



A novel AR remote collaborative platform for sharing 2.5D gestures and gaze

Yue Wang¹ · Peng Wang¹ · Zhiyong Luo¹ · Yuxiang Yan²

Received: 25 June 2021 / Accepted: 15 January 2022 / Published online: 21 January 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Although sharing gestures and gaze can improve AR remote collaboration, most current systems only enable collaborators to share 2D or 3D gestures, and the unimodal HCI interface remains dominant. To address this problem, we describe a novel remote collaborative platform based on 2.5D gestures and gaze (2.5DGG), which supports an expert who collaborates with a worker (e.g., during assembly or training tasks). We investigate the impact of sharing the remote site's 2.5DGG using spatial AR (SAR) remote collaboration in manufacturing. Compared to other systems, there is a key advantage that it can provide more natural and intuitive multimodal interaction based on 2.5DGG. We track the remote experts' gestures and eye gaze using Leap Motion and aGlass, respectively, in a VR space displaying the live video stream of the local physical workspace and visualize them onto the local work scenario by a projector. The results of an exploratory user study demonstrate that 2.5DGG has a clear difference in performance time and collaborative experience, and it is better than the traditional one.

Keywords Remote collaboration · Physical tasks · Augmented reality · Multimodal interaction · Gestures and gaze

1 Introduction

Remote collaboration enables local workers and remote experts to collaborate in real time without geographic restrictions. There are many such collaboration scenarios in our lives, for example, in telehealth, remote technical support, teleducation, and in providing emergency repair or training [1–3]. Thus, remote collaboration has become an essential part of most professional activities in particular during the COVID-19 outbreak. Moreover, augmented reality (AR) can be used as an effective tool in the COVID-19 epidemic owing to its features such as a powerful visualization tool, and annotation by providing real-time descriptions in the virtual world [4]. Drawing on this background, we concentrate on AR remote collaboration that can enable the guidance from an expert to a novice by means of 2.5D gestures and gaze (2.5DGG).

Over the years, the method of adding annotations on a video (POINTER) [5–8], gaze [9, 10], and gestures [11, 12] has been successfully applied in remote collaboration. However, when using these systems, users must wear head-mounted displays (HMDs), which is unsuitable for some tasks. In addition, other systems have to use a 2D monitor in front of local workers to show instructions. This will produce fractured ecologies having negative impacts on the performance [7]. In general, users in a remote site often use a 2D Windows-Icon-Menu-Pointer (WIMP).

Multimodal human–computer interaction can endow systems with more compelling interactive experiences. It can improve efficiency and decision making that follows from using each modality for a specific appropriate task, increase the likelihood of remote communication proceeding smoothly address answer specific problems about the collaborative task (e.g., which part is to be taken in the complex background), and support the naturalness and perceptibility of human–human communication during remote collaboration. Despite the evident potential advantages of multimodal interaction, the knowledge of how to actually serve AR remote collaboration is still in its very preliminary stages. Therefore, we take full advantage of combining gestures and gaze in order to develop a novel multimodal interaction for an AR remote collaborative platform.

✉ Peng Wang
ilovemymfandb@163.com

¹ School of Advanced Manufacturing Engineering,
Chongqing University of Posts and Telecommunications,
Chong Qing 400065, China

² Cyber-Physical Interaction Lab, Northwestern Polytechnical
University, Xi'an 710072, China

This research is motivated by earlier research [13–16] and builds upon it by combining 2.5D gestures, gaze, and SAR. Hence, the contributions of the present study fall into three areas:

1. We propose a novel 2.5DGG-based interactive interface for SAR remote collaborative work.
2. We conduct the first formal user study exploring the use of 2.5DGG for SAR remote collaboration in industry.
3. We offer a good example of using SAR and multimodal interaction based on 2.5DGG for SAR remote collaborative work.

The rest of the paper is organized as follows. First, we summarize the related work in this area, then introduce our proposed 2.5DGG platform. Next, we report and discuss a formal user study. Finally, we draw conclusions and describe the limitations of the study, while suggesting how they could be addressed in future work on this topic.

2 Related works

2.1 The POINTER interface

In remote collaboration, video-based cues have become a fundamental element. Although they achieve some good results successfully, they also lose some important cues such as body language and gestures. Therefore, various approaches have been proposed to address these disadvantages. Fussell et al. [17] demonstrated the effects of POINTER, and the recent studies [5, 6] have also shown that this method can greatly improve performance. Despite the fact that the POINTER interface indeed facilitates remote collaboration, the remote expert still primarily depends on a 2D WIMP interface. Most importantly, it is limited in the ability to convey gestures and gaze.

