Mixed-Reality Communication System Providing Shoulder-to-shoulder Collaboration

Minghao Cai and Jiro Tanaka

Graduate School of Information, Production and Systems
Waseda University
Kitakyushu, Japan
Email: mhcai@toki.waseda.jp, jiro@aoni.waseda.jp

***Abstract*—In this paper, we propose a mixed-reality-based mobile communication system for two users placed in separate environments. The first is a remote user who physically travels to a shared environment with a mobile augmented reality setup, and the second is a local user who remains in another place while being immersed in a virtual reality view of the shared environment with the first user. The users are provided with a unique kind of collaboration, i.e., Shoulder-to-shoulder collaboration that simulates the users to walk shoulder-to-shoulder with viewing independence and bidirectional gesture communication. The major objective is to enhance co-located sensation. We introduce our prototype system as a proof of concept and perform evaluations of two user studies to verify system applicability and performance.**

Keywords–Remote collaboration; Shoulder-to-shoulder; Viewing independence; Gesture; Co-located Sensation.

I. INTRODUCTION

In recent years, remote communication has been extensively used at the workplace and in everyday life to increase productivity and to improve the performance of instant communication. The advantage in allowing users from different locations to communicate and collaborate as a team helps the remote communication system become a cost-effective and popular way that can help users to get an instant solution for problems.

Although commercial remote conferencing technologies are cost-effective and more immersive than traditional phone calls that use only voice, most of these systems mainly provide a mere capture of both the user's face and limited transition in terms of body language or the reference of ambiance, which also act as a great source of information [1]. When indulging in a physical collaborative task or conversation with context related to the surroundings, existing technologies offer limited ways for users to achieve effective gestural communication, as they tend to focus on face-to-face interaction experiences. When users wish to describe the objects or directions in a scene or show operations, use of hand gestures would be more understandable than mere voice.

Another problem is in the form of the camera used for real-time video capture. When using telecommunication systems with smartphones or tablets, users tend to switch between the front and back cameras or they might place the device in a fixed position to attain a wider range of view. In most cases, the camera needs to be moved around for the remote person to perceive the entire scene. Such constraints make it difficult for users to get a common perception or to feel connected with each other.

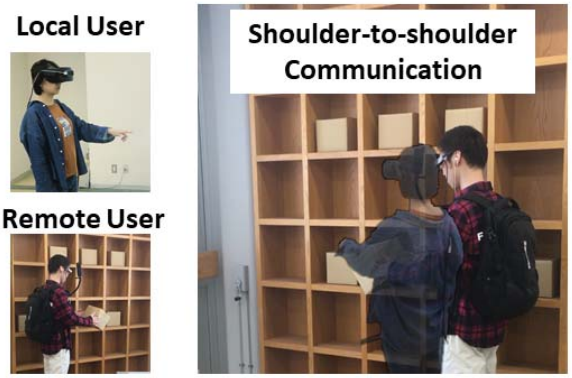

Figure 1. Shoulder-to-shoulder communication for two users

In this paper, we propose a solution to these problems in the form of our prototype that provides a mobile *shoulder-to-shoulder communication system* for using mixed-reality (MR) collaboration and communication. This unique type of communication can enhance the user-to-user interactions and co-located sensation between users.

The prototype is designed for use by two users who are in different locations (as shown in Figure 1). For convenience, we refer to the user who goes to a remote environment, which would be shared, as the remote user, and the other one who stays in a local indoor workspace and remotely views the shared world as the local user, even though the roles may be reversed. We try to offer both the users with a shared feeling that they are going shoulder-to-shoulder together using gesture communication. Wearing a head-mounted display (HMD) with a virtual reality (VR) experience, the local user perceives the remote environment with viewing independence, while the remote user wears a see-through smart glass for an augmented reality (AR) experience.

To address the existing problems, as mentioned earlier, we create the following design requirements for the shoulder-to-shoulder communication prototype:

1 Offer the local user an independent view of the remote environment with control of his or her own viewpoint.

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

2 The local user should be able to easily see the remote partner's action and the direction of attention.
3 Offer an appropriate visual representation of the local user so that the remote AR user is aware of the attention and a improvement of the understanding and fidelity of the remote communication.
4 Provide mutual free-hand gesture communication
5 Offer visual assistance cue to enhance user interactions

The main contributions of this work are as follows:

- The design of shoulder-to-shoulder collaboration and a software system that supports MR collaboration between two users.
- The implementation of a prototype as a proof of concept (POC) that includes mobile setup for the remote VR user and a wearable setup for the local AR user.
- An evaluation consisting of two user studies to test the usability of the proposed prototype and user performance.

In Section II, we introduce related works. In Section III, we introduce our shoulder-to-shoulder collaboration and the corresponding system design. In Section IV, we introduce the implementation of our prototype. In Section V, we describe the evaluation that consists of two user studies in which we compare our should-to-shoulder communication design against two comparative conditions and then, test the system in a practical scenario. In Section VI, we discuss potential applications. In Section VII, we draw our conclusion to this work.

II. Related Work

A. Remote Communication for Users Located in Different Places

Currently, it is not unusual to get instant contact with the use of commercial video conferencing systems (e.g., Skype and Cisco WebEx Conferencing). Most of these systems provide remote communication with a face capture feature from disparate locations, however, they do not allow users to reference a common physical ambient or share a co-presence feeling. Previous research has tried to address this limitation with different approaches [2] including projecting interface [3] and virtual reality interface [4].

Several pieces of research have made a lot of effort in working toward remote video communication techniques that aim at realizing a remote collaborative work experience among users in separate places [5, 6]. Some of these works tested depth sensors to extract and analyze body motions and interactions to support users to work in the same media space.

