ORIGINAL ARTICLE

Can you notice my attention? A novel information vision enhancement method in MR remote collaborative assembly

YuXiang Yan[1] · Xiaoliang Bai[1] · Weiping He[1] · Shuxia Wang[1] · XiangYu Zhang[1] · Peng Wang[2] · Liwei Liu[1] · Bing Zhang[1]

Abstract

In mixed reality (MR) remote collaborative assembly, remote experts can guide local users to complete the assembly of physical tasks by sharing user cues (eye gazes, gestures, etc.) and spatial visual cues (such as AR annotations, virtual replicas). At present, remote experts need to carry out complex operations to transfer information to local users, but the fusion of virtual and real information makes the display of information in the MR collaborative interaction interface appear messy and redundant, and local users sometimes find it difficult to pay attention to the focus of information transferred by experts. Our research aims to simplify the operation of remote experts in MR remote collaborative assembly and to enhance the expression of visual cues that reflect experts' attention, so as to promote the expression and communication of collaborative intention that user has and improve assembly efficiency. We developed a system (EaVAS) through a method that is based on the assembly semantic association model and the expert operation visual enhancement mechanism that integrates gesture, eye gaze, and spatial visual cues. EaVAS can give experts great freedom of operation in MR remote collaborative assembly, so that experts can strengthen the visual expression of the information they want to convey to local users. EaVAS was tested for the first time in an engine physical assembly task. The experimental results show that the EaVAS has better time performance, cognitive performance, and user experience than that of the traditional MR remote collaborative assembly method (3DGAM). Our research results have certain guiding significance for the research of user cognition in MR remote collaborative assembly, which expands the application of MR technology in collaborative assembly tasks.

1 Introduction

With the continuous development of network communication technology and computer technology, augmented reality (AR)/virtual reality (VR)/mixed reality (MR) are playing an increasingly important role in remote cooperative physical tasks due to their unique advantages [1-3]. Mixed reality can seamlessly blend the real world and the virtual world using augmented reality and virtual reality, which enable making remote experts and local users feel that they are in the same cognitive space [2]. Therefore, mixed reality can significantly improve performance and user experience in many remote collaboration scenarios. In some complex production scenarios, such as assembly [4, 5], emergency maintenance [6, 7] and training [8, 9], local operators often face problems when assembling or disassembling parts due to their limited domain knowledge. They lack the experience to identify the causes of the problems and do not know how to solve them correctly and efficiently. Therefore, they need to consult experts for assistance. However, in today's society of the Internet of Things and globalized production, experts are not always present at the production site. Therefore, through a remote collaboration, remote experts can overcome geographical restrictions to help and supervise local users complete production tasks [10]. In the remote collaboration tasks, a key is how to enable remote experts and local users to share the cooperation status of the same space and how to clearly and effectively convey the instructions of remote

✉ Xiaoliang Bai
    bxl_nwpu@163.com

✉ Weiping He
    weiping@nwpu.edu.cn

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

2   School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

experts to local users [11]. The traditional video conference method has been widely used in remote collaboration tasks because it can convey the information of audio and 2D video streams. However, due to the uncertainty of audio information expression and the lack of depth information in the 2D video stream, the collaboration intention among users is difficult to convey in the complex operation task and the lack of a 3D spatial reference environment [1]. MR remote collaboration provides a "human-centered" design space for remote collaboration. MR remote collaboration is a type of computer-supported collaborative work that uses mixed reality technology to enable remote users to interact with each other and with physical objects in a shared virtual environment [12]. MR remote collaboration enables sharing of spatial information cues and collaboration status, as well as conveying and expressing collaboration intentions through human–computer interaction [11]. In specific physical tasks, MR remote collaboration can allow remote and local users to jointly carry out activities in MR space across geographical constraints, even across different time zones and cultures [12, 13]. Different from the traditional video conferencing method, MR remote collaboration can integrate different viewpoints of users in virtual reality and augmented reality, and add visual cues of virtual cues to the real world through MR technology to communicate in a natural and intuitive interactive way [12].

In order to better transfer the knowledge and experience of experts to local users, some MR collaborative pointing and dynamic gestures research uses non-verbal cues to guide communication, such as using pointers, arrows, gestures, eye gaze, virtual avatars, and other user cues [7, 14-18]. These studies have greatly improved the collaboration efficiency, coexistence awareness, user attention, and user collaboration experience in MR collaboration [2, 19]. However, according to Polanyi's paradox, the information that humans can express is far less than their skills and knowledge [20].

The existing way of remote expert transferring information is very cumbersome, and sometimes, some operations of experts are difficult to describe clearly (such as assembly space position, assembly direction, assembly attribute), which is difficult for operators to understand. In addition, in some assembly scenarios, it is difficult for operators to remember the operation of experts. Therefore, in many cases, it is necessary for the expert to describe in words or gesture for many times, so that the operator can understand the operation intention of the expert. We believe that it is necessary to adjust the visual form according to the attention of experts to strengthen the operation of experts and focus on strengthening the information that experts want to express.

Therefore, different from the previous work, our research is from the perspective of user cognition, which provides experts with great freedom of operation in MR Remote Collaborative Assembly, aiming at strengthening the expression of information that experts want to convey to local users. Our research aims to simplify the operation of remote experts in MR remote collaborative assembly and to enhance the expression of visual cues that experts' attention, so as to promote the expression and communication of collaborative intention that user has and improve assembly efficiency.

In addition, in the field of manufacturing and assembly, 3D CAD design system plays an important role in the assembly process [21]. 3D CAD models of most manufactured parts are stored in the repository [22]. In MR remote collaboration, 3D CAD model (3D virtual replicas) relieves the burden of remote experts to express information to a certain extent due to its intuitive spatial visual information expression [23-25]. Furthermore, some researches integrate visual forms such as gestures, eye gaze, virtual avatars, and 3D CAD to share information [8]. It can not only make use of the advantages of user cues such as shared gestures to achieve more intuitive and expressive interaction, but also reuse the existing 3D CAD models in industry to express information [8, 26]. To simplify the operation of experts, Wang et al. [27] developed a new adaptive MR remote cooperation architecture, which enables remote experts to simplify the demonstration task of guiding user operations. Remote users can activate instructions through simple and intuitive interaction, and then display clear instructions in MR (local) and VR (remote) views, so that local workers can operate tools according to these instructions.

Inspired by these research works, our research is based on these previous works. Our method can not only share gaze information, gesture information, and spatial visual information in MR remote collaborative assembly, but also sense the expert's attention through hand-eye collaborative interaction and adjust the visual form of information to visually enhance the expert's operation. By strengthening the expression of expert information, our research enables users to focus on the information that experts want to express, thus simplifying the operation of experts and strengthening the cognition of local users. Compared with previous research work, our research makes the following novel contributions:

- Proposing an information vision enhancement method based on expert attention for the first time in MR remote collaborative assembly, which senses expert behavior through hand eye interaction so that experts can control the expression of information to convey important information.
- Designing an information hierarchy division method based on assembly semantic association model in MR assembly.
- Implementing a remote collaboration system (EaVAS) based on expert attention visual enhancement. EaVAS supports multimodal data fusion information cues combining hand-eye user cues and virtual replica space cues

and considers expert attention to adjust the visual form of assembly guidance information.

- Exploring the impact of enhanced visual information based on expert attention on users in MR remote collaborative assembly tasks.

The experimental evaluation shows that our method is feasible. In the engine assembly task, compared with the traditional MR remote collaborative assembly, our system improves the assembly efficiency and significantly improves the user's attention, confidence, focus, and user experience in the collaborative assembly task.

In the rest of this paper, we first review the previous relevant research work, and then describe our system, mainly focusing on the hierarchical design of assembly process information based on assembly semantic association model and the visual presentation of expert attention in MR remote collaborative assembly. Next, we design an engine co-assembly user study experiment and discuss the experimental results we found. Finally, we draw some conclusions and look forward to future research work.

2 Related work

In this section, we will review the methods of MR remote collaboration and the sharing of gesture and eye gaze user cues, spatial visual cues, and multimodal data fusion information cues in MR collaboration. Previous studies have explored two main methods of remote collaboration: traditional video/audio-mediated communication and mixed-reality communication based on sharing MR cues [12]. Section 2.1 provides a detailed comparison of the two methods of remote collaboration. Sections 2.2, 2.3, and 2.4 introduce various MR communication cues and their benefits for remote collaboration. From previous studies, we found that few studies focused on the impact of information control and attention perception in MR remote collaboration on user cognition. Our work combines and extends earlier research on MR remote collaboration and enhances the presentation of information to explore the impact on user cognition through the information vision enhancement method based on expert attention.

2.1 MR remote collaboration

Previous studies of traditional remote collaboration methods focused on sharing voice and video cues through telephone and video conferencing [28, 29]. Traditional telephone and video conference methods have enhanced remote collaboration by being economical and convenient in the context of modern society and communication technology [29, 30]. Traditional remote collaborative technology has limitations in providing visual information. It cannot fuse the physical task space and the enhanced communication space with virtual and real information. As a result, some important nonverbal cues in remote collaborative work are lost, such as gestures, gaze, and depth perception of the task environment [1]. As science, technology, and information progress, MR-based remote collaboration becomes more important for remote collaboration of physical tasks. Mixed reality remote collaboration technology differs from traditional voice and video conferencing in that it integrates real and virtual environments and objects. This provides a richer and more natural way of interaction and improves user experience and task performance by sharing MR non-verbal communication cues (pointing, annotation, gaze, gesture, empathy, etc.) [31-34]. Compared to traditional remote collaboration technology, which often fails to simulate the real spatial relationship and causes isolation and communication barriers, mixed reality remote collaboration technology can enhance the cooperation effect by making participants feel a stronger sense of presence and co-presence in a shared space [12]. In addition, mixed reality remote collaboration technology enables participants to switch between different perspectives and roles flexibly, which improves collaboration efficiency and quality [12].

MR remote collaboration technology can be applied in many remote physical scenarios. These scenarios are asymmetric, as remote experts with knowledge and experience collaborate with local users who have physical tools and better workspace to complete tasks [7, 19, 35]. Remote experts need to pass complex operation instructions to local users to guide them to operate tools to complete tasks. Remote experts can provide effective instructions to improve remote collaboration performance by adding AR annotations [1, 36] or sharing gaze and gesture cues [37, 38] on the shared view of the task space. This can effectively reduce the user's response time and mental workload in many application scenarios (such as manufacturing, assembly and telemedicine, remote education).

Choi [39] et al. proposed a context-based MR remote collaboration method, which can provide more effective AR space for remote collaboration. Their AR collaboration based on real-time video with synchronous VR mode can provide more effective and accurate 3D annotation by synchronizing virtual objects with physical objects. Wang [19] proposed a gesture-based MR remote collaboration platform, which can project the gestures of remote experts into the real workspace of local users to improve performance, co-presence awareness, and user collaboration experience. Lee et al. [40] developed a prototype system that shares gaze cues between remote experts and local users. Their experimental results show that sharing gaze cues improves focus awareness and collaboration experience. MR remote collaboration is a powerful and intuitive way for remote

experts. They can use various visual cues to provide real-time help to users who face operational difficulties. However, it is not easy for local users to focus on the remote experts' attention in the mixed-reality workspace, which may hinder their understanding of the experts' operation intention in complex task collaboration. To the best of our knowledge, few studies have focused on how information control and attentional cognition affect users in remote collaboration. Different from the previous methods, we aim to enhance the expert's information control and expression by using attention-based cues from a cognitive perspective, which can help local users focus on the information that the expert wants to convey, thus simplifying the expert's operation and improving the local user's cognition. Next, we will review the use of visual cues in MR remote collaboration from the following three aspects.

2.2 Presenting gesture and eye gaze cues

In remote collaboration, non-verbal cues such as human body language can convey a lot of information. With the rapid development of eye tracker, Kinnect somatosensory sensor, Leapmotion gesture recognition and other human detection devices, user-centered body language cues (such as eye gaze, head points, virtual avatars, gestures) can provide natural and intuitive visual information in remote collaboration tasks [41]. As one of the most used human body languages, gestures can express the interaction intentions of remote experts through natural interaction such as finger pointing and dynamic gestures [33, 42-44]. Gestures play an important role in many fields, such as scientific research and commercial applications, and have become a pervasive technology in the field of collaborative cooperation [19]. Li et al. [45] demonstrated that incorporating gesture information in remote collaboration not only enhances task performance but also improves user experience. Kiek et al. [46] found that gesture interaction can affect natural collaboration performance and the grounding process in remote collaboration. In order to reduce the distraction of users between gesture instructions and shared 2D videos, Wang et al. [47] proposed to project the gestures of remote experts to the real work site, which greatly improved performance, coexistence awareness, and user collaboration experience.