2.2 Sharing gaze and gestures

Recently, researchers have developed new methods that use eye gaze and gestures in remote collaborative work. Brennan et al. [18] proved that using gaze cues could improve performance for users who could understand their partners' intentions in a visual search task. Gupta et al. [9] demonstrated the impact of sharing gaze cues in remote collaboration on a Lego construction, and showing that the gaze-based cues could enhance co-presence. However, they shared the gaze-based cues from the local site. Akkil et al. [14] developed a SAR remote collaborative system called GazeTorch, which provides the remote user' gaze awareness for the local user. Recently, continuing in this exploratory direction, they investigated the influence of gaze in video-based

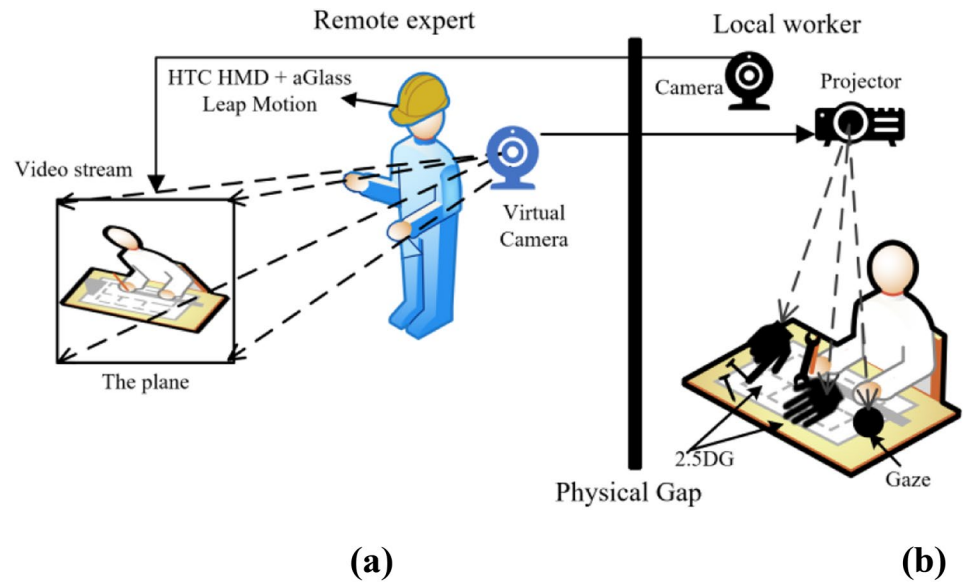
remote collaboration for a physical task, compared with the POINTER interface (e.g., a mouse-controlled pointer) [16]. Similar work was carried out by Higuchi et al. [19] and Wang et al. [20–22].

With the continuous innovation in gesture recognition, Sun et al. [23] designed a new collaborative platform, namely, OptoBridge, which can support sharing gestures. They conducted a user study to qualitatively and quantitatively evaluate the effects of different viewpoints on task performance and subjective feelings in AR remote collaboration. Kirk et al. [24] found that gestures-based cues have a positive impact on collaborative behavior. Moreover, they observed that shared gestures have some advantages, which includes supporting rich and free expression, decreasing the costs of word interpretation, improving the user experience of natural interaction, and bringing about better task performance. However, these systems used a flat graphical user interface, and the remote expert's interaction interface has a fractured ecology which may increase the workload. Recently, Amores et al. [25] proposed an immersive mobile remote collaborative system: ShowMe, which allows a remote user to provide assistances for a novice user on physical tasks using audio, video, and 3D gestures. Similar to this concept is the system BeThere [26]. The drawback of this approach is that the whole device is too heavy, thus forcing the user to use a monopod [25]. Wang et al. [11] proposed an MR-based telepresence system for facilitating remote medical training. Nevertheless, it only supports sharing four commonly used gestures. Kim et al. [27] investigated how two different kinds of factors (i.e., hands-in-air style and hands-on-target) influence AR collaboration for object selection and object manipulation in terms of performance and mental effort. In these researches, the unimodal interaction is still dominant, which easily leads to user fatigue over a long period of time. Moreover, local workers have to wear HMD. This may create some inconvenience for a certain task in industry specifically when operating some large or heavy parts. In order to overcome this shortcoming, Wang et al. developed a SAR remote collaborative platform supporting the sharing of 2.5D gestures [15] and proposed an mixed reality (MR) remote collaborative system sharing 3D gestures and CAD models [28].