B. Remote Collaboration with Mixed Reality

Since the emergence of technology that supports remote communication [7, 8], researchers have started exploring remote collaboration with different degrees of user-to-user interactions. Reality is the user perception of the real environment. Introduced as a mix of both augmented reality and augmented virtuality (Figure 2 illustrates the reality continuum [9]), recently mixed-reality technique has been proven valuable for applications that involve a single user. It is believed that MR applications can provide users with a seamless combination of the virtual world and the real, physical world along with an enhancement of reality, which are the two major issues in traditional computer-supported collaborative work (CSCW). Researchers have started exploring the introduction of MR technology when constructing remote collaborations, where the VR representation of a local physical task environment is captured and shared with a remote-access collaborator who can view it in a separate remote place. Apart from verbal communication, the remote-access user is allowed to communicate with the local user with a certain degree of visual interactions.

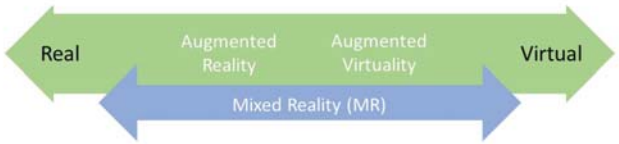

Figure 2. Virtuality continuum

Some previous works explored the use of pre-prepared 3D models to build virtual representations of the physical environment [10], or share 3D reconstructions of remote physical scenes on the user's desktop computer or mobile device [5]. Although these prior attempts with the use of static 3D reconstruction techniques can provide the spatial structure of the physical environment, they have limitations in terms of updating dynamical changes in the shared media space and moreover, in this case, the visual quality is typically inferior to a video image.

C. View Sharing in Remote Collaboration

A few researchers have tried using live video streams to share a view in an MR remote collaboration. Prior prototypes used handheld controls or touch screens to help a user create notes or draw annotations as visible cues in a 2D video stream [11–13]. Some system tried to introduce eye gaze or gesture [14] to improve communication efficiency. However, most of these works provide an egocentric viewpoint through a 2D video stream, in which the remote-access user's perspective was dependent on the motion of the camera capturing the surroundings.

Different approaches have been proposed to overcome the viewpoint limitation. Nuernberger et al. [15] demonstrated a system saving keyframes of the scene for later viewing. Fussell et al. [16] tried to place a camera fixed in the environment for remote collaboration. Some other attempt works explored utilizing remote presence and robotic techniques to offer remote-access users a certain control of the camera [17–20], but these approaches still have limitations with the field of view and delay in remote controlling the view.

Other researchers investigated using 360 panoramas [21] to help the remote-access user get a much larger field of view, or sharing panoramic images as an enhancement element of the 2D video streams in collaboration. In a demonstrated system, the researcher used a 360°camera to share the user's surroundings to a remote viewer who used mobile devices to

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

access an egocentric view [22]. They found the remote-access user had difficulties in communicating location and orientation information due to the lack of sharing gestures and other non-verbal communication cues.

D. Gesture Interaction

Hand gesture has been shown as an irreplaceable part for conversation, as it is treated as a cognitive visible awareness cue and provides rich contextual information that other body cues cannot reveal, which contributes significantly to a recipient's understanding [23, 24]. Over the past several years, some researchers have paid attention to support gestural interactions in a shared media space using different approaches. A study confirmed that over a third of the users' gestures in a collaborative task was performed to engage the other users and express ideas [25]. Kirk et al. [26] demonstrated the positive effect of using gestures and visual information in promoting the speed and accuracy in remote collaborative activities. Another work by Fussell et al. [27] demonstrated that users tend to rely more on visual actions than on speech in collaborative work.

E. Depth-based Gesture Recognition

Some researchers began to explore the idea of conveying gestures over a certain distance. A prior work [28] explored sharing live images of captured the arm action of one side's user on a remote shared tabletop screen for gesture collaboration. The gesture interaction in this work is still limited and the system only provides 2D images of hands or arms without any structural depth information. Several systems have captured users' hands in 3D and shared hand embodiments in a shared media space [29, 30]. However, these works require both local and remote users to remain within specific areas, which constrains the applications.

With the development of wearable devices and tracking sensors, some researchers have started exploring the use of a combination of depth cameras and head-mounted devices in experimental designs to realize remote collaboration in a reconstructed virtual-reality environment [30, 31]. These systems provide virtual hand gesture cues either captured with a depth camera [5] or represented by virtual hand models [31].

Previously, using depth-based gesture recognition, we built a remote sightseeing prototype that supported gestural communication to realize a gesture communication between two separated users [32, 33]. It was investigated by providing users with an approach to achieve a spatial navigation and direction guidance during mobile sightseeing. The positive evaluation results of this work encouraged us to support a mid-air gesture interaction for improvements to users' interactions in remote collaborations.