In addition, eye tracking, as an attention agent for specific AR information, can tell us what we are interested in [48]. Fixation is the basic output measure of interest, which can show what the eye is looking at and select virtual elements. The gaze point can be sensed through a sensor to dynamically track the intention and state of the expert. In face-to-face MR collaboration, eye gaze is an important communication cue, especially for the focus of attention [2]. When collaborating on a

physical task, providing information that expresses the viewpoint indicating where the expert is looking is more important than providing convincing face-to-face eye contact [49, 50]. Research has shown that gaze cues can increase the sense of co-existence of collaborators [51] and are implicit pointers to promote communication [51, 52]. Gaze cues can enhance the performance of visual search tasks and enable operators to capture the focus of experts' eyes [53, 54].

Furthermore, in order to synthesize the advantages of gesture and eye gaze in MR collaboration, some researchers have proposed methods for hand-eye collaborative interaction [55]. Wang et al. [19] created a remote collaboration system 2.5DHANDS, which utilizes virtual reality and spatial augmented reality to support remote experts to provide guidance through instructions based on gestures and gaze cues for pump assembly tasks. By incorporating user cues for gestures and eye gaze, the researchers' systems significantly improved assembly efficiency and collaboration experience by increasing attention and reducing errors [8]. Piumsomboon et al. [2] explored the effect of different combinations of three non-verbal cues (head/eye gaze, gesture, frustum) on the object search task in VR/AR interfaces. They found that displaying a user's eye gaze and frustum significantly improved user performance and preferences. Recently, Bai et al. [56] proposed a MR remote collaboration system that shares users' eye gaze, gestures, and other cues, with a real-time 3D panorama of their surroundings as one of the shared cues. They found that by combining eye gaze and gestural cues, the remote collaboration system was able to provide a strong sense of co-presence between experts and local users in spatial communication compared to using individual eye gaze cues. However, we have found that the existing methods of MR remote collaboration that rely solely on gesture cues and gaze cues have limited time to attract users' attention in complex assembly environments. When remote experts guide users, they may need to repeat gesture and gaze guidance tirelessly to ensure that they attract the attention of local users. This increases the operational burden of experts and the cognitive load of local users to some extent. Similar to the previously mentioned methods, our system also uses user cues combining gestures and eye gaze to enhance information exchange between remote experts and local users. Different from previous studies, our method provides remote experts with significant information control interaction space while sharing gesture and gaze cues. Remote experts can freely control the display form of virtual element information in the VR workspace to enhance the key operational information they want to convey to local users. This helps to continuously attract local users' attention and improve their cognition.

2.3 Presenting spatial visual cues

According to the classification of Ref. [12], non-verbal cues mainly include spatial visual cues such as AR annotations, cursor pointer, and virtual replicas or physical proxy in addition to user cues such as gestures, eye gaze, and virtual avatars. Previous research work [24, 57, 58] has shown the importance of AR annotations cues, such as virtual pointers or markers, in supporting effective communication. Remote experts can effectively improve task performance in collaborative systems and reduce the task response time and mental workload of users by sharing AR annotations or a cursor pointer on the shared task space view [1, 36]. Although AR display marks or mouse cursors can enhance the visual expression in remote collaborative assembly, the presentation of assembly guidance information is arbitrary, and the accuracy of collaborative intent expression needs improvement. To avoid miscommunication, some researchers have studied interaction and visualization techniques in AR/MR remote collaboration, using 3D virtual replicas for maintenance/assembly tasks [24, 25]. Elvezio et al. [23, 24] developed an AR remote assistance system where a remote expert can use a 3D CAD virtual replica to provide guidance for 6DOF alignment operations of an aircraft engine combustion chamber. Their findings suggest that 3D CAD virtual replicas can improve user efficiency in remote collaboration and reduce error-prone interactions. Kritzler et al. [14] created the RemoteBob system to allow remote experts to use 3D virtual replicas and AR annotations to provide instructions to operators on site in order to avoid miscommunication and reduce errors. Sukan et al. [25, 59] demonstrated a new interaction and visualization method in which a remote expert provides real-time guidance by controlling the rotation operation of a 3D CAD model through a handle, and they found that using the clearly visible rotation provided by the 3D CAD model can make it easier for users to understand the operation. For the assembly industry, most 3D CAD models of components used for assembly are stored in repositories [21, 22], so 3D-CAD models of parts are available to developers. Therefore, building on previous research, this paper introduces a 3D CAD virtual replica in MR remote collaboration to assist remote experts in focusing on expressing the information they want to convey. This approach can help to better complete the task of guiding assembly. However, in a complex assembly environment, local users in MR space may not always be able to pay attention to the designated virtual replica information operated by remote experts. This is due to the interference of complex information from the fusion of real physical task assembly sites and various complex 3D virtual replicas of assembly parts. Different from previous studies, our study enables remote experts

to freely control the visualization form of the 3D virtual replica through hand-eye interaction to attract the user's attention.

2.4 Presenting multimodal data fusion information cues

From the above research, it can be seen that both user cues (gestures, eye gaze, virtual avatars, etc.) and spatial visual cues (AR annotations, virtual replicas, etc.) can improve the performance of remote collaboration tasks and user experience in MR remote collaboration. However, the clarity and accuracy of the expression of single-modal visual information cues in long-distance collaboration still needs to be improved, and the expression of assembly guidance information largely relies on the description of voice communication. Previous studies [8, 27, 47] have demonstrated that multimodal data fusion of information cues, combining user cues and spatial visual cues, exhibits fast, accurate, and rich visual expressiveness. The fusion of the two visual cues can comprehensively utilize their respective advantages in the triggering of commands, the selection of virtual objects, and the expression of information. Oda et al. [24] developed a remote collaboration system in which VR expert users can use gestures to point and manipulate virtual objects to help AR users with object assembly tasks. Ref. [41] described a collaborative assembly platform called SHARIDEAS, which integrated user cues and scene cues (objects, tools, and spaces) through a generalized gray correlation method. This system can infer the operator's working intention and display the information in an appropriate visual way to intuitively guide the local operator to assemble. However, this system was aimed at human–machine cooperation and ignored the impact of experts' experience and knowledge on local operators. Wang et al. [8] explored combining gesture cues and graphics in complementary ways, enabling remote experts in virtual reality to provide guidance to local workers based on 3D gestures and CAD models. The results showed that the combination of 3D gestures and CAD models had great potential in assembly training. On the basis of this research, Zhang et al. [26] took the real-time 3D panorama of their surrounding environment as one of the shared cues. Their system combined 3D gestures and CAD models with the real-time 3D panorama of the surrounding environment. This integration enabled remote experts to interact with the 3D model in the real environment, which greatly improved their guidance ability. Multi-modal data fusion information cues expand the representation of information, which is more natural and efficient than traditional single-channel interaction methods. The visual presentation of information cues from multimodal data fusion can improve the freedom of collaborative intent expression. Unlike previous studies that only fuse multimodal visual cues, our study considers

expert attention from the perspective of information cognition. Our method allows remote experts to freely control the display form of information through hand-eye interaction, thereby enhancing the key information experts want to convey to local users. This is achieved based on multimodal data fusion information cues that combine hand-eye user cues and virtual replicated spatial cues, which attract user attention and enhance the expression of collaborative intention.

2.5 Summary

From the research discussed above, it can be seen that hand eye collaborative interaction can help remote experts to trigger commands quickly and accurately and select virtual objects. The user cues of hands and eyes have rich visual expression, which can enhance the sense of co-existence of experts and users. In addition, remote experts can provide instructions to operators on site to avoid miscommunication and reduce errors by using 3D virtual replicas and AR annotated spatial visual cues. The multi-mode data fusion information cues, which combine user cues and spatial visual cues, expand the expression of information and can enhance the freedom of information expression of remote experts. However, previous studies have not considered the focus of information that remote experts want to express from the perspective of information cognition, nor have they studied how to visually enhance the operation of the remote

expert to improve information cognition in MR remote collaboration. From the perspective of cognition, this research attempts to establish visual information adjustment rules by perceiving the hand-eye interaction behavior of remote experts on the basis of multi-modal data fusion information cues combining hand-eye user cues and virtual replica space cues, so that remote experts can freely control the display form of information. The purpose of this study is to reduce the burden of experts in information exchange and improve users' cognition by enhancing the key information that remote experts want to express to users.

3 Prototype system

In this section, we present the structure and implementation detail of our Expert-attention Vision Augmentation System (EaVAS). EaVAS takes into account the attention of the expert to adjust the visualization of the assembly guidance information to enhance the key information the expert wants to convey to the local user. We developed EaVAS through a method that is based on the assembly semantic association model and the expert operation visual enhancement mechanism that integrates gesture, eye gaze, and spatial visual cues. Our system is mainly composed of four modules: assembly process information hierarchy module, expert attention perception module, data processing module,
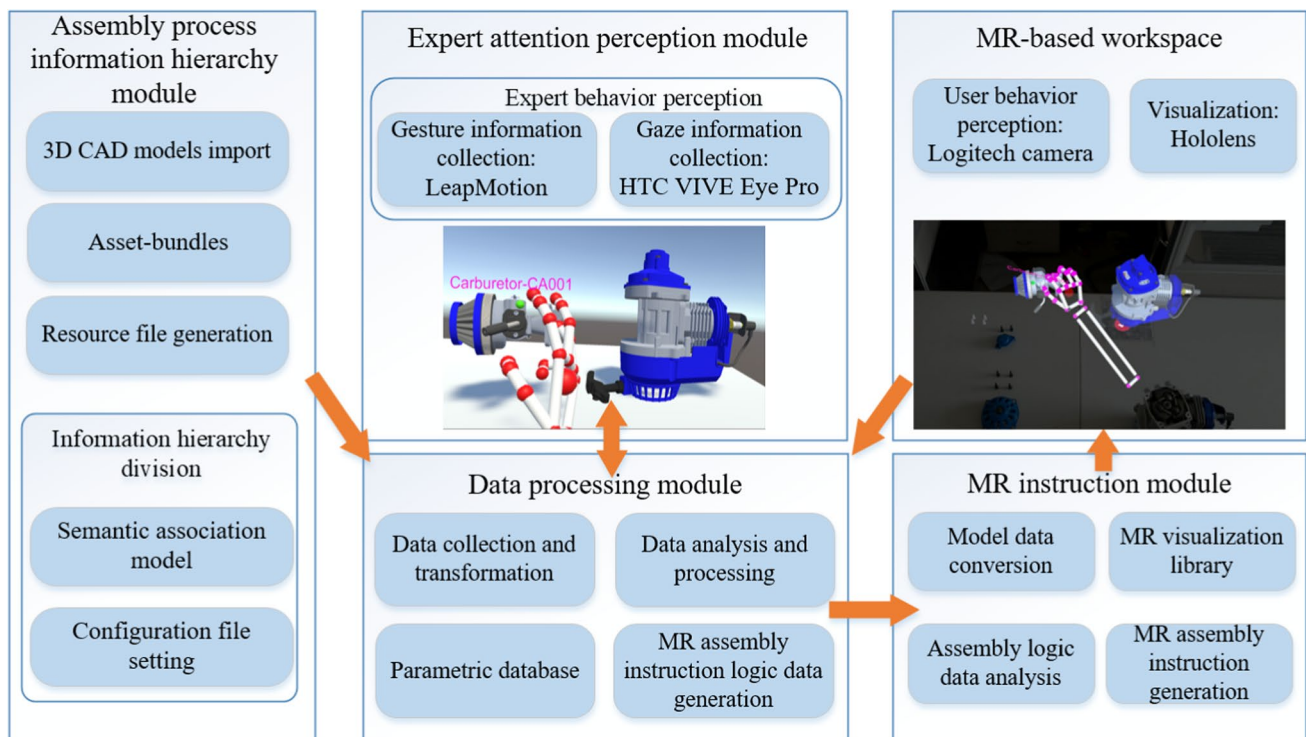

Fig. 1 Expert-attention Vision Augmentation System module

and MR instruction visualization module (see Fig. 1). In the assembly process information hierarchy module, we have designed a new interface so that experts can divide the assembly process information according to the information hierarchy method based on the semantic association model. Experts can use gestures to control the visibility of different process information levels and the degree of information display to convey key information. The expert attention perception module can perceive the expert's gaze information and gesture information and analyze the expert's hand-eye interaction behavior to trigger MR visual assembly instructions. Remote experts can share gesture and gaze cues as well as 3D virtual replica copy spatial cues to conduct visual expression of attention. In addition, our system can enable experts to visually enhance their operations by providing an enhanced representation of assembly process information and important operational behaviors, as well as an adaptive visual presentation of operational details. The data processing module can perform data processing and share data between the remote VR side and the local MR side. The MR instruction visualization module is located on the local MR side to generate MR assembly instructions that can change the visualization form of the assembly process. We will focus on explaining the functions of each module and the presentation of expert attention and the implementation process of EaVAS (see Fig. 2).