2.3 Multimodal interaction

In AR/MR remote collaboration, much researches have focus on multimodal interaction. Higuchi et al. [19] studied the effect of sharing helper's 2D gestures and gaze with the local worker. In this research, the remote instructions based on gestures and eye gaze fused on a 2D monitor screen that shows a physical scenario and then visualized onto the local worker side by a projector or an HMD. However, it is necessary for the remote user to gesture at the desk when

Fig. 1 The prototype 2.5DGG platform framework. (a) The remote user's workspace and setups. (b) The local user's workspace and setups. The prototype becomes the 2.5DG platform framework when the gaze is disabled



watching the 2D WIMP monitor in front of users. This may lead to a fractured ecology, thus increasing the work load [7]. Moreover, the eye-tracking information will be lost, especially when the remote user needs to keep a close watch on movements of gestures in some cases. Therefore, it may be difficult for the remote user to synchronize the information between the video and other non-verbal cues. To overcome this shortcoming, we propose a novel SAR remote collaborative platform combining the merits of VR and AR to support the multimodal interaction based on 2.5DGG. Additionally, our platform can support collaboration among users by overcoming the shortcomings of separating gestural instructions and the communication space, and lack of some important communication cues (e.g., speech, gestures, and gaze.), which are instead used with the same ease as during collocated collaborative work. Recently, Bai et al. [29] conducted a formal user study on MR remote collaboration sharing gaze and gesture. Our work is similar to this study, but we focus on the AR site, while eliminating the HMD in order to improve the user experience of the local collaborator.

2.4 Summary

Three conclusions can be drawn from the research discussed above. (1) The remote interface is mainly a 2D WIMP or an HMD, (2) the interface in the local site can be classified into three types: 2D WIMP-, HMD-, or spatial AR(SAR)-based interface, and (3) there have been few previous examples of AR/MR remote collaborative systems using multimodal interaction. We find no earlier research that supports multimodal interaction based on 2.5DGG for SAR remote collaboration, and that has conducted user studies. Therefore, our research fills this gap

by evaluating a novel remote collaborative platform sharing 2.5DGG to support effective communication in tightly coupled physical tasks.

3 Our approach

In this section, we provide an overview of the prototype system, and we introduce the key technologies and implementation details.

3.1 Basic concepts

Definition 1. 2.5DGG (2.5D gestures + gaze). In our work, remote users provide instructions using 3D gestures and gaze in a VR device attached with aGlass and Leap Motion. Then, the gestures and gaze are projected onto the local workspace. Therefore, we define this concept as 2.5DGG as a way to distinguish it from other research (see Fig. 1).

Definition 2. 2.5DG (2.5D gestures). To evaluate our work, we implemented the new application for SAR remote collaboration on physical tasks [15, 19]. Their [15, 19] remote collaborative platform supports sharing 2D gestures and 2.5D gestures, respectively. Wang et al. [15] overcame two key disadvantages in their research [19]. They transformed 2D gestures into 2.5DG in a local site and fused the gestural instructions space and the communication space in the remote site. Therefore, we define this concept as 2.5DG (see Fig. 2). This system only moves the eye-tracker (aGlass) of the 2.5DGG system.

Fig. 2 A Leap Motion, aGlass, and an HTC HMD



3.2 Implementations

Our system includes the following hardware: an aGlass, a desktop computer, a Leap Motion, a camera, an HTC Vive, and a projector (see Figs. 2 and 3). Additionally, it was implemented on Unity3D 2018.4.30f1, Leap Motion Assets 4.3.3, aGlass-DKII, and OpenCV2410.

Figure 1 shows the prototype system diagram. The system is running on a desktop PC (Intel Core i7-10700 K CPU 3.80 GHz, 16 GB DDR4 RAM, NVIDIA GeForce RTX 2070 SUPER), and it has a local client and a remote client. The HTC VR device, Leap Motion, and aGlass were deployed on the remote user's side. We rendered the video from the local site on a plane in the Unity3D scene; thus, the remote expert can know the local situation in real-time in the HMD. The position of the plane should be still after calibration, and the VR user can move freely. Nevertheless, the VR users have to stand in front of the plane so that the 2.5DGG-based guidance information can be projected on the virtual plane in the VR environment. Moreover, we transformed the scene of the virtual camera in the 3D VR space to the local projector.

3.3 Technical challenges

In our current study, we simplified the system to concentrate on the core elements using co-located collaboration to emulate remote collaboration, as in many prior related systems [9, 10, 12, 14, 15, 19, 30] and in several publications reported in the review of related work [1]. There were some technologies that we needed to further investigate. The first involves dynamic gesture and gaze recognition and visualization. We wanted to explore how multimodal interaction based on gestures and gaze influences remote collaboration. The algorithm of the recognition of dynamic gestures and eye gaze is a sound foundation for our proposed research. To address these problems, we chose Leap Motion in order to achieve a good dynamic recognition. Moreover, we tracked eye gaze using aGlass. Thus, the key is to share the gestures and eye gaze with the local worker. The framework of gaze visualization will be introduced in the next subsection. The second technology involved calibration which will also be introduced in the subsection below on projector-camera pair calibration.

3.3.1 Gaze visualization

We used the eye tracker of aGlass to track eye gaze. More information about the aGlass is available on its website.¹ The key steps of eye gaze visualization are as follows. The first step is calibration. User calibration should be done for each participant individually, and we used the default process (9-point calibration). Second, it is necessary to obtain the gaze coordinate. This is the gaze point mapped onto a plane at a specific location. The third step is the gaze visualization. The algorithm was developed to show the ray along the direction of straight ahead from the remote VR user's eye coordinate. Then, by collision detection, the system can obtain the intersecting point between the ray and virtual objects and show the point with the red ball.