III. Shoulder-to-shoulder Collaboration

In this section, we introduce our proposed shoulder-to-shoulder collaboration and the system design that supports MR collaboration. This section consists of the following main aspects:

- A Overview of the prototype system
- B Shoulder-to-shoulder viewing independence
- C Shoulder-to-shoulder Gesture Communication
- D Tele-presence of the Local User's Head Motions
- E Virtual Pointing Assistance

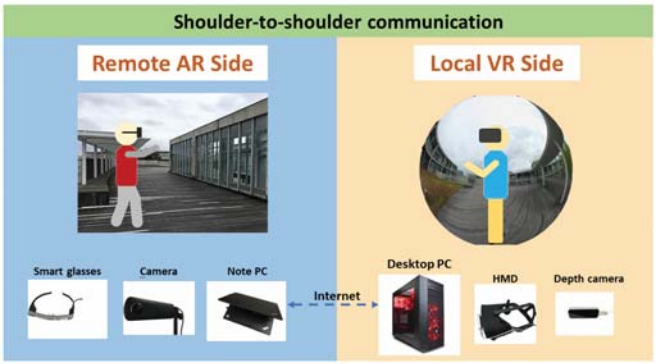

Figure 3. Prototype overview

A. Prototype overview

Figure 3 illustrates an overview of the prototype. Shoulder-to-shoulder communication is an MR collaboration that provides shoulder-to-shoulder viewing independence and shoulder-to- shoulder gesture communication between two users. A remote user wears AR smart glasses and carries a 360 camera for capturing the remote environment. The 360° view is shared with a local user via the Internet. The local user utilizes an HMD as the display to observe the remote view and attains an immersive VR feeling. A depth camera is used to capture the local user's hand gestures for mutual gesture interactions.

B. Shoulder-to-shoulder Viewing Independence

To capture and share the real-time remote environment, we choose a 360°camera that provides a high-resolution video with a range of 360°in both horizontal and vertical directions. Unlike previous view sharing systems, where the camera was usually put on the remote user's head or cheek [34], our chosen 360°camera is mounted on the remote user's shoulder using a holder. The real-time 360°video is streamed back to the local site via the Internet and displayed in the HMD worn by the local user.

As the camera is fixed to the shoulder, its orientation is prevented from being influenced by the remote user's head movements. The local user has independent control over the viewing direction that can be manipulated by the head movements. As shown in Figure 4, the local user can simply turn the head to naturally change the viewpoints. Using this design, the local user immerses in the virtual remote world and perceives a sensation of personally standing next to the remote user and viewing the same scene.

C. Shoulder-to-shoulder Gesture Communication

In our proposed system, we provide the users with an approach to achieve a bidirectional gesture interaction during mobile communication. On one hand, a shoulder-looking capture of the hand gestures of the remote user is included in the local user's virtual viewing. On the other hand, a pair of virtual hands based on the depth-based recognition reappear during the local user's gestures in the remote user's field of view.

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Figure 4. Independent control of the viewing direction for the local user

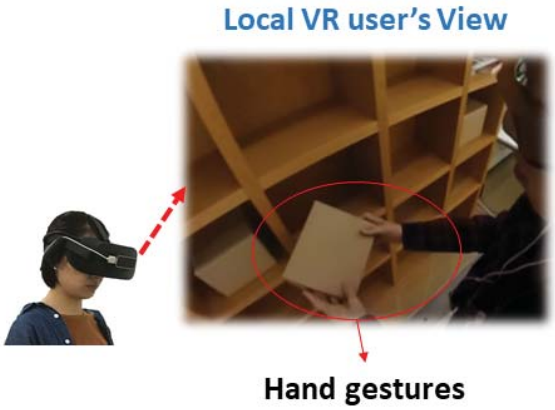

Figure 5. Local user's field of view: the remote user is making gestures

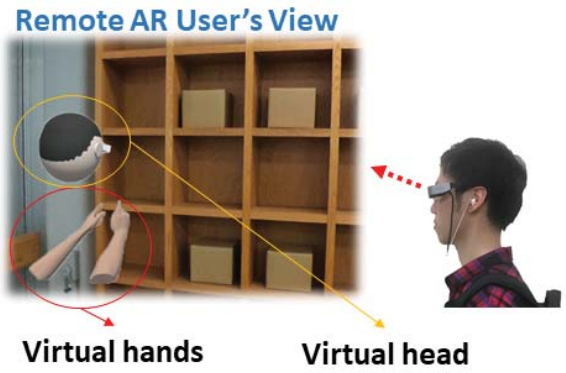

Figure 6. Remote user's field of view: the local user is making gestures. Red circle shows the virtual hands and yellow circle shows the virtual head representing the local user

1) Remote Gestures to Local User: As introduced in Section III-B, the local user has a 360°independent viewing of the remote world with a visual perspective obtained from the remote user's shoulder. This design allows the local user to see the remote user's hand gestures as well as the profile face. As shown in Figure 5, the local user simply looks leftward, and can directly see the remote partner performing hand gestures with an object (grabbing a box using the hands).

2) Local Gestures to Remote User: One of the important contributions of the proposed system is the reappearance of the local user's hand gestures in the remote world, as the local user is in a physically separate environment. We implement the hardware to extract the user's hand motion and the software to render it in the remote user's see-through smart glasses. Being considered as an accurate and convenient way, depth-based recognition has been used in current researches for hand motion extraction [29, 35]. A depth sensor is attached to the front side of the local user's HMD to extract a fine 3D structure data of both hands in real time. The local user can perform hand gestures without any wearable or attached sensors on the hands, which improve the freedom of hand motions and comfort. The system extracts the raw structure data with almost 200 frames per second with the help of the Leap Motion SDK [36]. We construct a pair of 3D hand models, which include the palms and the different finger joints. This pair of 3D hand models is matched with the latest hand structural data. Thereafter, the current reconstructed hands are sent to the remote side via the Internet and rendered in the remote user's AR smart glasses as an event to update the previous hands. Therefore, once the local user makes hand gestures, the models change to match the same ones, almost simultaneously appearing in the remote user's field of view as well (Figure 6).

D. Tele-presence of the Local User's Head Motions

As we aim to enhance a co-located sensation by improving the interaction between users, we try to help the users by letting them easily know where their partner is exactly looking. It would improve the efficiency of communication when the user tries to join the same field of view to find out common interests or initiate a discussion. As we introduced in Section III-B, the local user can easily tell the remote user's viewing direction in the virtual scene. As the local user is in a physically separated environment, we construct a virtual head model to show his/her head motions in the remote user's view.