3.1 System architecture

3.1.1 Assembly process information hierarchy module

In our system, in order to allow experts to express the important information that experts want to convey to local users among the numerous process information, we provide a process information editing client as a platform for expert to edit and hierarchically divide assembly process information. This client uses Dell Alienware 17 (ALW17C-D2758) laptop as hardware. It uses NVIDIA GeForce GTX 1070 graphics card and Corei7 7700HQ 2.8 Ghz CPU processor and provides WIFI connection function. The expert can import the 3D CAD model of the assembly into the game engine Unity 3D to generate a prefab and edit the process
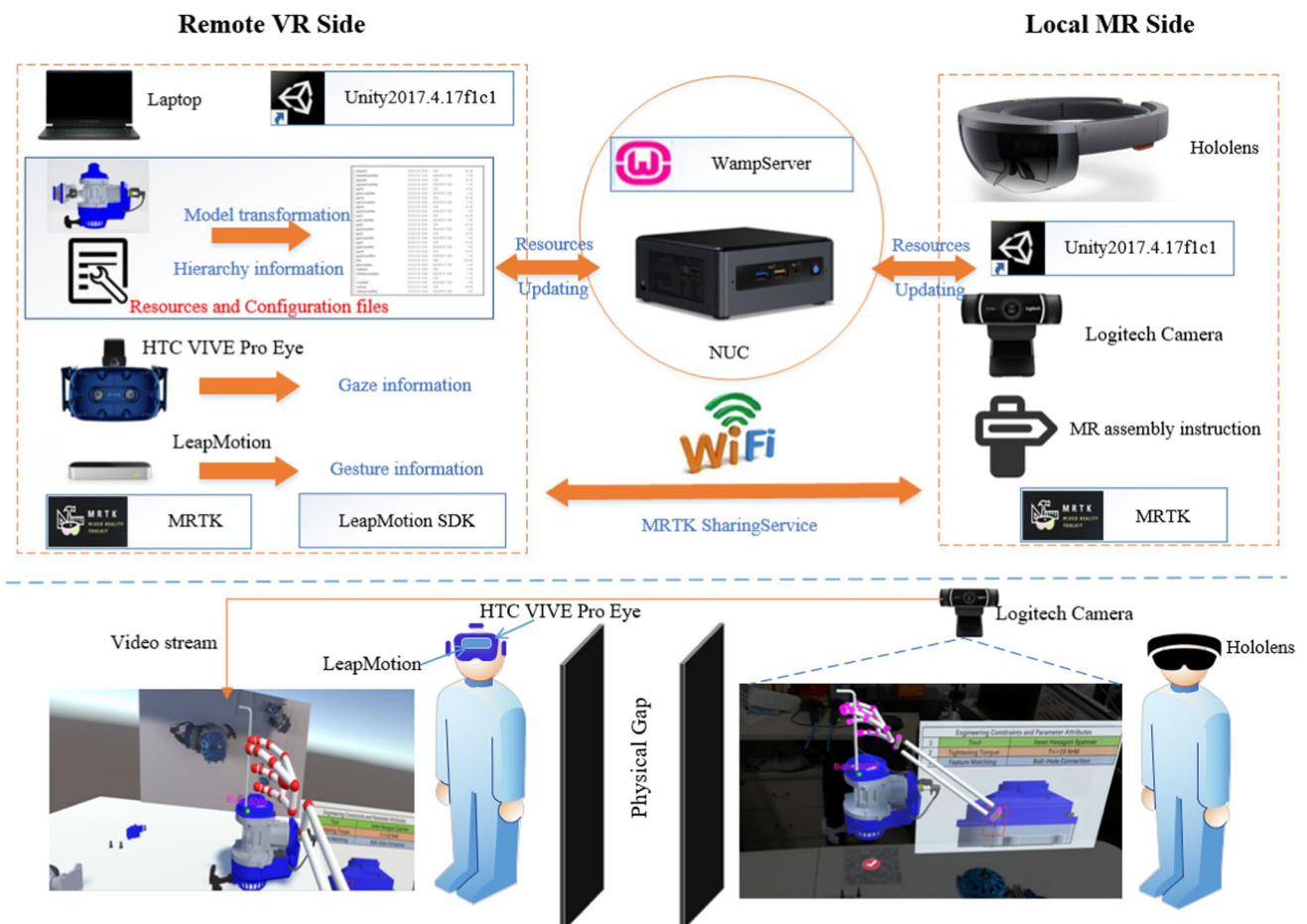

Fig. 2 Expert-attention Vision Augmentation System workflow

information and generate Asset-bundle resource files in Unity3D. We designed a new interface so that the expert can divide the information according to the information level division method based on the semantic association model. See Sect. 3.2.1 for details on the information hierarchy division method based on the assembly semantic association model. In addition, the expert can define the configuration file according to the information classification designed based on the semantic association model, including the logical data of the assembly process. In Unity3D, our system parses the assembly logic data into an XML format file, which converts the assembly logic data into tree hierarchical data supporting MR assembly instructions, and then packs it together with the resource file and submits it to the data processing module through WiFi. For details of our work, please refer to [60]. Different from the previous work of our team, we have designed a new interface so that the expert can divide the assembly process information according to the information hierarchy division method based on the semantic association model.

3.1.2 Expert attention perception module

The expert attention perception module is the core of EaVAS to obtain expert behavior information from collaborative assembly scenarios. The expert attention perception module is located at the remote VR client. It consists of HTC VIVE Pro Eye Kit VR display, LeapMotion, and computer processor. The computer uses Dell Alienware 17 (ALW17C-D2758) laptop. At the remote VR client, experts can observe the operation behavior of local MR client users in the form of video streams. The expert's behavior data collection is mainly realized through HTC VIVE Pro Eye Kit VR display and LeapMotion. The HTC VIVE Pro Eye Kit VR display mainly collects the expert's eye gaze information, while the LeapMotion mainly collects the expert's gesture behavior information. The expert attention perception module perceives the expert attention through the collected eye gaze data information and gesture data information of the expert and the interaction behavior analysis of hand-eye collaboration to trigger the change of MR visualization instructions. The presentation of expert attention is described in detail in Sect. 3.2.2. It is worth noting that the behavior data of the expert's eye gaze and gesture collected by the expert attention perception module will be transmitted to the system data processing module for data processing and analysis.

3.1.3 Data processing module

The data processing module is located on the client side of the work data server. It contains some parameter data library. It is mainly responsible for the communication and data processing and data sharing between the remote VR client and the local MR client. It contains MR assembly instruction logic data that can be used to generate and visualize MR assembly instructions. The hardware of the server client is an Intel NUC-7I7BNH microcomputer, which uses an Intel ceroi7 7567u 3.5 GHz CPU and an Intel GMA HD 650 graphics card, and has the WiFi network connection function. The server can receive eye gaze data information from experts collected by HTC VIVE Pro Eye Kit VR display and gesture information collected by LeapMotion. The data processing module analyzes and processes the collected eye gaze information, gesture recognition, and other data. It associates this information with the CAD models, assembly process hierarchy information, and other information in the assembly process information hierarchy module according to certain rules. This MR assembly instruction work logic data that can change the visual form of virtual elements. The main function of the data processing module is to use these information for association and data analysis and processing to form a set of MR assembly instruction working logic that supports changing the visual form of assembly process information, and send it to the remote expert attention perception module and MR instruction visualization module together with the resource files and configuration files of parts through WiFi using WampServer. In addition, it can also transmit the collected state information of the user of the local MR client and the real-time data of the assembly scene to the remote VR client. Experts then adjust the guidance mode in real time according to the assembly status of local users, so as to realize the closed-loop sharing of data among various clients.

3.1.4 MR instruction visualization module

The MR instruction visualization module is located on the local MR client. Our team used Hololens as a local MR client display due to its wearable portability and good 3D visual display functions. In addition, our team chose Logitech camera as the data collection hardware of the local MR client to collect the status information of local users and the video stream data of assembly scenes. After receiving the MR assembly instruction generation logic, the part resource file, and the configuration file from the server client, the MR instruction visualization module parses them into MR assembly instructions that can change the visual form of assembly process information. Also, it should be noted that the detailed procedures for virtual-real registration and calibration were performed using the process described by Piumsomboon et al. [61].

3.2 Presentation of remote expert attention

3.2.1 Assembly process information hierarchy

The purpose of hierarchical division of assembly process information is to facilitate the remote expert to focus on

expressing the important information he wants to convey to local users. Our team has utilized the assembly semantic association model-based information hierarchy division method to categorize the assembly process information into various levels of information. The assembly semantic association model is shown in Fig. 3. Assembly semantics is an abstract description of the assembly relationship and assembly process information between assembly features in an assembly, such as assembly fit relationships, assembly hierarchy, assembly action, assembly sequence, assembly rules, and parameters (including dimensions) [62]. Assembly semantics has the simplicity of expression, which is closer to the habit of engineers to communicate design ideas [62]. It is more suitable for designers to express assembly intentions through virtual reality interaction methods (such as gesture or gaze) in a virtual environment [63]. Given the extensive use of 3D CAD in current MR assembly, the assembly semantic association model incorporates the spatial position information of 3D CAD models in the virtual world, as well as the associated objects and assembly types during assembly. In addition, through research and investigation, we found that there are two types of constraints between the parts to be assembled: positioning constraints and engineering constraints. Detailed definitions and explanations are as follows:

1. Spatial position information: the final assembly position of the current assembly parts in the virtual world, mainly used for coarse positioning between parts.
2. Associated objects and assembly types: the parts associated with the current assembly part (bolts, pins, keys, etc.) and assembly types (such as clearance fit, transition fit, interference fit).
3. Positioning constraints: mainly used for geometrically accurate positioning between parts.
4. Engineering constraints: mainly include the matching relationship of assembly features between mating parts (hole-shaft fit, etc.), assembly precautions (tools, etc.) of parts, and assembly parameter attributes.

Therefore, this paper defines the assembly semantic association model as an abstract expression of the assembly relationship between parts, which contains the spatial position information of parts, associated objects and assembly types, and the positioning constraints and engineering constraints between assembly parts. According to the assembly semantic association model, the expert uses the interface set by our system in Unity 3D to divide the process information of the currently assembled parts into different levels of information by setting different labels. Remote experts can display the spatial position, associated objects and assembly types, positioning constraints, engineering constraints, and other information of the current assembly parts through different
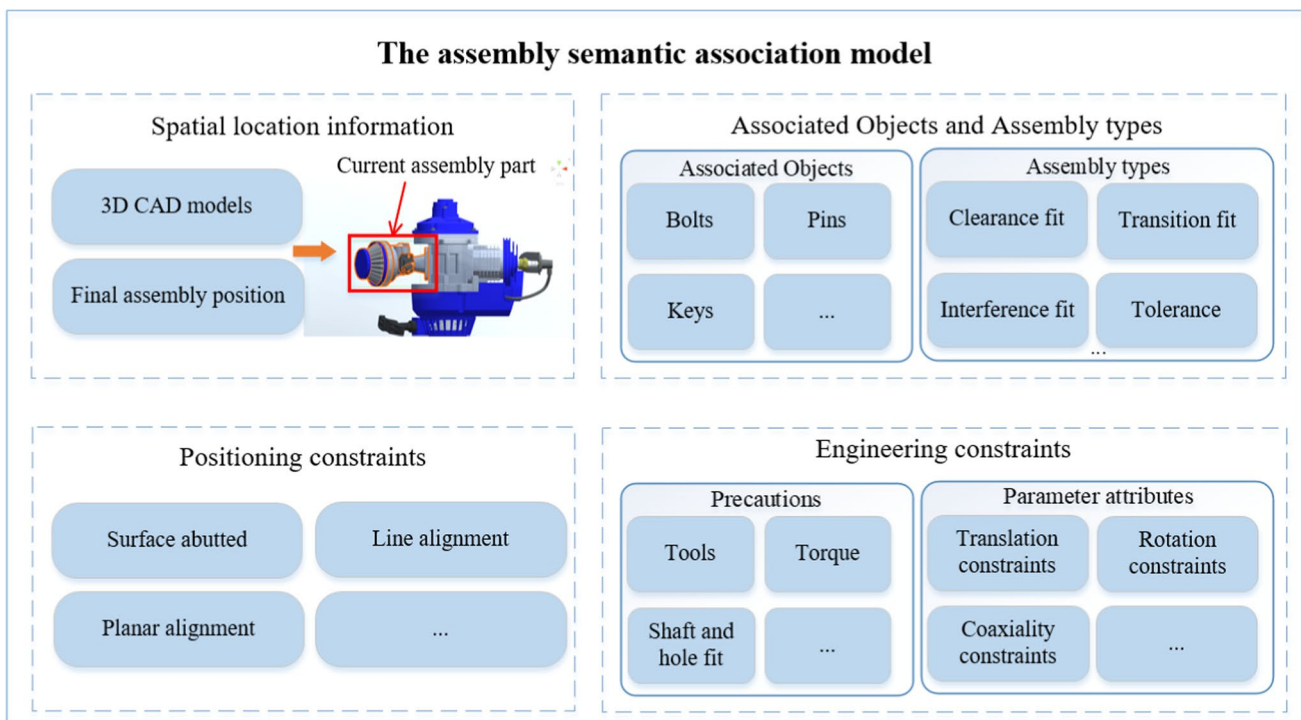

Fig. 3 The assembly semantic association model

gesture controls and can expand the hierarchical display of this information through gazing.