3.3.2 Projector-camera pair calibration

Before our prototype system can be used online, it requires some preparatory steps to ensure that the 2.5DGG-based instructions project to the right place. The detailed process for calibration was carried out referring to Wang et al. [21].

4 User study

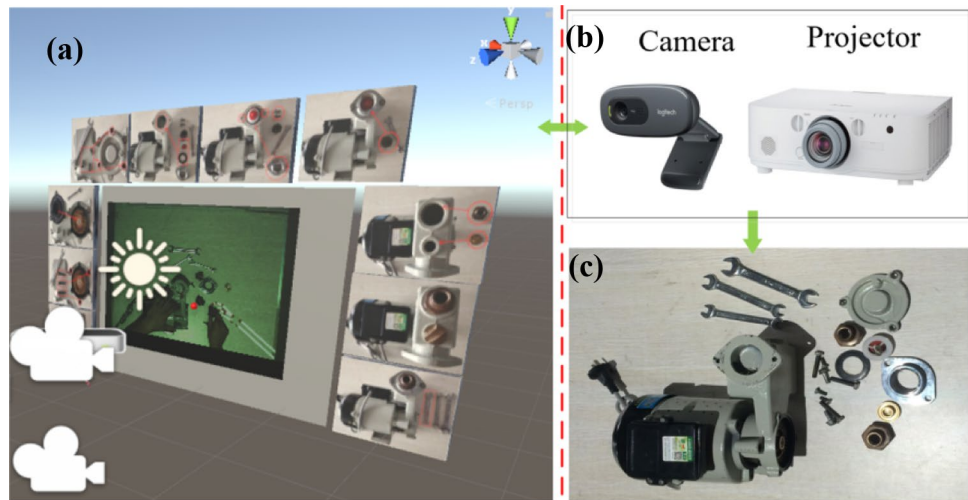
To evaluate the proposed 2.5DGG platform, we conducted an initial user study. The main purpose was to test the usability of the platform and the user experience.

4.1 Experiment design

To evaluate the prototype, we conducted a within-subjects pilot study by comparing two different conditions: the 2.5DGG interface and the 2.5DG interface. The physical task is to assemble a pump (see Fig. 4). Figure 4 shows remote collaboration using the 2.5DGG and 2.5DG interface, respectively. To encourage collaboration and achieve better counter-balance experimental conditions, we added four constraints. Firstly, to confirm the objectivity of the experimental data, we included a pair of participants who were not familiar with

¹ www.aglass.com.

Fig. 3 The prototype platform configurations. (a) is the remote user’s view using VR HMD, while (b) and (c) are the local user’s workspace and setups; (b) is the projector and camera and (c) shows the parts to be assembled



each other. Secondly, the local users began the clock when all participants were ready and then stopped the clock when the assembly is completed. Thirdly, the local user should take only one part at a time from the part’s area to the main workspace with the remote expert’s guidance. Finally, during assembling, the parts should be tightened first by hand and then using the wrench to tighten them further.

For our study, we formulated two hypotheses:

H1: Performance. Two different interfaces have a significant difference on the task performance, and the 2.5DGG interface takes less time and decreases operational mistakes.

H2: User Interaction. The 2.5DGG system is more intuitive, natural, and flexible to express instructions than the 2.5DG interface.

4.2 Demographics and procedure

We recruited twenty-four participants from 22 to 28 years old (15 males, 9 females, average age 24.4, SD 1.23). All participants had some experience using video streaming interfaces (e.g., Wechat) and limited AR/MR experience. Their university majors are mechanical manufacturing, robot

engineering, industrial design, and mechatronics engineering. Due to the COVID-19 pandemic, we provided participants with strict safety protections during the entire process of the study. The VR HMD and the parts were carefully disinfected with paper towels, and volunteers’ hands were sanitized before each session.

To counterbalance the learning effect, the order of conditions was arranged following a Latin square sequence. To achieve consistency in training, participants have to personally assemble the pump according to the assembly processes at the stage of training. Moreover, during the experiment, we provided the remote users with some photographs of the key assembly process (see Figs. 3a and 4c) for they could better describe the process to the local worker. For each condition, all users having the experience of remote collaborative work give priority to being as remote users. When the task is completed, VR users should guide the local SAR users to assemble the pump according to the assembly process.

