/10.1007/s12193-017-0250-2
15. Yang P, Kitahara I, Ohta Y (2015) Remote mixed reality system supporting interactions with virtualized objects. ISMAR. https://doi.org/10.1109/ISMAR.2015.22
16. Takeda Y, Matsuda A, Rekimoto J (2020) 3D human reconstruction from an image for mobile telepresence systems. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops 2020:772–773. https://doi.org/10.1109/VRW50115.2020.00237
17. Huang R, Zhang S, Bai X, Xu C, Huang B (2015) An effective subpart retrieval approach of 3D CAD models for manufacturing process reuse. Comput Ind 67:38–53. https://doi.org/10.1016/j.compind.2014.12.001
18. Zhang Y, Zhang S, Huang R, Huang B, Yang L, Liang J (2021) A deep learning-based approach for machining process route generation. Int J Adv Manuf Technol. https://doi.org/10.1007/s00170-021-07412-9
19. Kritzler M, Murr M, Michahelles F (2016) Remotebob: support of on-site workers via a telepresence remote expert system. Proceedings of the 6th International Conference on the Internet of Things 7–14. https://doi.org/10.1145/2991561.2991571
20. Elvezio C, Sukan M, Oda O, Feiner S, Tversky B (2017) Remote collaboration in AR and VR using virtual replicas. ACM SIGGRAPH 2017 VR Village 1–2. https://doi.org/10.1145/3089269.3089281
21. Stork S, Schuboe A (2010) Human cognition in manual assembly: theories and applications. Adv Eng Inform 24(3):320–328. https://doi.org/10.1016/j.aei.2010.05.010
22. Makris S, Karagiannis P, Koukas S, Matthaiakis AS (2016) Augmented reality system for operator support in human–robot collaborative assembly. CIRP Ann 65(1):61–64. https://doi.org/10.1016/j.cirp.2016.04.038
23. Deshpande A, Kim I (2018) The effects of augmented reality on improving spatial problem solving for object assembly. Adv Eng Inform 38:760–775. https://doi.org/10.1016/j.aei.2018.10.004
24. Rios H, Hincapié M, Caponio A, Mercado E, Mendívil GE (2011) Augmented reality: an advantageous option for complex training and maintenance operations in aeronautic related processes. Int Conf Virtual Mix Real 2011:87–96. https://doi.org/10.1007/978-3-642-22021-0_11
25. Ojer M, Alvarez H, Serrano I, Saiz FA, Barandiaran I, Aguinaga D, Querejeta L, Alejandro D (2020) Projection-based augmented reality assistance for manual electronic component assembly processes. Appl Sci 10(3):796. https://doi.org/10.3390/app10030796
26. De Amicis R, Ceruti A, Francia D, Frizziero L, Simões B (2018) Augmented Reality for virtual user manual. Int J Interact Des Manuf 12(2):689–697. https://doi.org/10.1007/s12008-017-0451-7
27. Funk M, Lischke L, Mayer S, Shirazi SA, Schmidt A (2018) Teach me how! Interactive assembly instructions using demonstration and in-situ projection. Assist Augment 2018:49–73. https://doi.org/10.1007/978-981-10-6404-3_4
28. Neb A, Strieg F (2018) Generation of AR-enhanced assembly instructions based on assembly features. Procedia CIRP 72:1118–1123. https://doi.org/10.1016/j.procir.2018.03.210
29. Bhattacharya B, Winer EH (2019) Augmented reality via expert demonstration authoring (AREDA). Comput Ind 105:61–79. https://doi.org/10.1016/j.compind.2018.04.021
30. Su Y, Rambach J, Minaskan N, Lesur P, Pagani A (2019) Stricker D (2019) Deep multi-state object pose estimation for augmented reality assembly. IEEE Int Symp Mix Augment Real Adjunct 2019:222–227. https://doi.org/10.1109/ISMAR-Adjunct.2019.00-42

31. Chang MML, Nee AYC, Ong SK (2020) Interactive AR-assisted product disassembly sequence planning (ARDIS). Int J Prod Res 58(16):4916–4931. https://doi.org/10.1080/00207543.2020.1730462
32. Dimitropoulos N, Togias T, Zacharaki N, Michalos G, Makris S (2021) Seamless human–robot collaborative assembly using artificial intelligence and wearable devices. Appl Sci 11(12):5699. https://doi.org/10.3390/app11125699
33. Dimitropoulos N, Togias T, Michalos G, Makris S (2021) Operator support in human–robot collaborative environments using AI enhanced wearable devices. Procedia Cirp 97:464–469. https://doi.org/10.1016/j.procir.2020.07.006
34. Adcock M, Anderson S, Thomas B (2013) RemoteFusion: real time depth camera fusion for remote collaboration on physical tasks. Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry 235–242. https://doi.org/10.1145/2534329.2534331
35. Dai A, Nießner M, Zollhöfer M, Lzadi S, Theobalt C (2017) Bundlefusion: real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. ACM Transact Graph 36(4):1–18. https://doi.org/10.1145/3054739
36. Izadi S, Newcombe R, Kim D, Hilliges O, Molyneaux D, Hodges S, Kohli P, Shotton J, Davison A, Fitzgibbon A (2011) KinectFusion: real-time dynamic 3D surface reconstruction and interaction. ACM SIGGRAPH 2011 Talks 1–1. https://doi.org/10.1145/2037826.2037857