3.2.2 Visual presentation of remote expert attention

An interesting question is: how should the remote expert express his attention so that the local user can understand the important information he wants to convey? Our system mainly understands the attention of remote experts by perceiving the interaction behavior of experts and realizes the visual enhancement of expert operations by integrating gestures, eye gaze, and spatial visual cues. This section will focus on the visual presentation of the attention of the remote expert.

Visual presentation of gestures and eye gaze cues We implement gesture recognition and sharing using LeapMotion attached to the HTC VIVE Pro Eye headset. Through MRTK communication architecture, gesture data collected by LeapMotion can be shared to local MR clients. At the local MR client, we use the gesture structure of LeapMotion to create a virtual hand model and render it for display. When receiving the gesture data collected by LeapMotion of the remote VR client, the local MR client first decodes the shared gesture data and then uses it to control the virtual hand model. Finally, as shown in Fig. 4(a, d), the remote client's hand gestures are mapped to the local client hand model in real time.

In addition, remote experts can express collaborative attention through shared eye-gaze (EG) cues. The local user can perceive the area of concern of the remote expert through the shared EG. They can find an object in the collaborative space through the jumping and smooth tailing of the EG to shift their attention to the new target object.

We use the EG tracking function of the VR head display HTC VIVE Pro Eye to obtain the EG viewpoint coordinate data. After obtaining the EG viewpoint coordinate data, we combine these coordinate data with the user's gaze and head orientation to make a virtual ray and calculate the intersection of the ray and the object in the virtual assembly space. Then, we visualize the intersection data and use the server to transmit the data to the local MR client. And on the local MR side, the remote expert's EG is displayed through the method of virtual and real fusion.

Eye gaze information has the characteristics of "what you see is what you get [64]," which can solve the problem of ambiguous reference in remote collaboration and can represent the interaction intention of the collaborators. In addition, the remote collaboration based on gesture interaction is natural and intuitive in expressing collaborative attention, so the information sharing of hand-eye collaboration can improve the accuracy of the remote expert's expression of attention, as shown in Fig. 4(b, e).

Presentation of spatial visual cues In our research, the presentation of spatial visual cues is mainly aimed at 3D CAD models (virtual replica) with good 3D spatial visualization. We first import the CAD model in FBX format into Unity 3D and generate the prefab, and adopt the corresponding technical method of XML to realize the abstract reconstruction of the tree hierarchy information in the virtual assembly scene. When using XML to describe the virtual assembly scene abstractly, we use customized data tags to organize the semantic information of XML documents and build an association mapping with the physical entities and digital virtual entities of the assembly task. According to the generation rules and characteristics of XML, we define the vocabulary of XML nodes corresponding to the built virtual assembly
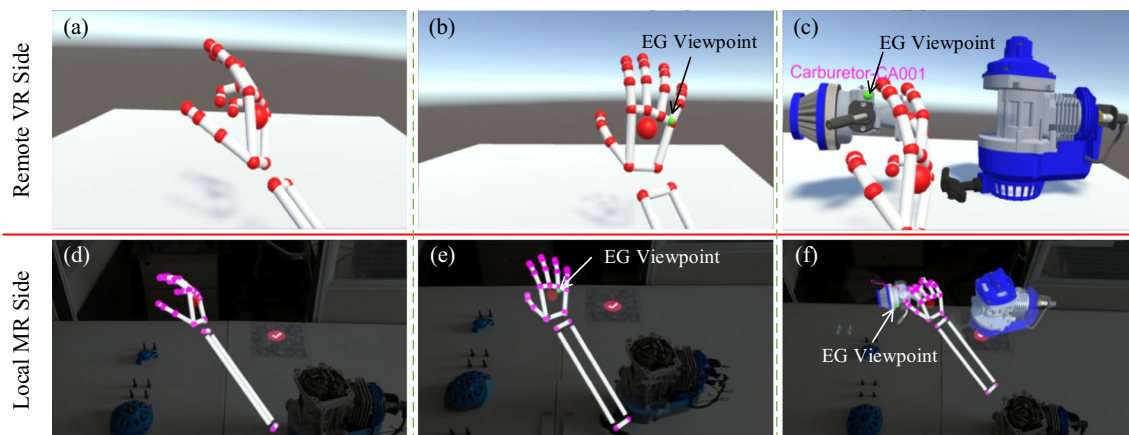
Fig. 4 The visual presentation of gestures and eye gaze cues and spatial visual cues. (**a**, **d**) The visual presentation of gestures cues. (**b**, **e**) The visual presentation of gestures and eye gaze cues. (**c**, **f**) The presentation of the fusion of gesture and eye gaze as well as spatial visual cues. (**a**–**c**) The HTC VIVE Pro Eye view on the remote VR side. (**d**–**f**) The Hololens view on the local MR side

scene and combine the logical relationship between objects to build the tag structure of XML documents, so as to generate scene data describing virtual assembly in XML. The generated virtual assembly scene data is stored on the server. The server needs to parse the XML document and generate MR assembly instruction working logic data in combination with other information, and then share it to VR and AR client programs for data processing through network communication. In VR and AR clients, the information described in the XML document can be reproduced as a tree hierarchical structure of the Unity 3D virtual assembly scene in VR and AR clients according to the inverse processing of the working logic data of the shared MR assembly instructions. Then, relevant 3D CAD models in the virtual assembly scene can be loaded through the server transmission to reproduce the preset virtual assembly scene. The presentation of the fusion of gesture and eye gaze as well as spatial visual cues is shown in Fig. 4(c, f).

Visual enhancement of expert operation Different from all previous studies, our system perceives the expert's attention through the expert's interactive behavior and uses visual enhancement to reinforce the important information the expert wants to convey to the user.

- Assembly process information enhancement presentation

We have designed an interactive area near the current virtual assembly parts to facilitate the interaction of remote experts. When it is detected that the expert's gesture is outside the interaction area, the expert can control when the process information at different levels is displayed through different gestures, so as to attract the user's attention to focus on the information the expert wants to express. When the expert's gesture is detected in the interaction area, it indicates that the expert may want to express an important operation intention, and we will describe the implementation details in the next section. Upload the collected expert gesture recognition data to the server. The data processing module maps this information with the divided assembly process level information in the resource file and generates MR assembly instruction working logic data with controllable information through data analysis and processing. The VR and AR clients can work on the inverse processing of logical data according to the shared MR assembly instructions, and the information described by the logical data can be reproduced as MR assembly instructions displayed at the information hierarchy level. Figure 5 shows the effect of assembly process information enhancement presentation of the remote expert.

- Important operational behavior presentation

Similar to the method described above, the difference is that when an expert's gesture enters the interaction area of the current virtual assembly part, our system perceives the expert's interaction behavior through eye tracking and gesture recognition and adjusts the visual form of the virtual part to achieve the presentation of the expert's important operation behaviors. We set the rules of selecting virtual objects through eye gaze and triggering MR visualization commands through gesture recognition to adjust the visualization form of virtual parts. On the server side, by acquiring the gaze information and gesture recognition information of remote experts in real time, the data processing module analyzes and processes this information and maps them with CAD models and MR visualization library to generate MR assembly instruction working logic data that can adjust the visual form of virtual assembly parts. In VR and AR clients, the information described by logical data can be reproduced as MR assembly instructions in the form of adjustable virtual part visualization according to the inverse processing of shared MR assembly instructions. Figure 6 shows the effect of the important operational behavior presentation of the remote expert.

- Operation details adaptive visual presentation

Sometimes, the expert needs to repeat the same operation tirelessly to guide the user in the assembly operation. Our system presents the details of assembly operation adaptively according to the interaction behavior of experts, so as to simplify the expert operation and the communication between experts and users. Remote experts only need to activate instructions through simple and intuitive interaction, and then our system can display clear MR assembly instructions in AR (local) and VR (remote) views, so that local users can perform assembly operations according to the instructions. Similar to our previous work [27], for specific physical tasks, we have established a parametric database for relevant operation guidance. The difference is that we set the rules for selecting virtual objects through gaze and triggering MR visualization commands through gesture recognition. When an expert's gesture is detected to enter the interaction area of the current virtual assembly part, our system perceives the expert's intention through eye tracking and gesture recognition to adaptively display detailed information such as assembly considerations for the current assembly part, rather than just displaying animation instructions. For details of our work implementation, please refer to Ref. [27]. As shown in Fig. 7c, the operator is completing the installation of the bolts on the engine. The remote expert selects the current virtual assembly bolt by staring and indicates that the bolt needs to be tightened through the "rotation gesture." Our system detects the interactive operation of the remote expert to
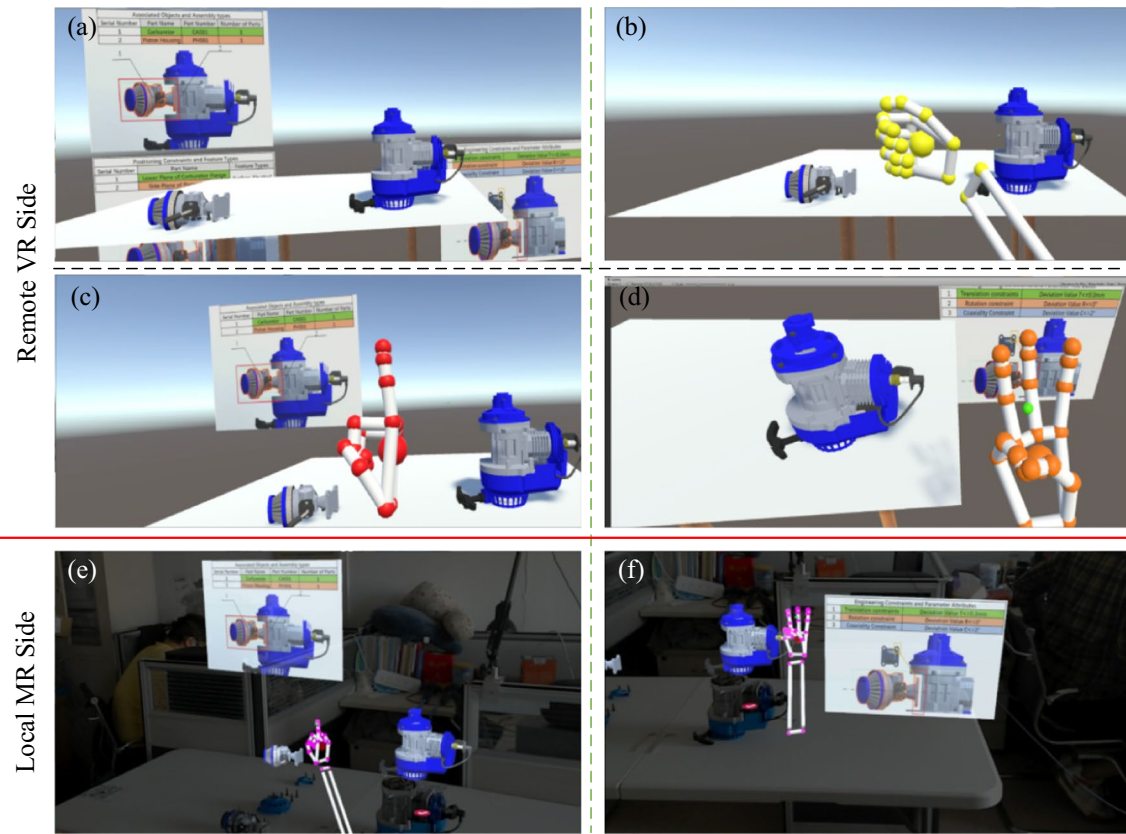
Fig. 5 The effect of assembly process information enhancement presentation of the remote expert. (a) All levels of information are displayed. (b) Gesture "zero" makes all information disappear. (c) Gesture "one" shows information on the first level. (d) Gesture "three" shows information on the third level. (a–d) The HTC VIVE Pro Eye view on the remote VR side. (e–f) The Hololens view on the local MR side
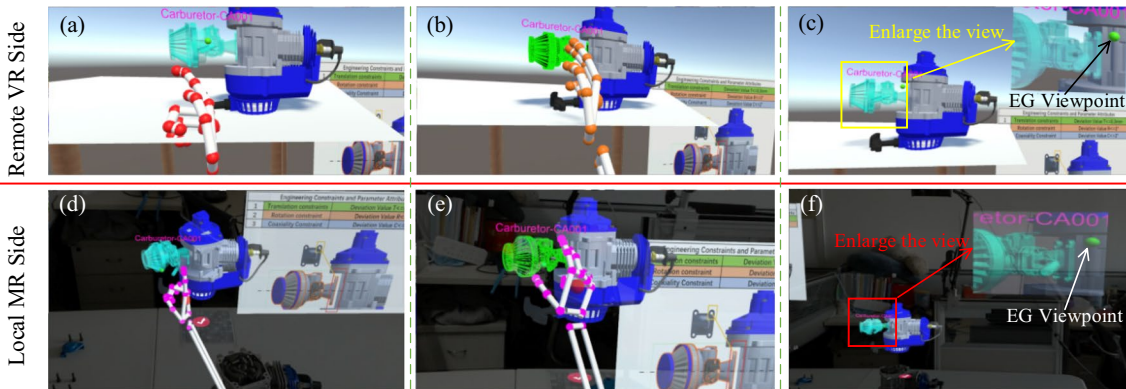

Fig. 6 The effect of the important operational behavior presentation of the remote expert. (a, b) Remote experts express important operational information by making the CAD model transparent and wireframe to attract users' attention, respectively. (c) Remote experts express important operation information by enlarging the staring area. (a–c) the HTC VIVE Pro Eye view on the remote VR side. (d–f) the Hololens view on the local MR side

trigger the demonstration animation of the current bolt installation in the parameterized database and the detailed information such as operation tools and assembly precautions, as shown in Fig. 7.