At the end of each condition, the performance time was recorded, and users were asked to fill in a questionnaire in terms of the quality of the telecollaboration

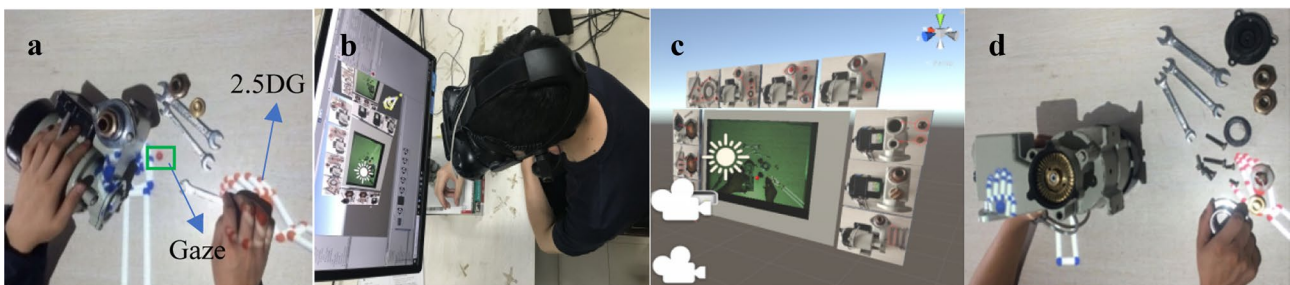


Fig. 4 The remote collaborative interface. (a)(b) are the 2.5DGG interface; (c) is the VR users’ view in the HTC HMD; in (a) the local user assembles the pump according to the remote user’s 2.5DGG; in

(b) the remote expert highlights the target parts by 2.5DGG; in (d) the local user assembles the pump according to the remote user’s 2.5DGG

Table 1 Evaluation of the collaborative experience

Q*	Statement: 1 (I entirely disagree) to 7 (entirely agree)
Q1	I was able to focus on the work actively
Q2	Information from the partner was helpful
Q3	I was able to express myself clearly
Q4	I can help my partner when he/she needed assistance
Q5	The manipulation is more natural and intuitive
Q6	I felt very confident during the assembling

(see Table 1). The questionnaire was designed based on related works [9, 31] but slightly modified to focus on our research aim. Finally, all users ranked the two different conditions.

4.3 Results

4.3.1 Performance

We compared the performance in the two conditions. Figure 5 shows task completion time for the two different interfaces. Descriptive statistics showed that on average the users took more time using the 2.5DG (Mean = 268.25 s, SD = 9.08 s) interface than our 2.5DGG (Mean = 254.17 s, SD = 8.94 s). The paired *t*-test ($\alpha = 0.05$) revealed that there was a statistically significant difference between the two different interfaces on performance time ($t(11) = 4.329$, $p = 0.001$).

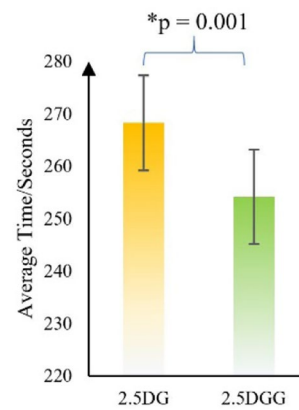
4.3.2 Error evaluation

To investigate the impact of 2.5DGG, we recorded the number of wrong operations (e.g., the incorrect work piece taken (IPT) and the incorrect motion processed (IMP)). Figure 6 shows the number of errors for local users. We conducted the Wilcoxon signed rank test ($\alpha = 0.05$) and determined that there is a difference in IPT ($Z = -2.471$, $p = 0.013$), but not IMP ($Z = -0.849$, $p = 0.396$) using the two systems. The results show the gaze-based cues can help local users accurately specify the part and position based on 2.5D gestures.

4.3.3 Collaborative experience

To compare the ratings between the two different interfaces, we conducted the Wilcoxon signed rank test ($\alpha = 0.05$). The statistical results are shown in Fig. 7.

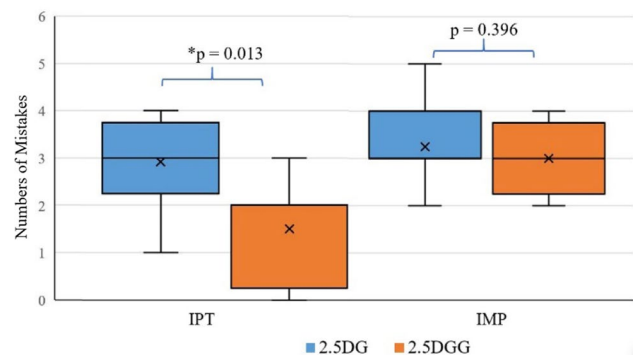
The result for local workers is shown in Fig. 7a. The 2.5DGG-based cues had a significant effect for local workers in terms of Q2 and Q6, but not of Q1, Q3, Q4, and Q5. In addition, the 2.5DGG interface had a significant