A motion tracking sensor is used to extract the head motion that is used to rotate the virtual head model. In Figure 6, the

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

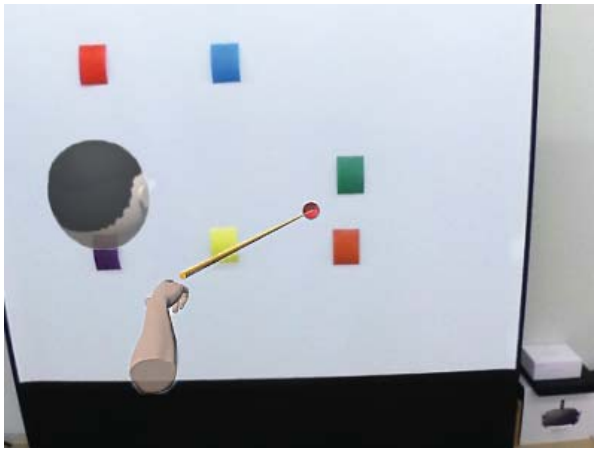

Figure 7. Remote user's view: Pointing cue for instructions

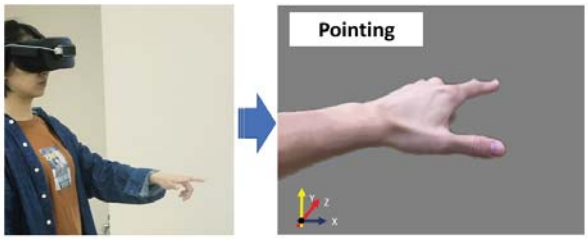

Figure 8. Zoomed in view of the pointing gesture

model present on the left side of the vision shows the remote user's precise facing direction.

E. Pointing Assistance

Previous research has shown that utilizing finger-pointing assistance can benefit cooperation and passing of instructions between users, especially when spatial information is involved in conversations [5].

In our shoulder-to-shoulder communication system, we allow the local user to use pointing assistance using fingers. The user performs a free hand pointing gesture that uses a virtual 3D arrow for showing the specific direction information in the remote user's view. This 3D arrow is treated as a spatial cue that assists a navigation or selection task during the communication (see Figure 7).

Our system uses a heuristic technique for gesture recognition. Using the depth sensor, our system can keep tracking the 3D structure of the user's hands including the different finger joints and can extract both the 3D position and orientation of the local user's fingers. The proposed system does not require calibration or precedent training. To activate the pointing technique, the user only needs to extend the thumb and index finger and keep the angle between them larger than the set threshold (see Figure 8).

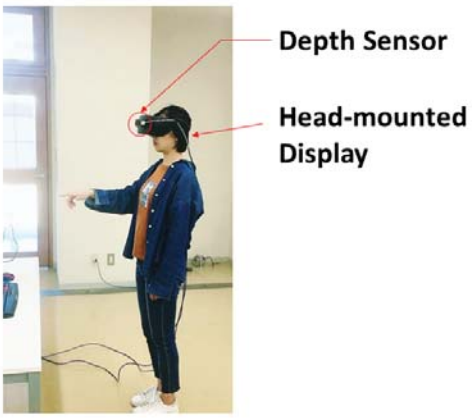

Figure 9. Local user's wearable setup: a Head-mounted Display with a depth sensor mounted on its front side

IV. Implementation

A. Hardware

Our system's hardware involves two parts: the local user side and the remote user side.

1) Local User's Side: The equipment in the local user's side includes wearable devices (see Figure 9) and a desktop PC. The desktop PC (Intel Core i5, RX480 Graphics Card, 8GB RAM) on the local user's side is used to analyze data and as an engine for the core system. We use Unity as the engine to render and process the incoming data from both the remote and local sides, as well as to generate the graphical user interface (GUI) for both users. The headset that we chose as the local user's head-mounted display uses a pair of low persistence OLED screens, that provide a 110°field of view (FOV) [37]. A point tracking sensor is used to provide six total degrees of freedom in terms of rotational and positional tracking of the head movements. For hand motion tracking, the depth sensor used is light enough and introduces a gesture tracking system with sub-millimeter accuracy [38].

2) Remote User's Side: The integrated wearable device in the remote user's side consists of AR smart glasses, a 360°camera, and a notebook computer (see Figure 10). The AR glasses present a semitransparent display on top of the physical world, thus, allowing the user to view the physical world simultaneously. It is packed with a motion-tracking sensor for detecting the direction that the user is facing and a wireless module to exchange information with the local user's side via the Internet. It is also provided with an audio output with an earphone. The camera is connected to a notebook computer to generate a live stream so that the live video data can be sent to the desktop PC on the local user side using real-time messaging protocol (RTMP). The streaming uses an H.264 software encoder.

B. Software

We develop the software of our proposed system using Unity game engine [39] with Oculus Integration for Unity [40], Leap motion SDK [36], and MOVERIO AR SDK [41].