4 User study

In this section, we conducted a user study on EaVAS to investigate the benefits and limitations of adjusting the
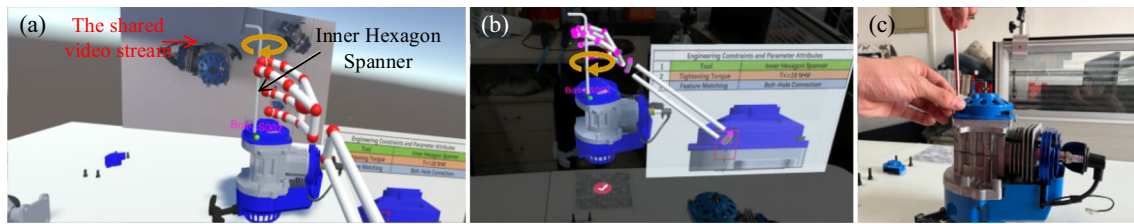
Fig. 7 The effect of the operation details presentation of the remote expert. a Remote experts trigger bolt installation animation and process details information through "tightening" gesture. **b** The Hololens view on the local MR side. **c** Assembly of physical parts on the local MR side

form of information visualization based on expert attention in MR remote collaborative assembly. We will describe experimental design details, summarize hypothesis testing, and report experimental results. We were interested in (1) how our system affects the task performance of experts and local users, (2) the effectiveness of our system in attracting user attention, and (3) the user experience effect of our system. Considering the experimental conditions and the actual assembly, we focused our research on the targets of our interest. Similar to previous studies, we used co-location instead of geographical separation in MR remote collaborative assembly.

4.1 Study design

In this research, we selected two experimental conditions:

1. 3DGAM [8]: A common assembly method in MR remote collaborative assembly. The system only supports sharing gestures and 3D CAD models;
2. EaVAS: An assembly method based on expert attention to adjust the visual form of information in MR remote collaborative assembly. The system not only supports the sharing of gestures, eye gaze, 3D CAD models, but also supports experts to adjust the information visualization form.

This user study used the within-subject design. A cross design was used to test the performance of 3DGAM and EaVAS. Dependent variables included task completion time, number of assembly errors, cognitive load, and user experience. We used the System Usability Scale questionnaire (SUS) [65] to verify the usability of our system. The results of the SUS scoring are shown in Fig. 8. For remote experts, the mean SUS score was 84.625 (SE = 2.643), while for local users, the mean SUS score was 80.250 (SE = 2.967). The results showed that our system was in the "good usability [65]" category for remote experts and local users.

The NASA-TLX questionnaire [66] was used to measure the subjective cognitive load of remote experts and local users. We designed a seven-point Likert scale

to evaluate the user experience. These questionnaires were collected after the experimenters completed the assembly task.

4.2 Experimental task

In order to simulate the remote collaborative assembly environment, we set our experimental space in a large enough room (6.1 m by 4.3 m). As shown in Fig. 9, the remote expert VR client and the local user MR client were separated by a physical partition. Remote experts and local users could communicate via voice. An experimental assembly platform was placed on the local MR client. Our task was to complete the engine assembly on this platform. The assembly task included completing the assembly of engine parts such as clutch shield, cap, carburetor, and bolts.

The reason why we chose this assembly task was to simulate the most common assembly tasks in the actual assembly process. The remote expert could observe the user's assembly status in real time and guide the user to complete the assembly task.
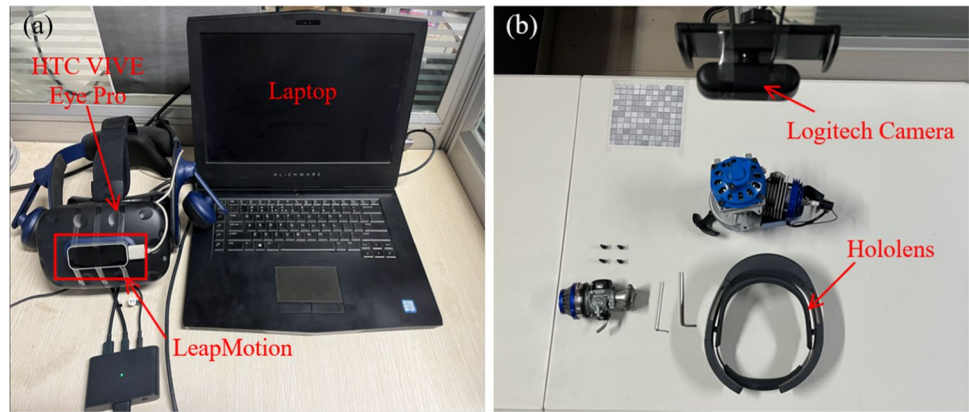
4.3 Hypotheses

What we were more interested in was whether the visual form of information adjusted by the system through
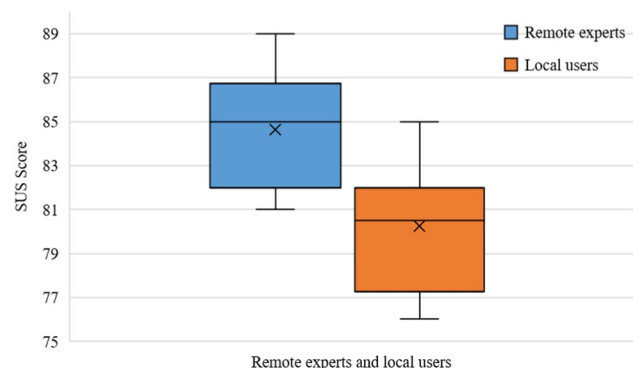
Fig. 8 The SUS questionnaire results of remote experts and local users

Fig. 9 **a** remote VR expert side setup, **b** local MR user side setup

perceiving the behavior of experts would affect the task performance of experts and users, and whether the system was effective in reducing the cognitive burden and attracting users' attention. Effective remote collaboration requires the use of communication cues, which help users understand tasks more easily. As pointed out in Ref. [67], rich and efficient communication cues are essential for effective remote collaboration. In remote collaboration, it is important for everyone to be able to communicate their intentions accurately [68]. In mixed reality environments, where real and virtual information are fused, local users may become confused due to information overload [69] and have difficulty understanding the operation intentions of experts. Angelo et al. [70] proposed that monitoring expert attention can help local users reduce cognitive load and improve task efficiency. To attract users' attention, our EaVAS system supports remote experts in focusing on enhancing key information to improve the expression of expert collaborative intentions. We enhanced the 3DGAM system proposed by Ref. [8] with gaze cues and visual enhancement based on expert attention. It is worth noting that gaze cues positively affect MR remote collaboration by enhancing users' attention, efficiency, and quality [2]. In addition, Ref. [71] demonstrates that effective visual expression can capture users' attention and improve their cognitive ability. Based on this and the results of previous studies [72], we proposed the following four hypotheses:

H1: Time. The EaVAS will be more efficient than the 3DGAM in task completion time.
H2: Error. Using EaVAS will reduce operating errors.
H3: User cognitive load. The cognitive load of using EaVAS for both experts and local users is lower than that of 3DGAM.
H4: User Experience (UX). EaVAS will provide a better user experience than 3DGAM.
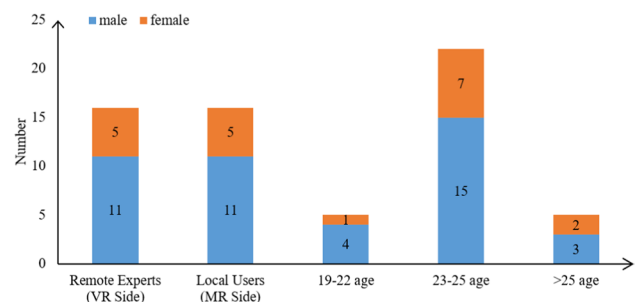
4.4 Participants

We invited 32 participants (16 pairs) from Northwestern Polytechnical University, including 22 males and 10 females and aged from 22 to 29 years ($M = 25$, SD = 2.4). We sought participants with AR/VR/MR experience to reduce the impact of novelty effects. Figure 10 shows more details of the participants in the experiment.
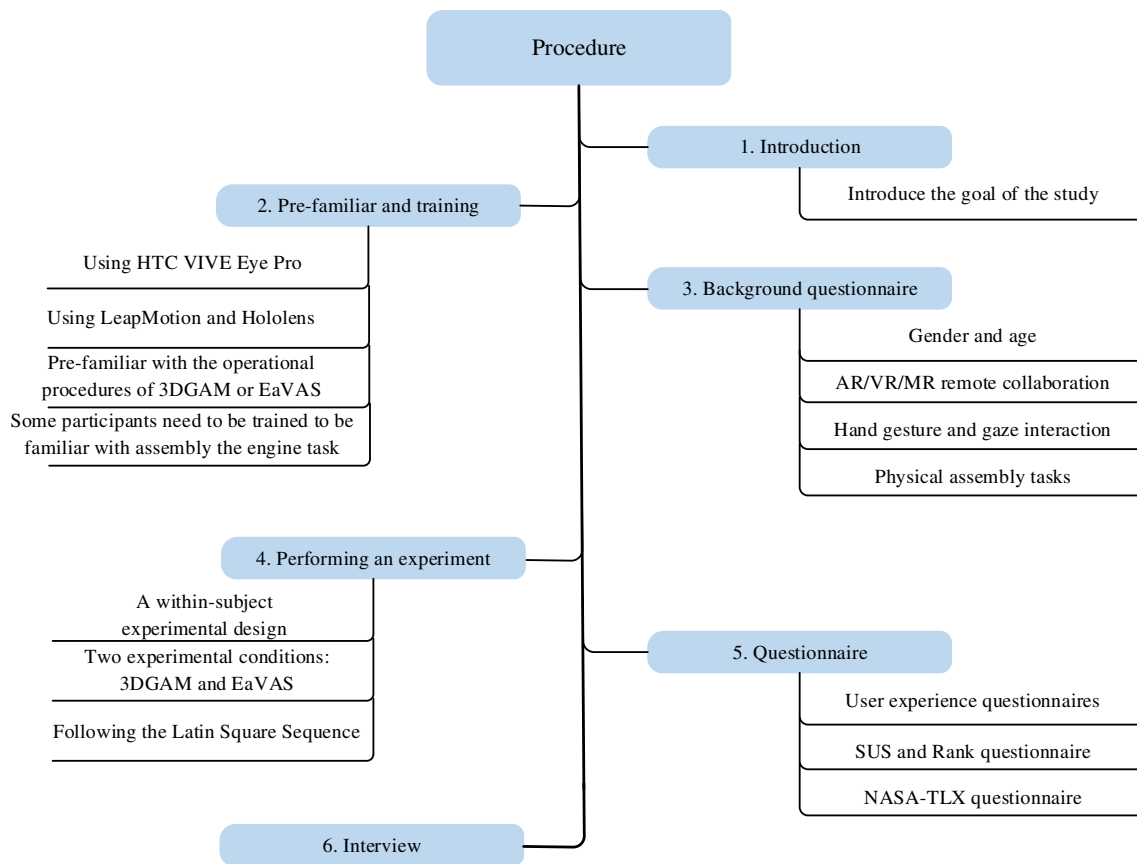
4.5 Procedure

The user study experiment procedure followed the six steps shown in Fig. 11. Each participant pair performed two rounds (e.g., 3DGAM, EaVAS) of an experiment. They were randomly assigned to expert groups or local user groups, and their roles did not change during the experiment. Before the experiment, the participants were informed of the objectives of the experiment and were familiar with the operation process of 3DGAM and EaVAS in advance.