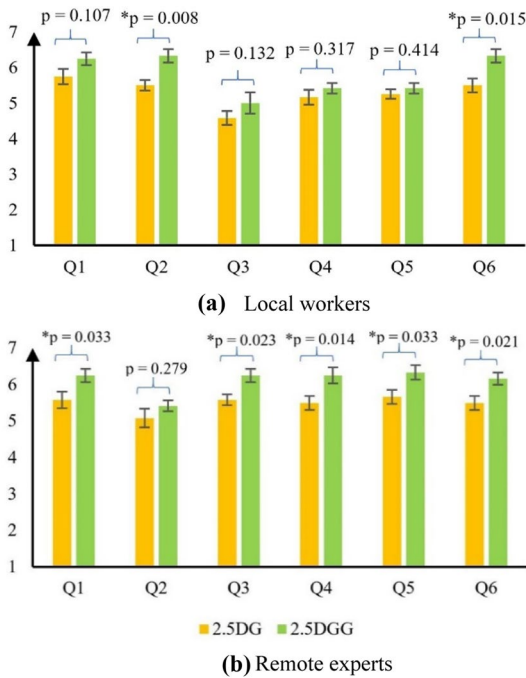
**Fig. 5** The time needed to assemble the water pump (error bar: \pm SD)

impact not only on guidance information (Q2, $Z = -2.640$, $p = 0.008$), but also on the sense of confidence (Q6, $Z = -2.428$, $p = 0.015$). Moreover, although the average of items Q1, Q3, Q4, and Q5 is generally greater for the 2.5DGG interface, there was no significant effect.

For VR side users (see Fig. 7b), a significant influence of the 2.5DGG factors was found in Q1, Q3, Q4, Q5, and Q6, but not Q2. This statistical result shows that, compared to the 2.5DG interface, the 2.5DGG interface had a significant impact on expressing ideas (Q3: $Z = -2.271$, $p = 0.023$; Q4: $Z = -2.460$, $p = 0.014$), and on sense of being focused ($Z = -2.126$, $p = 0.033$) and self-confidence ($Z = -2.309$, $p = 0.021$). Furthermore, the 2.5DGG condition is more natural and intuitive (Q5, $Z = -2.309$, $p = 0.021$).

The statistical results indicate that the 2.5DGG-based cues allow remote users to express themselves more easily and that these cues are also more understandable for local users. We believe that these findings can enhance collaborative efficiency to some extent. With reference to users' preferences, all users preferred the 2.5DGG condition.

**Fig. 6** The number of mistakes in assembly



Error bar: +/- SE, *: significant difference between the two interfaces

Fig. 7 Average rating of results of local users (a) and remote users (b) for items (1: entirely disagree, 7: entirely agree)

5 Discussion

In this study, we have demonstrated the value of multimodal interaction based on 2.5DGG for SAR remote collaborative assembly in manufacturing. The statistical analysis revealed that sharing 2.5DGG enhanced performance and collaborative experience.

At the end of each trial, we collected users’ feedback on how to enhance the prototype system. For remote users, all users perceived the impact of 2.5DGG, though some users were not fully aware that the shared gestures and gaze had influenced their behavior and made them more focused on task (Q1 in Fig. 7). The multimodal interaction based on

2.5DGG provided more natural and intuitive interaction than the 2.5DG condition (Q5 in Fig. 7b). Additionally, most VR users maintain that it is interesting to see their 3D hands and gaze in the HMD, while most SAR users also considered collaborating with the 2.5DGG interface to be a positive experience. The local users believed that they can more easily to understand their partners’ intentions (see Q2 in Figs. 7a and 8a–c). Thus, we conclude that all local/remote users preferred the 2.5DG condition for these reasons.

During the experiment, we found that gaze is a fast and precise way to point and specify the target part in a complex background (see Fig. 8d). Moreover, gaze can indicate the remote collaborator’s intentions (e.g., next instructions and the region of interest). Although the non-verbal cues (e.g., gestures and gaze) could decrease the amount of voice communication based on interviews, users always combined speech with gestures and gaze to provide explicit instructions. Specifically, 2.5DG and 2.5DGG can also improve communications by pointing and augmenting remote collaboration by means of deictic references.

More importantly, our proposed system paves the way for multimodal interaction in AR/MR remote collaboration, sharing the local workspace as shown in previous researches [5, 6] by using a 3D reconstruction in the local environment. Wang et al. [15] found that gesture-based instructions alone could lead to user fatigue over a long period of time. Our research could overcome this disadvantage to some degree by the multimodal interaction based on 2.5DGG.