37. Orts-Escolano S, Rhemann C, Fanello S, Chang W, Kowdle A, Degtyarev Y, Kim D, Davidson P, Khamis S, Dou M, Tankovich V, Loop C, Cai Q, Chou P, Mennicken S, Valentin J, Pradeep V, Wang S, Kang B, Kohli P, Lutchyn Y, Keskin C, Lzadi S (2016) Holoportation: virtual 3D teleportation in real-time. Proceedings of the 29th Annual Symposium on User Interface Software and Technology 741–754. https://doi.org/10.1145/2984511.2984517
38. Prisacariu VA, Kähler O, Cheng MM, Ren CY, Valentin J, Torr PH, Reid ID, Murray DW (2014) A framework for the volumetric integration of depth images. 1-18 arXiv preprint http://arxiv.org/abs/1410.0925
39. Nuernberger B, Turk M, Höllerer T (2017) Evaluating snapping-to-photos virtual travel interfaces for 3D reconstructed visual reality. Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology 1–11. https://doi.org/10.1145/3139131.3139138
40. 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. https://doi.org/10.1007/s00170-018-03237-1
41. 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. IEEE Third VR Int Workshop Collab Virtual Environ 2016:9–12. https://doi.org/10.1109/3DCVE.2016.7563559
42. Wang P, Bai X, Billinghurst M, Zhang S, He W, Han D, Wang Y, Min H, Lan W, Han S (2020) Using a head pointer or eye gaze: the effect of gaze on spatial AR remote collaboration for physical tasks. Interact Comput 32:153–169. https://doi.org/10.1093/iwcomp/iwaa012
43. Wang P, Zhang S, Bai X, Billinghurst M, Zhang L, Wang S, Han D, Lv H, Yan Y (2019) A gesture-and head-based multimodal interaction platform for MR remote collaboration. Int J Adv Manuf Technol 105(7):3031–3043. https://doi.org/10.1007/s00170-019-04434-2
44. Al-Khafajiy M, Baker T, Chalmers C, Asim M, Kolivand H, Fahim M, Waraich A (2019) Remote health monitoring of elderly through wearable sensors. Multimed Tools Appl 78(17):24681–24706. https://doi.org/10.1007/s11042-018-7134-7
45. García-Pereira I, Portalés C, Gimeno J, Casas S (2020) A collaborative augmented reality annotation tool for the inspection of prefabricated buildings. Multimed Tools Appl 79:6483–6501. https://doi.org/10.1007/s11042-019-08419-x
46. Adcock M, Ranatunga D, Smith R, Thomas B (2014) Object-based touch manipulation for remote guidance of physical tasks. Proceedings of the 2nd ACM Symposium on Spatial User Interaction 113–122. https://doi.org/10.1145/2659766.2659768
47. Tait M, Billinghurst M (2014) View independence in remote collaboration using AR. ISMAR. 309–310.
48. Pierce J, Stearns B, Pausch R (1999) Voodoo dolls: seamless interaction at multiple scales in virtual environments. Proceedings of the 1999 Symposium on Interactive 3D Graphics 141–145.
49. Wang P, Bai X, Billinghurst M, Zhang S, Han D, Lv H, He W, Yan Y, Zhang X, Min H (2019) An MR remote collaborative platform based on 3D CAD models for training in industry. 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) 91–92. https://doi.org/10.1109/ISMAR-Adjunct.2019.00038
50. Fakourfar O, Ta K, Tang R, Bateman S, Tang A (2016) Stabilized annotations for mobile remote assistance. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems 1548–1560. https://doi.org/10.1145/2858036.2858171
51. Kim S, Lee G, Huang W, Kim H, Woo W, Billinghurst M (2019) Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems 1–13. https://doi.org/10.1145/3290605.3300403
52. Wang P, Bai X, Billinghurst M, Zhang S, Han D, Sun M, Wang Z, Han D, Sun M, Wang Z, Lv H, Han S (2020) Haptic feedback helps me? A VR-SAR remote collaborative system with tangible interaction. Int J Hum Comput Interact 36(13):1242–1257. https://doi.org/10.1080/10447318.2020.1732140
53. Gao L, Bai H, He W, Billinghurst M, Lindeman R (2018) Real-time visual representations for mobile mixed reality remote collaboration. SIGGRAPH Asia 2018 Virtual & Augmented Reality 1–2. https://doi.org/10.1145/3275495.3275515
54. Liu M, Tuzel O, Veeraraghavan A, Chellappa R (2010) Fast directional chamfer matching. IEEE Comput Soc Conf Comput Vis Pattern Recognit 2010:1696–1703. https://doi.org/10.1109/CVPR.2010.5539837
55. Brooke J (1996) SUS-a quick and dirty usability scale. Usability Eval Ind 189(194):4–7
56. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Advances in Psychology, Vol. 52. Elsevier, North-Holland, pp 139–183
57. Harms C, Biocca F (2004) Internal consistency and reliability of the networked minds measure of social presence
58. 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):1–14. https://doi.org/10.3389/frobt.2019.00018