In addition, we would explain the meaning of each process parameter and other data to the participants to ensure that they fully understand the content of the instructions provided. Then, participants were asked to complete a short questionnaire about the research background. In our experiments, participants completed the assembly task in two conditions (3DGAM and EaVAS). Figure 12

Fig. 10 Statistical data of participants in the experiment

Fig. 11 The procedure of user study

shows the collaborative cooperation scenario where our EaVAS guides the completion of engine assembly tasks in our experiments. Remote VR experts could control the presentation of information and present the operation intention in the form of visual enhancement (as shown in Fig. 12(a–d)). Local MR users could complete the assembly of the engine under the guidance of visual information shared by experts (as shown in Fig. 12(e–h)). The main process of assembling the engine is shown in Fig. 12.

We used timers to record the time taken by remote and local participants to complete the engine assembly task and counted the number of assembly errors (e.g., WPA is the number of wrong parts assembled, and IGP is the number of incorrect guidance provided). We changed the condition between the 3DGAM and the EaVAS following a Latin Square Sequence to reduce learning effects. After the assembly task was completed, both the remote expert and the local user were required to complete the user experience questionnaires (see Table 1). Participants were then asked to rank the two experimental conditions according to preference. Then, we asked both remote experts and local users to complete the NASA-TLX questionnaire. Finally,

each participant was asked to conduct an interview based on the content of the experiment.

4.6 Results

This section reports the results of the analysis of the data from our experimental measurements. We conducted a normal analysis on all dimensions of data. The paired t-test was used when the data met the assumption of normality, and the Wilcoxon signed-rank test was used when the data did not meet the assumption of normality.

4.6.1 Performance time

We hoped to explore whether the EaVAS interface was more efficient in task performance than the 3DGAM interface. Therefore, we compared the time performance of the two methods in assembly tasks. Table 1 shows the average time performance under different conditions. A paired t-test ($\alpha = 0.05$) revealed that there were a statistically significant difference between the EaVAS condition and 3DGAM condition on performance time ($t(15) = 10.366$, $p < 0.001$).

Remote VR Side
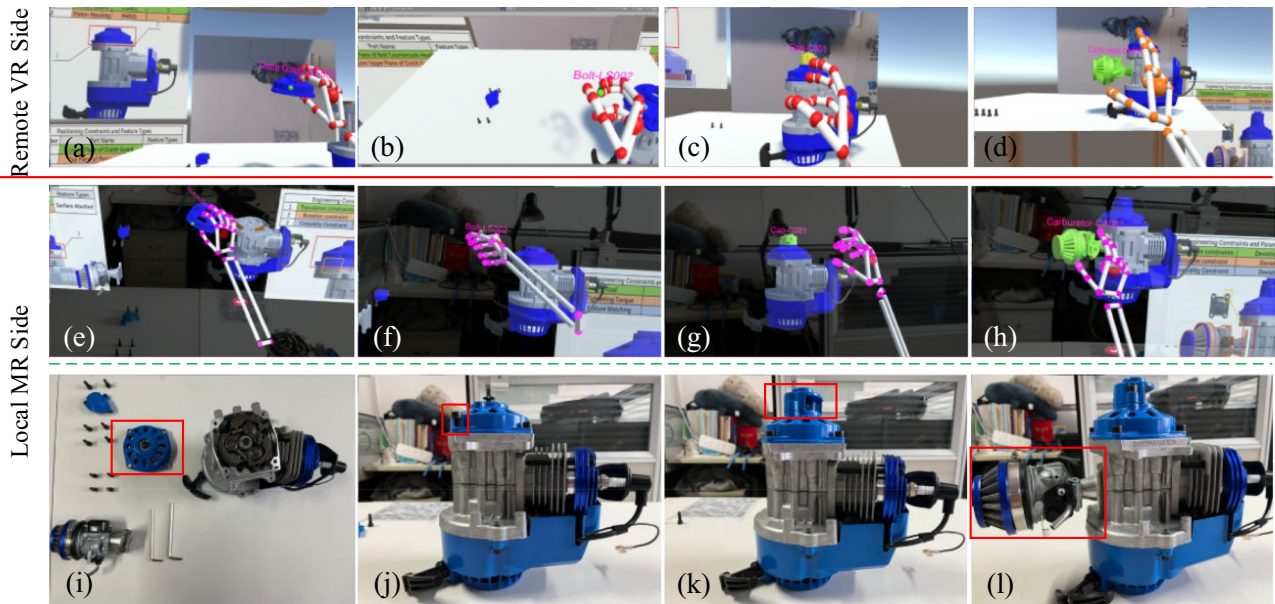
Local MR Side

Fig. 12 The main process of assembling the engine. (a–d) The HTC VIVE Pro Eye view on the remote VR side. (e–h) The Hololens view on the local MR side. (i–l) Assembly of physical parts on the local MR side

Moreover, according to the statistical data, the average time to complete the assembly task using EaVAS (M = 474.310, SE = 1.527) interface was significantly shorter than that of 3DGAM (M = 500.060, SE = 1.745).

4.6.2 Error evaluation

We had hoped to explore whether the use of EaVAS in assembly tasks could reduce the rate of assembly errors. To our surprise, according to the Wilcoxon signed rank test ($\alpha = 0.05$), there were no statistically significant differences in IGP ($Z = -1.342$, $p = 0.180$) and WPA ($Z = -1.732$, $p = 0.083$) between the EaVAS and the 3DGAM interface in engine assembly task. However, we found that the remote and local participants using our EaVAS interface (IGP: M = 0.188, SE = 0.099; WPA: M = 1.063, SE = 0.139) had fewer errors than the 3DGAM interface (IGP: M = 0.438,

SE = 0.249; WPA: M = 1.250, SE = 0.167), as shown in Table 1.

4.6.3 Cognitive load

Cognitive load was an important measure of the effectiveness of our EaVAS system. We used the NASA-TLX questionnaire to measure cognitive load. Through paired t-test ($\alpha = 0.05$), we explored the influence of the EaVAS condition and 3DGAM condition on global cognitive load. We found that there were statistically significant differences in cognitive load for both remote experts ($t(15) = 6.780$, $p < 0.001$) and local workers ($t(15) = 13.500$, $p < 0.001$) between each experimental condition, as shown in Table 1. For remote experts and local users, 3DGAM (remote experts: M = 12.507, SE = 0.259; local users: M = 11.959, SE = 0.268) brought a heavier cognitive load than EaVAS

Table 1 Assembly performance data results for the two experimental conditions reported by remote experts and local users

	3DGAM		EaVAS		
	Remote experts	Local users	Remote experts	Local users	
Performance time(s)	500.060 ± 1.745		474.310 ± 1.527		$t(15) = 10.366^*$
Number of ICP	0.438 ± 0.249		0.188 ± 0.099		$Z = -1.342$
Number of WPA	1.250 ± 0.167		1.063 ± 0.139		$Z = -1.172$
Cognitive load	12.507 ± 0.259	11.959 ± 0.268	10.469 ± 0.231	7.861 ± 0.172	$t_e(15) = 6.780^*$ $t_u(15) = 13.500^*$

*$p < 0.05$ and there is a significant difference between two different conditions. The $t(15)$ indicates the t-value in the paired t-test. The Z indicates the Z-value in the Wilcoxon signed rank test. The $t_e(15)$ indicates the t-value in the paired t-test among remote experts. The $t_u(15)$ indicates the t-value in the paired t-test among local users

(remote experts: $M = 10.469$, SE = 0.231; local users: $M = 7.861$, SE = 0.172).

4.6.4 User experience

User experience was critical to the availability of a system. As shown in Table 2, we designed a seven-point Likert scale to assess the impact of the EaVAS interface and 3DGAM interface on the user experience of experts and user participants. We evaluated the user experience of participants from twelve aspects (presence (Q1), efficiency (Q2), feeling (Q3), responsiveness (Q4), confidence (Q5), collaboration (Q6), attention (Q7), cognition (Q8), helpfulness (Q9), convenience (Q10), focus (Q11), usability (Q12). Q7, Q8, and Q9 are only for local users, while Q10, Q11, and Q12 are only for remote experts). We used the Wilcoxon signed rank test ($\alpha = 0.05$) to explore whether there was a difference in user experience between the EaVAS interface and the 3DGAM

interface. The statistical data results are shown in Figs. 13 and 14.

For remote VR experts (as shown in Fig. 13), there were statistically significant differences in terms of efficiency (Q2: $Z = -2.041$, $p < 0.05$), feeling (Q3: $Z = -2.848$, $p < 0.01$), confidence (Q5: $Z = -2.694$, $p < 0.01$), collaboration (Q6: $Z = -2.873$, $p < 0.01$), convenience (Q10: $Z = -2.682$, $p < 0.01$), focus (Q11: $Z = -2.831$, $p < 0.01$), and usability (Q12: $Z = -2.699$, $p < 0.01$). No significant differences between the two experimental conditions were observed for the other two factors (i.e., presence (Q1: $Z = -1.890$, $p = 0.059$), responsiveness (Q4: $Z = -1.508$, $p = 0.132$)).

For local MR users (as shown in Fig. 14), there were statistically significant differences in terms of presence (Q1: $Z = -2.877$, $p < 0.01$), efficiency (Q2: $Z = -2.555$, $p < 0.05$), feeling (Q3: $Z = -2.825$, $p < 0.01$), responsiveness (Q4: $Z = -2.354$, $p < 0.05$), confidence (Q5: $Z = -2.410$, $p < 0.05$), collaboration (Q6: $Z = -2.911$, $p < 0.01$), attention

Table 2 Likert scale rating questions for user experience

Question	Question statement	Evaluation factors
Q1	I could easily notice the **presence** of my partner	Presence
Q2	This interface allowed me to **quickly** complete tasks	Efficiency
Q3	I **enjoyed** using this interface to complete tasks	Feeling
Q4	I could **react quickly** according to the partner's operation	Responsiveness
Q5	I was **confident** that I had completed the task correctly	Confidence
Q6	My partner and I could work **together** well to complete the task	Collaboration
Q7, Q8, Q9 are provided to local users		
Q7	This interface allowed me to capture the **attention** of my partner	Attention
Q8	This interface allowed me to **easily understand** the instruction information	Cognition
Q9	The MR instruction information of this interface was very **helpful** for me to complete the task	Helpfulness
Q10, Q11, Q12 are provided to remote experts		
Q10	This interface allowed me to **easily provide** clear instruction information	Convenience
Q11	This interface allowed me to **focus** on the information I want to convey	Focus
Q12	The information provided by this interface was **intuitive** and **useful**	Usability