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

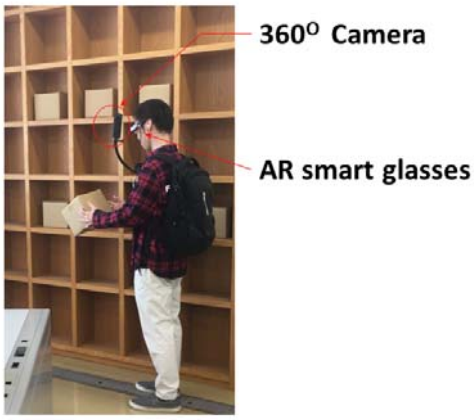

Figure 10. Remote user's setup: a AR smart glasses and a 360°mounted on the shoulder

V. Evaluation

In this section, we introduce our evaluation methodology that includes two studies to examine our Shoulder-to-shoulder prototype and test the design requirements, which were mentioned in the introduction section (Section I). In Study 1, we examine the effects of the viewing perspective of the proposed shoulder-to-shoulder viewing against two comparative conditions. This study assesses the design requirements for providing shoulder-to-shoulder viewing independence. In Study 2, we evaluate our proposed system in a more realistic collaboration. The purpose of this study is to investigate how the shoulder-to-shoulder communication affects the remote communication experience and co-presence sensation.

A. Evaluation Procedure

The two user studies are performed in the following order: Study 1→Study 2.

Before starting a study, the researcher explains all the equipment involved with the participants. The participants are then asked to try out the devices and fit the wearable equipment. At the beginning of each part, the researcher explains the purpose of the study and the participants' role in completing the tasks. The preparation time takes approximately 15 min for each participant. Further details about each part of the study are given in the following sections.

B. Study 1: Viewing Perspectives

In this study, we compare the different levels of viewing dependencies of the local user. We are interested in finding out how the difference in viewing perspective affects the remote-access user's spatial awareness level and social connection with the collaborators. Participants stayed indoors and did the test as the local VR user.

1) Workspace: In this study, we set up our experimental workspace in a room (see Figure 11). The workspace consists of a desk, a white partition, and a shelf which has multiple lattices. The local VR user and the remote AR user perform verbal communication over IP voice calls.

Figure 11. Experimental workspace for Study 1

2) Participants: For our evaluation, we recruited 12 participants from our department. They were between the ages of 20 to 26 years. All participants possessed average computing skills and had some experience with AR or VR interfaces, which could reduce the novelty effect for the test results and will provide potential insight into our system from their experiences.

3) Study Design and Tasks: This study is a within-subject design, where we compare our shoulder-to-shoulder viewing with two other conditions (as shown in Figure 12(a)) of the local user's viewing perspective of a remote environment: (a)Dependent condition and (b)Stand-in condition.

- In Dependent condition, the participants, as the local user, use an egocentric viewpoint. The viewing perspective is dependent on the control of our researcher, the remote user. They see what the remote user see of the surroundings. In this condition, a capture of the surroundings is provided using a fixed forward camera of the smart glasses worn by the remote user, which always makes the viewer's viewpoint synchronously follow that of the recorder's. The participants browse the video in an HMD without viewpoint control.
- In Stand-in condition, participants, the local user, could see a 360°video of the workspace in a consistent orientation, viewing independently of the 360°camera's rotation. Under this condition, the capture of the surroundings is provided by a 360°camera mounted on the recorder's head

The whole study had two parts. In Part 1, the participants, as the local VR users, were asked to learn the remote surroundings under the direction of an actor, as the remote AR user, and figure out the object of interest that the remote AR user was randomly assigned to find. The object could be either a letter (on one of the boxes on the desk), a box (on a shelf), or color (on a partition). The remote user was not allowed to directly tell the participants what the object was, and they had to search the workspace together and find the object of interest as fast as they could. This task simulates a situation wherein it is difficult to verbally describe the spatial arrangement and the object of interest in the scene to the collaborator, for example, a workspace full of similar items.

During the test, the participants could ask the actor any binary questions that the actor could answer using "yes/true" or

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

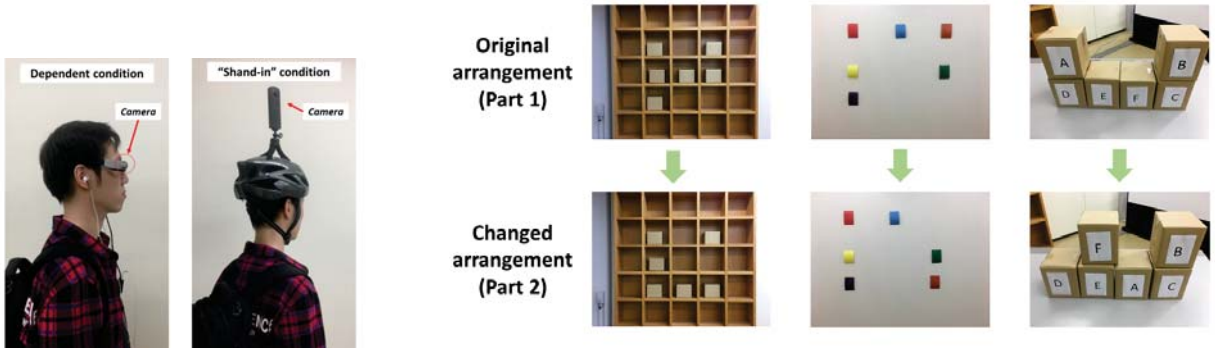

Figure 12. (a)Comparative conditions (b) Samples of changes of spatial arrangement task

Figure 13. Results of Study 1.

"no/false". A wrong answer for each binary question asked was counted as an error. The participants were given two training trials for each condition.

In Part 2, to evaluate spatial understanding, the participants were given a spatial arrangement task. After a collaboration test in Part 1, our researcher randomly made 6 changes to the arrangement of the workspace by changing a set of the experimental objects' locations. To score test points, the participants had to move the objects back to match the original arrangement shown in Part 1 of the test.

Each participant was assigned four experimental trials for each condition and allowed a learning trial before the experiments. The order of conditions was counterbalanced between participants. The study took about 30 min for each participant.