6 Conclusions, limitations, and future work

In this research, we focused on exploring how 2.5DGG impacts remote collaboration on assembly tasks in manufacturing. Consequently, we proposed a multimodal interaction for AR/MR remote collaboration based on 2.5DGG. To the best of our knowledge, this research presents one of the first remote collaborative prototypes to provide 2.5DGG-based cues from a remote VR user to a local

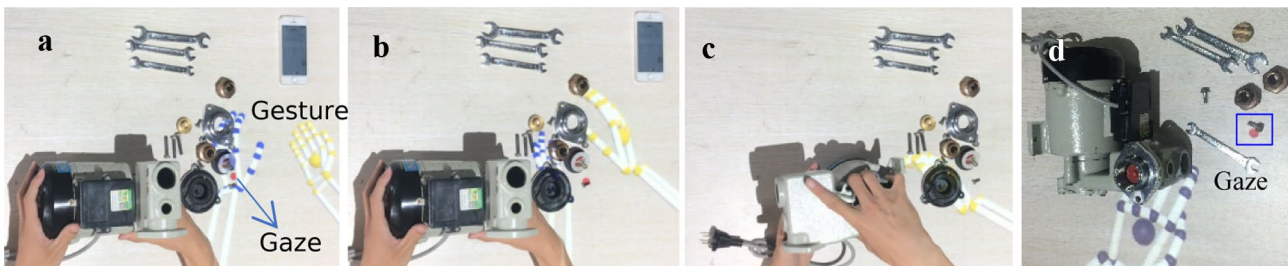


Fig. 8 The operations using the 2.5DGG interface: in (a), (b), and (c) the rotation is vertical to the desktop. The left hand must fix the left side of the pump, meanwhile the right hand must fix the right lower

side of the pump during operations; in (d) the eye gaze precisely specified the bolt. The 2.5DGG interface is more flexible because it supports gesture- and gaze-based instructions

SAR worker for remote collaboration in manufacturing. Furthermore, the exploratory results show that 2.5DGG-based cues could support hypothesis H1 (performance) and H2 (interaction).

There are some aspects of our research to be improved in future work based on this prototype. First, we will explore the use of multiple projectors and cameras to address occlusion problems. In the current work, we used only one fixed projector; therefore, sometimes the local user's arm occluded the objects, which makes it difficult for remote VR users to find a certain object and project the correct place. Second, we studied only one–one remote collaboration. In the future, we would like to investigate one-many VR/AR remote collaboration (e.g., one remote expert and two or more local workers) on assembly training using different devices such as PCs and mobile phones. Furthermore, although our results are encouraging and indicate that our research is heading in the right direction, it would be interesting to investigate whether the interface could support haptic feedback and a multiscale MR telepresence interface as shown by recent research [32–34].

Acknowledgements This research was sponsored by the national key research and development program (2020YFB1710300), Aeronautical Science Fund of China(2019ZE105001), and General Project of Chongqing Natural Science Foundation(cstc2019jcyj-msxmX0530). We would like to express our appreciation to the anonymous reviewers for their constructive suggestions for enhancing this paper. We also thank Huizhen Yang and Qichao Xia, who help improve the English language and put forward many constructive suggestions during the manuscript revision process. We would also like to thank members of Chongqing University of Posts and Telecommunications and Northwestern Polytechnical University for their participation in the experiment.

Author contribution Yue Wang and Peng Wang finished the draft; Yue Wang and Zhiyong Luo improved the paper. Moreover, Peng Wang and Yuxiang Yan designed the AR prototype system.

Funding This research was financially sponsored by the National Key Research and Development Program (2020YFB1710300), the Aeronautical Science Fund of China(2019ZE105001), and the General Project of Chongqing Natural Science Foundation (cstc2019jcyj-msxmX0530).

Availability of data and materials All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval Not applicable.

Consent to participate All participants consented to participate in the experiment.

Consent to publish All participants consented to the publication of the related data, and all authors consented to publishing the research.

Competing interests The authors declare no competing interests.