Fig. 13 User experience results (mean ± SE) for two experimental conditions reported by remote experts, *p indicates a significant difference between two different conditions
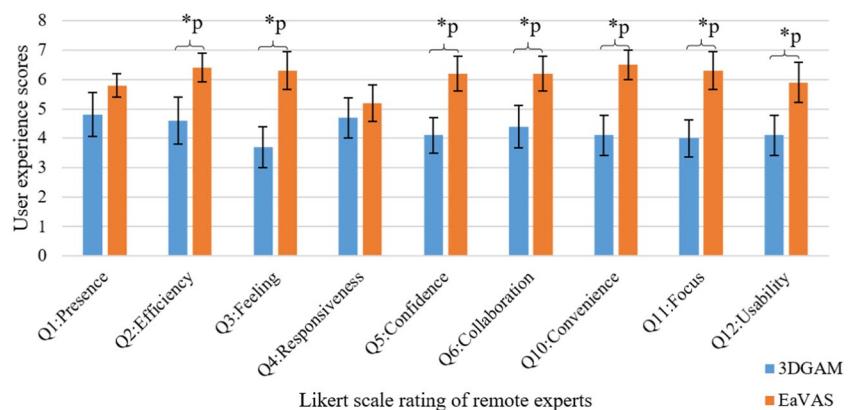
Fig. 14 User experience results (mean ± SE) for two experimental conditions reported by local users, *p* indicates a significant difference between two different conditions
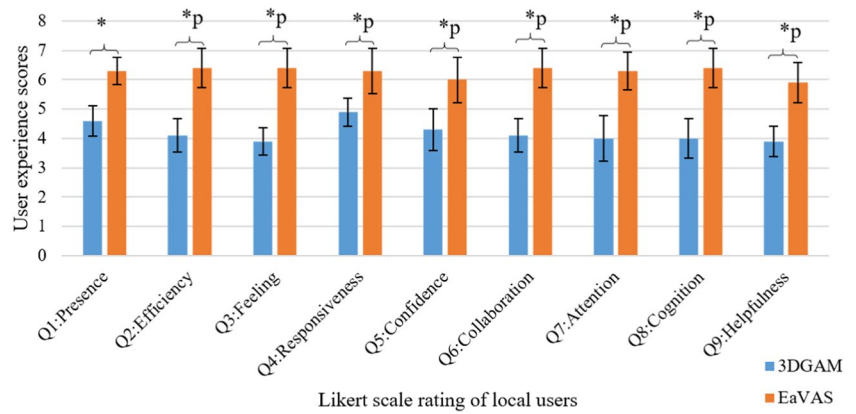
(Q7: $Z = -2.684$, $p < 0.01$), cognition (Q8: $Z = -2.820$, $p < 0.01$), and helpfulness (Q9: $Z = -2.332$, $p < 0.05$).
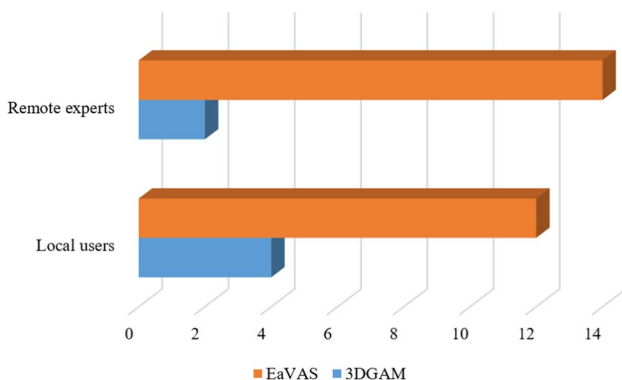
4.6.5 User preferences

By analyzing the statistical user preference data, we could know which interface users prefer to use. As shown in Fig. 15, we asked participants to complete the preference questionnaire to rank the two conditions of the experiment. The results showed that for both remote VR experts and local MR users, most participants preferred the EaVAS interface to the 3DGAM interface.

5 Discussion

5.1 Task performance

In our research, the indicators of task performance include performance time and error evaluation. We tested the performance time of EaVAS and 3DGAM interfaces in the engine assembly task to verify hypothesis 1. The results described in Sect. 4.6.1 show that it takes significantly less time to

Fig. 15 User preference results for the two experimental conditions reported by remote experts and local users

complete the engine assembly task using the EaVAS interface than 3DGAM, which proves that the EaVAS interface is more efficient (see Table 1). The feedback of Q2 also supports this view (see Figs. 13 and 14). According to our statistical time data and the feedback results of Q2 and Q4, we found that the operation response time of local users was directly related to the effectiveness of information transmitted to users by remote experts. That may be because local users need to know the correct assembly process information and precautions before assembling parts. Therefore, we have reason to believe that the more difficult the cognitive information transmitted by the remote expert, the longer it takes for the local user to complete the assembly task. "I don't have to think about what the experts mean anymore," said a local user who participated in the experiment, "I can easily find the information using this system, I can see the operations that the experts want me to complete, and I can just assemble according to the prompts." A reasonable explanation for these results is that EaVAS has introduced the information hierarchy division method based on assembly semantic association model and the visual enhancement mechanism for expert operation. Remote experts can adjust the visual form of assembly guidance information to enhance the key information that experts want to transfer to local users, which speeds up the efficiency of local users to obtain effective information, so they can have a more efficient performance. So hypothesis 1 is accepted.

Our team previously believed that the operation of visual enhancement experts could make local users pay more attention to the operation guidance of experts, thus reducing the operation error rate, so we proposed hypothesis 2. However, the results described in Sect. 4.6.2 indicate that there are no statistically significant differences in IGP and WPA between the EaVAS and the 3DGAM. In essence, remote experts affect the key information that local users pay attention to through visual cues to form their own psychological representation in the memory of local users, and ultimately affect the user's assembly operation. In engine assembly tasks, ICP and WPA using the

EaVAS interface were lower than those using the 3DGAM interface, although there was no statistically significant difference between the two interfaces. This shows that our EaVAS interface has a certain effect on users. The feedback results of Q5 and Q6 also support this view (see Figs. 13 and 14). Through interviews with remote experts and local users and analysis of ICP and WPA data, we speculate that the information hierarchy division method based on assembly semantic association model and the visual enhancement mechanism of expert operation may affect the psychological representation of local users. This requires further research to explore the relationship between visual cues of key information shared by remote experts and the degree of distraction of local users. Therefore, hypothesis 2 is rejected.

5.2 Spatial cognition

In order to prove hypothesis 3, EaVAS enables experts to control the distribution of MR interface information and the adjustment of visual forms through information hierarchical display and expert operation visual enhancement mechanism, thereby reducing the cognitive burden of users. It can be seen from Table 1 and Sect. 4.6.3 that for both remote VR experts and local MR users, the EaVAS interface can effectively reduce the cognitive burden of users compared with the 3DGAM interface. EaVAS improves the availability of visual information transmitted by remote experts to local users through expert attention perception (Fig. 4), information hierarchical processing (Fig. 5) and intuitive virtual model visualization (Figs. 6 and 7). This allows local users to focus on the key information transmitted by remote experts while reducing the amount of information, so that they can correctly complete the assembly task under the guidance of remote experts. This is consistent with the results of Q8 and Q9 feedback (see Fig. 14). "This interface is really great. I can see the gestures and viewpoints of experts, as well as the assembly process information of parts and the visual changes of part models. I can understand the assembly operations that experts ask me to do without even listening to what they are saying," said a local MR participant. We speculate that this may be because EaVAS enables experts to freely adjust the visual changes of information, which reduces the amount of information displayed while ensuring the existence of necessary information. At the same time, EaVAS can also adjust the intuitive visualization of virtual parts to focus on expressing the information that experts want to convey. This reduces the difficulty of local users' information cognition, so that local users can confidently complete assembly tasks. Therefore, hypothesis 3 should be accepted.

5.3 Attention presentation

EaVAS realizes the presentation of expert attention through the enhanced presentation of assembly process information, the expression of important operational behaviors, and the adaptive visual presentation of details, so as to improve the efficiency of visual information transmitted by remote experts to local users. It simplifies the operation of remote VR experts and improves the cognitive efficiency of local MR users. Our team believes that EaVAS provides an interface for remote experts to interact freely, so that remote experts can control the visual display of information to transmit to local users the key operation information. The results of the feedback from Q10 and Q11 also support this view (see Fig. 13, Sect. 4.6.4). In addition, according to Figs. 13 and 14 and Sect. 4.6.4, for both remote VR experts and local MR users, the two interfaces have significant effects on efficiency (Q2), feeling (Q3), responsiveness (Q4), confidence (Q5), and collaboration (Q6). The user experience of EaVAS in these aspects is better than that of 3DGAM interface. We speculate that this is mainly because EaVAS provides remote experts with great interaction freedom, enabling remote experts to express their attention through simple interaction, while local users can more easily understand experts' operations in rich visual information. Therefore, the user experience of the EaVAS interface is comprehensively superior to the 3DGAM interface. Therefore, hypothesis 4 should be accepted.

6 Limitations and future works

6.1 Visual display settings

In our system, remote VR experts can select virtual parts by eye gaze and change the visual form of virtual parts through gesture recognition. However, in the engine assembly experiment, some local MR participants complained that the sudden change in the visualization form of virtual parts had brought them some confusion. This also affects the user experience of EaVAS to some extent. In addition, our system at this stage changes the visualization form of the whole virtual part rather than the visual change of the virtual part in the local scope of the expert hand. This experiment cannot prove whether this visualization display setting affects the presentation of the expert's attention. Therefore, in the future, we can improve the visual display settings of our system to more accurately express the experts' attention, and add the time smooth transition settings of visual form changes based on the existing technology to reduce the troubles caused by sudden changes in the virtual model.

6.2 Multi-channel interactive settings

Through interviews with remote VR participants and local MR participants, we found that some participants hope our EaVAS interface can support annotation cues. A small number of participants suggested that our system should be able to display the virtual avatar of remote experts on the local MR client. This gives us some new inspiration. Therefore, in the future research work, we will try to add these functions to our system, so that remote VR experts can increase annotation cues through gesture interaction to enhance the communication of operational intent and improve the sense of co presence and the performance through the avatar.

6.3 Limited experimental conditions

Due to the severity of the COVID-19 epidemic during data collection, the experiment involved a relatively small number of participants. Despite our efforts to recruit more participants, we had to limit the number to ensure safety and comply with health guidelines. Therefore, the general applicability of our research results may be limited. In addition, the experiment was conducted in a controlled laboratory environment, which may not fully represent the real-world environment of the expected application of the technology. This research focuses on validating the technical feasibility and performance of the proposed method. While we have simulated real-world conditions as closely as possible, there may still be factors that we have not accounted for that could affect how the technology will perform in real-world situations. In the future, we will make our system available to more people in actual industrial production to verify its usability and practicality.

7 Conclusion

In this paper, for the first time, a method of sensing expert attention and visually enhancing expert operation (EaVAS) in MR remote collaborative assembly is proposed. This paper proves that EaVAS has higher time performance and better user experience than the traditional MR remote collaborative assembly method (3DGAM). This research aims to perceive the expert's operation attention according to the gaze and gesture interaction of the remote expert and to enhance the key information that the expert wants to convey to the local user by adjusting the visual form of the assembly guidance information. We developed EaVAS through the information hierarchy division method based on assembly semantic association model and the expert operation visual enhancement mechanism integrating gesture, eye gaze, and spatial visual cues. We designed an experimental case by imitating the actual engine assembly. To test the effect of

the experiment, 32 participants (16 pairs) were randomly assigned to different MR remote collaborative assembly systems (EaVAS and 3DGAM) groups. The experimental results were analyzed in terms of performance time, error evaluation, cognitive load, and user experience. All hypotheses except hypothesis 2 are accepted. Therefore, EaVAS is helpful for simplifying remote expert operations and improving the cognitive efficiency of local users.

Declarations

Competing interests The authors declare no competing interests.

References

1. Anton D, Kurillo G, Bajcsy R (2018) User experience and interaction performance in 2D/3D telecollaboration. Futur Gener Comput Syst 82:77–88. https://doi.org/10.1016/j.future.2017.12.055
2. Piumsomboon T, Dey A, Ens B, Lee G, Billinghurst M (2019) The effects of sharing awareness cues in collaborative mixed reality. Frontiers in Robotics and AI 6:5. https://doi.org/10.3389/frobt.2019.00005
3. Teo T, Lawrence L, Lee G A, 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
4. Huang W, Alem L, Tecchia F (2013) HandsIn3D: supporting remote guidance with immersive virtual environments. Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2–6, Proceedings, Part I 14. Springer, Berlin Heidelberg 2013:70–77. https://doi.org/10.1007/978-3-642-40483-2_5
5. Huang W, Alem L (2013) Gesturing in the air: supporting full mobility in remote collaboration on physical tasks. J Univ Comput Sci 19(8):1158–1174. https://doi.org/10.3217/jucs-019-08-1158

6. Marques B , Silva S, Alves J, Rocha A, Dias P, Santos B S (2022). Remote collaboration in maintenance contexts using augmented reality: insights from a participatory process. International Journal on Interactive Design and Manufacturing (IJIDeM), 1–20. https://doi.org/10.1007/s12008-021-00798-6

7. Gurevich P, Lanir J, Cohen B (2015) Design and implementation of teleadvisor: a projection-based augmented reality system for remote collaboration. Comput Support Cooperative Work (CSCW) 24(6):527–562. https://doi.org/10.1007/s10606-015-9232-7

8. 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. Multimedia Tools Appl 80(20):31059–31084. https://doi.org/10.1007/s11042-020-09731-7

9. Sasikumar P, Chittajallu S, Raj N, Bai H, Billinghurst M (2021) Spatial perception enhancement in assembly training using augmented volumetric playback. Front Virtual Real 2:698523. https://doi.org/10.3389/frvir.2021.698523

10. Piumsomboon T, Lee Y, Lee G A, Dey A, Billinghurst M (2017). Empathic mixed reality: sharing what you feel and interacting with what you see. In 2017 International Symposium on Ubiquitous Virtual Reality (ISUVR) (pp. 38–41). IEEE. https://doi.org/10.1109/ISUVR.2017.20

11. Zhang X, Bai X, Zhang S, He W, Wang S, Yan Y, Yu Q, Liu L (2023) A novel MR remote collaboration system using 3D spatial area cue and visual notification. J Manuf Syst 67:389–409. https://doi.org/10.1016/j.jmsy.2023.02.013