4) Data Collection: We collected both objective and subjective data. The objective variables were the number of errors that occurred during Part 1 of the study and the score from solving the arrangement task in Part 2. The subjective measure was questionnaire consisting of Networked Mind Measure of Social Presence questionnaire [42] (on Co-Presence aspect), Spatial Understanding based on Spatial Presence Questionnaire (MEC-SPQ) [43] with 6 item scale on Spatial Situation model (SSM), and user preferences. Social Presence and Spatial Understanding questionnaire were collected after each condition and user preference was collected at the end of the entire study. For all the conditions, our actor consistently performed the same way by looking toward the object of interest and utilizing hand gestures to assist verbal communication.

5) Hypotheses: We have the following hypotheses for this study:

- H1 Higher degree of viewing independence (Shoulder-to-shoulder perspective or Stand-in perspective) increases the spatial understanding and lowers the subjective mental effort and task difficulty.
- H2 Shoulder-to-shoulder perspective increases Social Presence score in terms of the Co-Presence (CoP).
- H3 Participants prefer using perspectives that could provide a higher degree of viewing independence.

6) Results: In this study, we used the Friedman and Wilcoxon signed-rank test to analyze the significance of the experiment results across the three conditions. Figure 13 illustrates the results of Study 1.

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

a) Performance: From the result, we observe that the average number of errors, in all conditions, in Part 1 were below, indicating that there was almost no error in this part.

Figure 13 illustrates the mean test score of Part 2 for the three conditions. Pairwise comparisons yielded significant difference as the *Shoulder-to-shoulder condition* and the *Stand-in condition* performed much better than the *Dependent condition* (both $p < 0.01$). However, no significant difference was found between the *Shoulder-to-shoulder condition* and the *Stand-in condition*.

b) Task difficulty: We found significant difference in pairwise comparisons between the *Shoulder-to-shoulder condition* and the *Dependent condition*, also between the *Stand-in condition* and the *Dependent condition*.

c) Spatial understanding: Pairwise comparisons yielded significant difference as the *Shoulder-to-shoulder condition* and the *Stand-in condition* got higher score than the *Dependent condition* (both $p < 0.01$). However, no significant difference was found between the *Shoulder-to-shoulder condition* and the *Stand-in condition*.

d) Co-presence: In terms of Co-Presence aspect, we found significant differences as the *Shoulder-to-shoulder condition* performed much better than the other two conditions (all $p < 0.01$). There was no significant difference between *Stand-in condition* and *Dependent condition*.

e) Preference: From the results, we found that, among three conditions of perspective, most of the participants preferred the *Shoulder-to-shoulder condition* (58%) followed by the *Stand-in condition* (33%).

f) Discussion: Our object and subject result strongly support our hypotheses H1, where the participants performed much better in spatial arrangement tasks and shown much better in the spatial understanding of the environment when they were provided with a higher degree of viewing independence. Our results also strongly support hypotheses H2 as there were significant differences in terms of Social Presence.

Most of the participants preferred having the shoulder-to-shoulder perspective and this supports our hypotheses H3, not only because it improved the users' confidence in spatial perception with view independence but also because it helped users in perceiving their partner's behaviors and hence, required less verbal communication.

C. Study 2 Collaborative Work

In this study, we evaluate the shoulder-to-shoulder communication under a more realistic collaboration scenario. The purpose of this study is to investigate how the shoulder-to-shoulder communication affects the remote communication experience and co-presence sensation.

1) Participants: For our evaluation, we recruited 12 participants from our department, which included six females. They were between the ages of 20 to 27 years with a mean age of 24 years. All of them possessed average computing skills and had some experience with AR or VR interfaces, which could reduce the novelty effect of the test results and also provide potential insight ino our system from their experiences. The participants were randomly grouped into six pairs. In each pair, one participant assumed the role of the local VR user, while the other participant assumed the role of the remote AR user.

TABLE 1. QUESTIONNAIRE

Q1. Did you observe interesting things independently?
Q2. Did you find it easy to tell your partner's viewing direction?
Q3. Did you feel gestural communication useful?
Q4. Did you feel the operation is easy enough to learn and use?
Q5. How much did you feel co-located with your partner during the test?

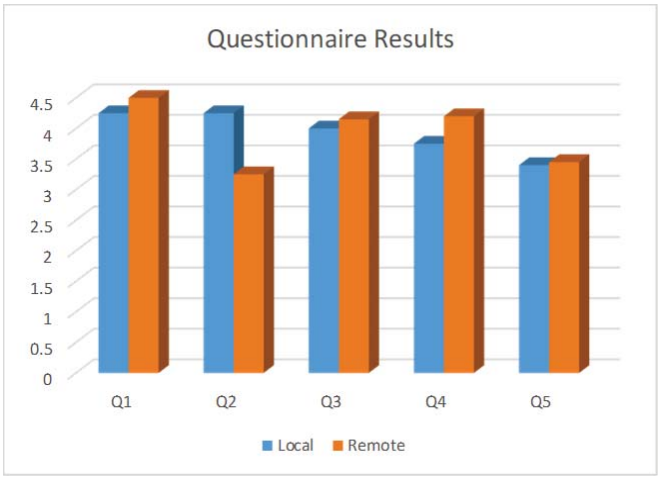

Figure 14. Questionnaire results

2) Study Design and Task: The experimental environment of the user study involved an indoor workspace for the local user and a departmental store, which was a larger space than the workspace used in Study 1, where the remote user stayed.