References

1. Wang P, Bai X, Billingham 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 Integr Manuf* 72(4):102071
2. Wang P, Zhang S, Billingham 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* 36:1715–1738
3. Calandra D, Cannavò A, Lamberti F (2021) Improving AR-powered remote assistance: a new approach aimed to foster operator's autonomy and optimize the use of skilled resources. *Int J Adv Manuf Technol* 114:3147
4. Asadzadeh A, Samad-Soltani T, Rezaei-Hachesu P (2021) Applications of virtual and augmented reality in infectious disease epidemics with a focus on the COVID-19 outbreak. *Inform Med Unlocked* 24:1–8
5. Anton D, Kurillo G, Bajcsy R (2018) User experience and interaction performance in 2D/3D telecollaboration. *Futur Gener Comput Syst* 82:77
6. Anton D, Kurillo G, Yang AY, Bajcsy R (2017) Augmented telemedicine platform for real-time remote medical consultation 77
7. Gurevich P, Lanir J, Cohen B (2015) Design and implementation of TeleAdvisor: a projection-based augmented reality system for remote collaboration [J]. *Comput Support Coop Work (CSCW)* 24(6):527–562
8. Fakourfar O, Ta K, Tang R, Bateman S, Tang A (2015) Stabilized annotations for mobile remote assistance [C]. *Proceedings of the 2016 CHI conference on human factors in computing systems. Comput Support Coop Work (CSCW)* 24:1548–1560
9. Gupta K, Lee GA, Billingham M (2016) Do you see what I see? The effect of gaze tracking on task space remote collaboration [J]. *IEEE Trans Visual Comput Graphics* 22(11):2413–2422
10. D'Angelo S, Gergle D (2018) An eye for design: gaze visualizations for remote collaborative work [C]. *Proceedings of the CHI Conference on Human Factors in Computing Systems* 2018:1–12
11. 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 [J]. *Sensors* 17(10):2294
12. Huang W, Alem L, Tecchia F, Duh HB (2018) Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration [J]. *J Multimodal User Interfaces* 12(2):77–89
13. Kytö M, Ens B, Piumsomboon T, Lee GA, Billingham M (2018) Pinpointing: precise head-and eye-based target selection for augmented reality [C]. *Proceedings of the CHI Conference on Human Factors in Computing Systems* 2018:1–14
14. Akkil D, James JM, Isokoski P, Kangas J (2016) GazeTorch: enabling gaze awareness in collaborative physical tasks [C]. *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems* 2016:1151–1158
15. Wang P, Zhang S, Bai X, Billingham M, He W, Sun M, Ji H (2019) 2.5 DHANDS: a gesture-based MR remote collaborative platform [J]. *Int J Adv Manuf Technol* 102(5): 1339–1353
16. Akkil D, Isokoski P (2018) Comparison of gaze and mouse pointers for video-based collaborative physical task [J]. *Interact Comput* 30(6):524–542
17. 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 [J]
18. Brennan SE, Chen X, Dickinson CA, Neider MB, Zelinsky GJ (2008) Coordinating cognition: the costs and benefits of shared gaze during collaborative search [J]. *Cognition* 106(3):1465–1477
19. Higuch K, Yonetani R, Sato Y (2016) Can eye help you? Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks [C]. *Proceedings of the CHI Conference on Human Factors in Computing Systems* 2016:5180–5190

20. Wang P, Zhang S, Bai X, Billingham M, He W, Wang S, Zhang X, Du J, Chen Y (2019) Head pointer or eye gaze: which helps more in MR remote collaboration? [C]. 2019 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE 1219–1220.
21. Wang P, Zhang S, Bai X, Billingham M, He W, Sun M, Chen Y, Lv H, Ji H (2019) 2.5DHANDS: a gesture-based MR remote collaborative platform. *Int J Adv Manuf Technol* 102(5):1339–1353
22. Wang P, Bai X, Billingham 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
23. Sun H, Liu Y, Zhang Z, Wang Y (2018) Employing different viewpoints for remote guidance in a collaborative augmented environment [C]. Proceedings of the sixth international symposium of Chinese CHI 64–70
24. Kirk D, Crabtree A, Rodden T (2005) *Ways of the hands* [C]. ECSCW Springer. Dordrecht 2005:1–21
25. Amores J, Benavides X, Maes P (2015) Showme: a remote collaboration system that supports immersive gestural communication [C]. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems 1343–1348
26. Sodhi R S, Jones B R, Forsyth D, Bailey B P, Maciocci G (2013) BeThere: 3D mobile collaboration with spatial input [C]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 179–188
27. Kim S, Jing A, Park H, Lee GA, Huang W (2020) Hand-in-air (HiA) and hand-on-target (HoT) style gesture cues for mixed reality collaboration [J]. *IEEE Access* 8:224145–224161
28. Wang P, Bai X, Billingham M, Zhang S, Wei S, Xu G, He W, Zhang X, Zhang J (2020) 3DGAM: using 3D gesture and CAD models for training on mixed reality remote collaboration. *Multimed Tools Appl* 1–26
29. Bai H, Sasikumar P, Yang J, Billingham M (2020) A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing [C]. Proceedings of the CHI conference on human factors in computing systems 2020:1–13
30. Wang P, Zhang S, Bai X, Billingham M, He W, Zhang L, Du J, Wang S (2018) Do you know what I mean? An MR-based collaborative platform [C]. 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE 77–78
31. Harms C, Biocca F (2004) Internal consistency and reliability of the networked minds measure of social presence [J]. Seventh Annual International Workshop: Presence 2004. Valencia: Universidad Politecnica de Valencia
32. Günther S, Kratz S, Avrahami D, Mühlhäuser M (2018) Exploring audio, visual, and tactile cues for synchronous remote assistance [C]. Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference 339–344
33. Piumsomboon T, Lee GA, Ens B, Thomas BH, Billingham M (2018) Superman vs giant: a study on spatial perception for a multi-scale mixed reality flying telepresence interface [J]. *IEEE Trans Visual Comput Graphics* 24(11):2974–2982
34. Wang P, Bai X, Billingham M, Zhang S, Han D, Sun M, Wang Z, Lv H, Han S (2020) Haptic feedback helps me? A VR-SAR remote collaborative system with tangible interaction. *Int J Hum-Comput Int* 36(13):1242–1257

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