12. Wang P, Bai X, Billinghurst M, Zhang S, Zhang X, Wang S, He W, Yan Y, Ji H (2021) AR/MR remote collaboration on physical tasks: a review. Robot Comput-Integrat Manuf 72:102071. https://doi.org/10.1016/j.rcim.2020.102071

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

14. Kritzler M, Murr M, Michahelles F (2016). Remotebob: support of on-site workers via a telepresence remote expert system. In Proceedings of the 6th International Conference on the Internet of Things (pp. 7–14). https://doi.org/10.1145/2991561.2991571

15. Wang P, Zhang S, Billinghurst M, Bai X, He W, Wang S, Sun M, Zhang X (2020) A comprehensive survey of AR/MR-based co-design in manufacturing. Eng Comput 36(4):1715–1738. https://doi.org/10.1007/s00366-019-00792-3

16. 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. 2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE). IEEE, 9–12. https://doi.org/10.1109/3DCVE.2016.7563559

17. Wu T Y, Gong J, Seyed T, Yang X D (2019) Proxino: enabling prototyping of virtual circuits with physical proxies. Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology 121–132. https://doi.org/10.1145/3332165.3347938

18. García-Pereira I, Portalés C, Gimeno J, Casas S (2020) A collaborative augmented reality annotation tool for the inspection of prefabricated buildings. Multimedia Tools Appl 79(9):6483–6501. https://doi.org/10.1007/s11042-019-08419-x

19. Wang P, Zhang S, Bai X, Billinghurst M, He W, Sun M, Chen Y, Lv H, Ji H (2019) 2.5 DHANDS: 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

20. Autor D (2014) Polanyi's paradox and the shape of employment growth. Natl Bureau Econ Res. https://doi.org/10.3386/w20485

21. Chen X, Gao S, Guo S, Bai J (2012) A flexible assembly retrieval approach for model reuse. Comput Aided Des 44(6):554–574. https://doi.org/10.1016/j.cad.2012.02.001

22. 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

23. 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

24. 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

25. Sukan M, Elvezio C, Feiner S, Tversky B (2016) Providing assistance for orienting 3D objects using monocular eyewear. Proceedings of the 2016 Symposium on Spatial User Interaction 89–98. https://doi.org/10.1145/2983310.2985764

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

27. Wang P, Bai X, Billinghurst M, Han D, Zhang S, He W, Zhang X, Yan Y (2019) I'm tired of demos: an adaptive MR remote collaborative platform. SIGGRAPH Asia 2019 XR 17–18. https://doi.org/10.1145/3355355.3361878

28. Gauglitz S, Lee C, Turk M, Höllerer T (2012). Integrating the physical environment into mobile remote collaboration. In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (pp. 241–250). https://doi.org/10.1145/2371574.2371610

29. Fussell S R, Kraut R E, Siegel J (2000). Coordination of communication: effects of shared visual context on collaborative work. In Proceedings of the 2000 ACM conference on Computer supported cooperative work (pp. 21–30). https://doi.org/10.1145/358916.358947

30. Kraut R E, Fussell S R, Siegel J (2003). Visual information as a conversational resource in collaborative physical tasks. Human–computer interaction, 18(1–2), 13–49. https://doi.org/10.1207/S15327051HCI1812_2

31. Günther S, Kratz S, Avrahami D, Mühlhäuser M (2018). Exploring audio, visual, and tactile cues for synchronous remote assistance. In Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference (pp. 339–344). https://doi.org/10.1145/3197768.3201568

32. Anton D, Kurillo G, Yang A Y, Bajcsy R (2017). Augmented telemedicine platform for real-time remote medical consultation. In MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4–6, 2017, Proceedings, Part I 23 (pp. 77–89). Springer International Publishing. https://doi.org/10.1007/978-3-319-51811-4_7

33. Wang S, Parsons M, Stone-McLean J, Rogers P, Boyd S, Hoover K, Meruvia-Pastor O, Gong M, Smith A (2017) Augmented reality as a telemedicine platform for remote procedural training. Sensors 17(10):2294. https://doi.org/10.3390/s17102294

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

35. Gurevich P, Lanir J, Cohen B, Stone R (2012). TeleAdvisor: a versatile augmented reality tool for remote assistance. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 619–622). https://doi.org/10.1145/2207676.2207763

36. Fussell S R, Setlock L D, Parker E M, Yang J. (2003) Assessing the value of a cursor pointing device for remote collaboration on physical tasks. CHI'03 Extended Abstracts on Human Factors in Computing Systems 788–789. https://doi.org/10.1145/765891.765992

37. Pejsa T, Kantor J, Benko H, Ofek E, Wilson, A (2016). Room-2room: enabling life-size telepresence in a projected augmented reality environment. In Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing (pp. 1716–1725). https://doi.org/10.1145/2818048.2819965

38. Piumsomboon T, Day A, Ens B, Lee Y, Lee G, Billinghurst M (2017) Exploring enhancements for remote mixed reality collaboration. SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications 1–5. https://doi.org/10.1145/3132787.3139200

39. Choi SH, Kim M, Lee JY (2018) Situation-dependent remote AR collaborations: Image-based collaboration using a 3D perspective map and live video-based collaboration with a synchronized VR mode. Comput Ind 101:51–66. https://doi.org/10.1016/j.compind.2018.06.006

40. Lee G, Kim S, Lee Y, Dey A, Piumsomboon T, Norman M, Billinghurst M (2017). Mutually shared gaze in augmented video conference. In Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2017 (pp. 79–80). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ISMAR-Adjunct.2017.36

41. Wang Z, Wang Y, Bai X, Huo X, He W, Feng S, Zhang J, Zhang Y, Zhou J (2021) SHARIDEAS: a smart collaborative assembly platform based on augmented reality supporting assembly intention recognition. Int J Adv Manuf Technol 115(1):475–486. https://doi.org/10.1007/s00170-021-07142-y

42. Tecchia F, Alem L, Huang W (2012). 3D helping hands: a gesture based MR system for remote collaboration. In Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry (pp. 323–328). https://doi.org/10.1145/2407516.2407590

43. Huang W, Alem L (2013). HandsinAir: a wearable system for remote collaboration on physical tasks. In Proceedings of the 2013 conference on Computer supported cooperative work companion (pp. 153–156). https://doi.org/10.1145/2441955.2441994

44. Li J, Wessels A, Alem L, Stitzlein C (2007) Exploring interface with representation of gesture for remote collaboration. Proceedings of the 19th Australasian Conference on Computer-Human interaction: Entertaining User interfaces 179–182. https://doi.org/10.1145/1324892.1324926

45. Alem L, Li J (2011) A study of gestures in a video-mediated collaborative assembly task. Advances in Human-Computer Interaction, 2011. https://doi.org/10.1155/2011/987830

46. Kirk D, Rodden T, Fraser D S (2017) Turn it this way: grounding collaborative action with remote gestures. Proceedings of the SIGCHI conference on Human Factors in Computing Systems 1039–1048. https://doi.org/10.1145/1240624.1240782

47. 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). IEEE 91–92. https://doi.org/10.1109/ISMAR-Adjunct.2019.00038

48. McNamara A, Boyd K, George J, Jones W, Oh S, Suther A (2019) Information placement in virtual reality. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 1765–1769. https://doi.org/10.1109/VR.2019.8797891

49. Kuzuoka H, Yamashita J, Yamazaki K, Yamazaki A (1999) Agora: a remote collaboration system that enables mutual monitoring. CHI'99 extended abstracts on Human factors in computing systems 190–191. https://doi.org/10.1145/632716.632836

50. Lee SB, Shin IY, Ho YS (2011) Gaze-corrected view generation using stereo camera system for immersive videoconferencing. IEEE Trans Consum Electron 57(3):1033–1040. https://doi.org/10.1109/TCE.2011.6018852

51. Gupta K, Lee GA, Billinghurst M (2016) Do you see what I see? The effect of gaze tracking on task space remote collaboration. IEEE Trans Visual Comput Graphics 22(11):2413–2422. https://doi.org/10.1109/TVCG.2016.2593778

52. Higuch K, Yonetani R, Sato Y (2016) Can eye help you? Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems 5180–5190. https://doi.org/10.1145/2858036.2858438

53. Špakov O, Istance H, Räihä K J, Viitanen T, Siirtola H (2019) Eye gaze and head gaze in collaborative games. Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications 1–9. https://doi.org/10.1145/3317959.3321489

54. 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(2):153–169. https://doi.org/10.1093/iwcomp/iwaa012

55. Pfeuffer K, Mayer B, Mardanbegi D, Gellersen H (2017) Gaze+ pinch interaction in virtual reality. Proceedings of the 5th symposium on spatial user interaction 99–108. https://doi.org/10.1145/3131277.3132180

56. Bai H, Sasikumar P, Yang J, Billinghurst M (2020) A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. Proceedings of the 2020 CHI conference on human factors in computing systems 1–13. https://doi.org/10.1145/3313831.3376550

57. Greenberg S, Gutwin C, Roseman M (1996) Semantic telepointers for groupware. Proceedings sixth Australian conference on computer-human interaction. IEEE, 54–61. https://doi.org/10.1109/OZCHI.1996.559988

58. Duval T, Nguyen TTH, Fleury C, Chauffaut A, Dumont G, Gouranton V (2014) Improving awareness for 3D virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues. J Multimodal User Interface 8(2):187–197. https://doi.org/10.1007/s12193-013-0134-z

59. Elvezio C, Sukan M, Feiner S, Tversky B (2015) [POSTER] Interactive visualizations for monoscopic eyewear to assist in manually orienting objects in 3D. 2015 IEEE International Symposium on Mixed and Augmented Reality. IEEE 180–181. https://doi.org/10.1109/ISMAR.2015.54

60. Wang Z, Bai X, Zhang S, He W, Wang P, Zhang X, Yan Y (2020) SHARIdeas: a visual representation of intention sharing between designer and executor supporting AR assembly. https://doi.org/10.1145/3415264.3431858

61. Piumsomboon T, Dey A, Ens B, Lee G, Billinghurst M (2017) [POSTER] CoVAR: mixed-platform remote collaborative augmented and virtual realities system with shared collaboration cues. 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct). IEEE, 218–219. https://doi.org/10.1109/ISMAR-Adjunct.2017.72

62. Hui W, Dong X, Guanghong D, Linxuan Z (2007) Assembly planning based on semantic modeling approach. Comput Ind 58(3):227–239. https://doi.org/10.1016/j.compind.2006.05.002

63. Tan J, Liu Z, Zhang S (2001) Intelligent assembly modeling based on semantics knowledge in virtual environment. Proceedings of the Sixth International Conference on Computer Supported Cooperative Work in Design (IEEE Cat. No. 01EX472). IEEE, 568–571. https://doi.org/10.1109/CSCWD.2001.942326

64. Jacob R J K (1990) What you look at is what you get: eye movement-based interaction techniques. Proceedings of the SIGCHI conference on Human factors in computing systems 11–18. https://doi.org/10.1145/97243.97246

65. Brooke J (1996) SUS-A quick and dirty usability scale. Usability Evaluation Industry 189(194):4–7. https://doi.org/10.1201/97814 98710411-35

66. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol North-Holland 52:139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

67. Kim S, Billinghurst M, Kim K (2020) Multimodal interfaces and communication cues for remote collaboration. J Multimodal User Interface 14:313–319. https://doi.org/10.1007/s12193-020-00346-8

68. Sutera J, Yang M C, Elsen C. (2014). The impact of expertise on the capture of sketched intentions: perspectives for remote cooperative design. In Cooperative Design, Visualization, and Engineering: 11th International Conference, CDVE 2014, Seattle, WA, USA, September 14–17, 2014. Proceedings 11 (pp. 245–252). Springer International Publishing. https://doi.org/10.1007/978-3-319-10831-5_36

69. Doswell J T, Skinner A (2014). Augmenting human cognition with adaptive augmented reality. In Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems: 8th International Conference, AC 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014. Proceedings 8 (pp. 104–113). Springer International Publishing. https://doi.org/10.1007/978-3-319-07527-3_10

70. D'Angelo S, Gergle D (2016). Gazed and confused: understanding and designing shared gaze for remote collaboration. In Proceedings of the 2016 chi conference on human factors in computing systems (pp. 2492–2496). https://doi.org/10.1145/2858036.2858499

71. Patterson RE, Blaha LM, Grinstein GG, Liggett KK, Kaveney DE, Sheldon KC, Havig PR, Moore JA (2014) A human cognition framework for information visualization. Comput Graph 42:42–58. https://doi.org/10.1016/j.cag.2014.03.002

72. Tang A, Boyle M, Greenberg S (2003) Display and presence disparity in Mixed Presence Groupware. https://doi.org/10.11575/PRISM/30675