The task of this study was joint shopping. The goal for the participants was to work collaboratively and look for a product (such as a pencil box) that could interest both participants. In each pair, both participants were allowed a free voice communication supported by Internet IP phone call. The remote participant walked around and communicated with the local partner, and the local participant indulged in the shopping activity via remote communication. The subsystem used in the local user's part was connected to the cabled Internet, and the remote user's subsystem used a wireless connection (LTE). After the pilot test, we observed that the duration of completion was primarily influenced by personal preference. Therefore, we did not enforce any time limitation. This study was open-ended, and the only requirement was that the participants had to arrive at an agreement when selecting a product.

We collected subjective feedback from the post-task questionnaire. After each trial, the participants were asked to fill out a questionnaire that formed the basis of the subjective feedback. The participants graded each question using a 5-point Likert Scale (1 = very negative, 5 = very positive).

3) Results: Table 1 shows the questions that make up our questionnaire. We calculate the average score for each question in each group. Figure 14 shows the results. The results are divided into two groups–the local user's group and the remote user's group.

Question 1 – *Did you observe interesting things independently?* This question is used to test whether our system could

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

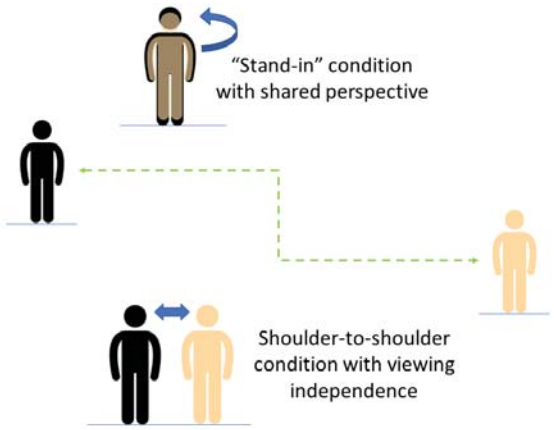

Figure 15. Comparison between the two types of remote communication

provide the users with viewing independence. According to the results, it was clear that both the users had independent control of viewpoint in remote view sharing.

Question 2 –*Did you find it easy to tell your partner's viewing direction?* This question indicates that the users could be aware of the partner's attention condition easily that made it possible for them to join in the same scene for further communication.

Question 3 – *Did you feel gestural communication useful?* This question is used to judge the practicability and effectiveness of the hand gesture communication through our system. It indicated that both the local user and the remote user found performing gestures to transmit their intentions usefully.

Question 4 – *Did you feel the operation is easy enough to learn and use?* This question is used to evaluate the ease of usability of our system. The result suggested that both the users generally found it effortless to achieve communication using our system.

Question 5 – *How much did you feel co-located with your partner during the test?* This question is aimed at investigating the overall performance and user experience. It demonstrates that during remote communication, both the users perceive a certain extent of co-located sensation.

4) Discussion:

a) Mutual Gesture: We also observed that the participants who played the role of local users graded slightly higher than their partners who played the role of the remote user. This difference, which indicates an incomplete equivalence of the gesture communication, benefits the local users more when compared to the remote users. After further communication with the participants during post-task interviews, we found that the difference was probably because the remote users could use hand gestures (such as touching, squeezing, or grasping) to interact with physical objects.

b) Shoulder-to-shoulder vs First-person Perspective: In traditional view sharing designs, which usually are found in previous computer-supported cooperative work (CSCW) [4], the local user mostly perceives the remote venue with the same field of view as that of the remote user. With such setting of first-person perspective (FPP) of the content, the remote user acts more like a "stand-in" of the local user rather than as a communicating partner (see Figure 15). It might lead to misunderstandings, thereby limiting the natural communication between the users. In contrast, our shoulder-to-shoulder communication simulates a shoulder-to-shoulder togetherness, which provides both the users with more independence and allows them to focus more on mutual interaction. This could enhance a co-located sensation, which is also supported by our user study results. In this evaluation, all the participants successfully finished the tasks. In each pair, the local participant and the remote participant could reach an agreement and pick up a target object after discussion. Each user was aware of their partners during the task, which provided the users with a close connection. We confirmed that both the users could enjoy the communication experience and generally received a certain level of co-located feeling.

VI. Potential Applications

The potential applications of our shoulder-to-shoulder collaboration prototype are not limited to the scenario used in our case studies. Our system is also suitable for other remote collaborative works or remote assistance. For example, in case of an emergency assistance scenario, an expert (the local VR user) tries to assist a worker (the remote AR user) in manual operations to handle problems for the first time; or, people with inconveniences (local VR user) can continue to stay in a comfortable environment and at the same time get a virtual sightseeing with their friends (remote AR user) to enjoy accompanying moments and rich lifelogging.

VII. Conclusion

In this paper, we introduced our design and implementation of a shoulder-to-shoulder communication prototype that aimed at enhancing a co-located sensation between two users in separate environments. This prototype supports users with viewing independence and bidirectional gesture communication. We also described our evaluation to investigate the system's usability and user performance. In Study 1, we examined the effects of viewing independence for our shoulder-to-shoulder communication system against two other conditions. In Study 2, we evaluated our system in a more realistic collaboration. The results demonstrated that both sides of the users could effectively transmit instructions relating to the physical world and could achieve a smooth remote collaboration, and finally could receive a certain degree of co-located sensation.

References

[1] Minghao Cai and Jiro Tanaka. Remote shoulder-to-shoulder communication enhancing co-located sensation. In *The Twelfth International Conference on Advances in Computer-Human Interactions (ACHI 2019)*, pages 80–85, 2019.

[2] Keisuke Tajimi, Nobuchika Sakata, Keiji Uemura, and Shogo Nishida. Remote collaboration using real-world projection interface. pages 3008–3013, 2010.

[3] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. Teleadvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 619–622. ACM, 2012.

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[4] Shunichi Kasahara and Jun Rekimoto. JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, 23(3):217–225, 2015.

[5] Rajinder S Sodhi, Brett R Jones, David Forsyth, Brian P Bailey, and Giuliano Maciocci. BeThere: 3D Mobile Collaboration with Spatial Input. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, pages 179–188, 2013.

[6] Seth Hunter, Pattie Maes, Anthony Tang, Kori Inkpen, and Sue Hessey. WaaZam ! Supporting Creative Play at a Distance in Customized Video Environments. *Conference on Human Factors in Computing Systems*, page 146, 2014.

[7] Sara A. Bly, Steve R. Harrison, and Susan Irwin. Media spaces: bringing people together in a video, audio, and computing environment. *Communications of the ACM*, 36(1):28–46, 1993.

[8] "Put-that-there": Voice and Gesture at the Graphics Interface. *Proceedings of the 7th annual conference on Computer graphics and interactive techniques - SIGGRAPH '80*, pages 262–270, 1980.

[9] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[10] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 405–415. ACM, 2015.

[11] Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19(3):273–309, 2004.

[12] Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 241–250. ACM, 2012.

[13] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 449–459. ACM, 2014.

[14] Keita Higuch, Ryo Yonetani, and Yoichi Sato. 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 - CHI '16*, pages 5180–5190, 2016.

[15] Benjamin Nuernberger, Kuo-Chin Lien, Tobias Höllerer, and Matthew Turk. Interpreting 2d gesture annotations in 3d augmented reality. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 149–158. IEEE, 2016.

[16] Susan R Fussell, Leslie D Setlock, and Robert E Kraut. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 513–520. ACM, 2003.

[17] Joel Lanir, Ran Stone, Benjamin Cohen, and Pavel Gurevich. Ownership and control of point of view in remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2243–2252. ACM, 2013.

[18] Nobuchika Sakata, Takeshi Kurata, Takekazu Kato, Masakatsu Kourogi, and Hideaki Kuzuoka. Wacl: Supporting telecommunications using wearable active camera with laser pointer. In *ISWC*, volume 2003, page 7th. Citeseer, 2003.

[19] Abhishek Ranjan, Jeremy P Birnholtz, and Ravin Balakrishnan. Dynamic shared visual spaces: experimenting with automatic camera control in a remote repair task. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1177–1186. ACM, 2007.

[20] Corey Pittman and Joseph J LaViola Jr. Exploring head tracked head mounted displays for first person robot teleoperation. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 323–328. ACM, 2014.

[21] VR QuickTime. An image-based approach to virtual environment navigation, shenchang eric chen, apple computer, inc. In *Siggraph, Computer Graphics Proceedings, Annual Conference Series*, pages 29–38, 1995.

[22] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. Collaboration in 360 videochat: Challenges and opportunities. Technical report, University of Calgary, 2017.

[23] Charles Goodwin. Gestures as a resource for the organization of mutual orientation. *Semiotica*, 62(1-2):29–50, 1986.

[24] Susan Wagner Cook and Michael K Tanenhaus. Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, 113(1):98–104, 2009.

[25] John C Tang. Findings from observational studies of collaborative work. *International Journal of Man-machine studies*, 34(2):143–160, 1991.

[26] David S Kirk and Danaë Stanton Fraser. The effects of remote gesturing on distance instruction. In *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pages 301–310. International Society of the Learning Sciences, 2005.

[27] Darren Gergle, Robert E Kraut, and Susan R Fussell. Action as language in a shared visual space. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 487–496. ACM, 2004.

[28] Anthony Tang, Carman Neustaedter, and Saul Greenberg. Videoarms: embodiments for mixed presence groupware. *People and Computers XX—Engage*, pages 85–102, 2007.

[29] Judith Amores, Xavier Benavides, and Pattie Maes. Showme: A remote collaboration system that supports immersive gestural communication. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1343–1348, 2015.

[30] Franco Tecchia, Leila Alem, and Weidong Huang. 3d helping hands: a gesture based mr system for remote collaboration. *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum*

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

and its Applications in Industry, pages 323–328, 2012.

[31] Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan, and Bruno Arnaldi. Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pages 9–12. IEEE, 2016.

[32] Minghao Cai and Jiro Tanaka. Trip together: A remote pair sightseeing system supporting gestural communication. *Proceedings of the 5th International Conference on Human Agent Interaction*, pages 317–324, 2017.

[33] Minghao Cai, Soh Masuko, and Jiro Tanaka. Gesture-based mobile communication system providing side-by-side shopping feeling. *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, pages 2:1–2:2, 2018.

[34] Gun A Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, page 14. ACM, 2017.

[35] Hani Karam and Jiro Tanaka. Finger click detection using a depth camera. *Procedia Manufacturing*, 3:5381–5388, 2015.

[36] LEAP MOTION. Leap Motion's SDK, 2019.

[37] Oculus. Oculus Rift, 2019.

[38] LEAP MOTION. LEAP MOTION, 2019.

[39] Unity 3D. Unity3D Game Engine, 2019.

[40] Oculus. Oculus Integration for Unity, 2019.

[41] EPSON. Moverio AR SDK, 2019.

[42] Frank Biocca Chad Harms. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annual International Workshop: Presence 2004*, pages 246–251, 2004.

[43] Peter Vorderer, Werner Wirth, Feliz Ribeiro Gouveia, Frank Biocca, Timo Saari, Lutz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, Tilo Hartmann, et al. Mec spatial presence questionnaire. *Retrieved Sept*, 18, 2004.

2019, © Copyright by authors, Published under agreement with IARIA - www.iaria.org