Remote Shoulder-to-shoulder Communication Enhancing Co-located Sensation

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 our mobile remote communication prototype between two users in separated environments – a remote user goes to a shared environment with mobile augmented reality setup and a local user stays indoor immersing in a virtual reality view to this shared environment. It realizes a kind of remote shoulder-to-shoulder communication, which simulates that the users go shoulder-to-shoulder with viewing independence and bidirectional gesture communication, and the major target is to enhance a shared co-located sensation. We also introduce our preliminary evaluation used to test the system usability and user performance.**

Keywords–Remote communication; Co-located Sensation; Viewing independence; Gesture communication.

I. INTRODUCTION

In recent years, remote communication is extensively used at work or in daily life to increase productivity and to improve the performance of the instant communication. It allows users from the different locations to communicate and collaborate together as a team. It is a cost-effective way that can truly help users to get an instant solution for any type of problem [1].

Although commercial remote conferencing technologies are cost-effective and more immersive than traditional phone calls with only voice, most of these systems mainly provide a mere capture of both user's face and limited transition in terms of body language or the reference of ambient, which also act as a great source of information. When facing 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 experiences. When users want to describe the objects or directions in the scene or showing operations, using the hand gesture would be more understandable than mere voice.

Another problem is derived from 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 camera or they might place the device in a fixed position in order to have a wild range of capture. In most cases, people have to take the camera and move around in order for the remote person to perceive the entire scene. Such constraints make it difficult for users to get a common perception or feel like staying together.

In this paper, we propose a solution with a prototype providing a mobile and immersive remote *Shoulder-to-shoulder Communication* between a local user and a remote user who are in different places. This type of communication can enhance a co-located sensation during the remote communication.

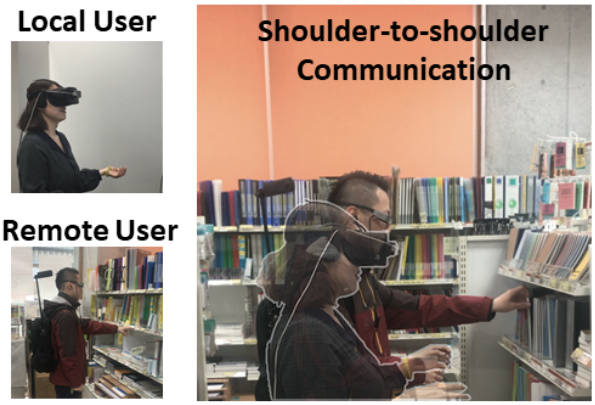

Figure 1. Remote Shoulder-to-shoulder Communication

The prototype is designed to be used by two users in different places (Figure 1). For convenience, we refer to the user who goes to a remote environment that would be shared as the remote user, and the other one who is staying in a local indoor workspace and remotely viewing the shared world as the local user, even though the roles may well be reversed. We try to offer both users a shared feeling that they are going shoulder-to-shoulder together with 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 glasses getting augmented reality (AR) experience.

The main contributions of this work are: (1) the implementation of the hardware prototype including the mobile setup for the remote user and wearable setup for the local user, and (2) the software system supporting virtual and augmented reality spatial interaction between two users, and (3) a preliminary evaluation carried out to test the usability of our prototype.

In Section II, we introduce the related works. In Section III, we introduce our system design. In Section IV, we introduce our implementation. In Section V, we introduce the preliminary evaluation. In Section VI, we discuss the difference between

 ISBN: 978-1-61208-686-6

our should-to-shoulder communication design and traditional remote communication design. In Section VII, we draw our conclusion to this work.

II. Related Work

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

Several pieces of research have spent effort on remote video communication techniques which aim to realize a remote collaborative work among users in separated places [5][1]. 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.

Hand gesture has been shown as an irreplaceable part for conversation, as it is treated as a cognitive visible awareness cue and provides rich context information that other body cues cannot reveal, which contributes significantly to a recipient's understanding [6][7]. Over the past several years, some researchers have paid attention to support gestural interaction in the shared media space with 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 [8]. Kirk et al. [9] demonstrated the positive effect of gestures and visual information in promoting the speed and accuracy in remote collaborative activities. Another work by Fussell et al. [10] demonstrated that users tend to rely more on visual actions than on speech in the collaborative work.

Previously, we built a remote sightseeing prototype supporting gestural communication to realize a gesture communication between two separated users [11][12]. It investigated providing users with an approach to achieve a spatial navigation and direction guidance during mobile sightseeing. The positive evaluation results of this work encourage us to support a mid-air gesture interaction for improvements of users' interactions in remote collaborations.

III. System Design

The system design consists of the following main aspects:

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

A. Shoulder-to-shoulder Viewing Independence

To capture and share the real-time remote environment, we choose a new generation camera that provides a high-resolution video with a range of 360°in both horizontal and vertical. Different from previous view sharing systems that usually put the camera on the remote user's head or cheek [13], this camera is fixed to one of the remote user's shoulder with the help of a steel support. The real-time 360°video is streamed back to the local side via the Internet and displays in the head-mounted display wore by the local user.

Figure 2. Independent control of the viewing direction for the local user

Since the camera fixed to the shoulder, its orientation is preventing from being influenced by the remote user's head motions. The local user is supported with independent control of viewing direction which can be simply manipulated by head movements. As shown in Figure 2, the local user simply turns the head and naturally changes the viewpoints. Through this design, the local user immerses in the virtual remote world, perceiving a sensation that personally standing next to the remote user and seeing the scene.

B. Shoulder-to-shoulder Gesture Communication

In our system, we provide users an approach to achieve a bidirectional gesture interaction during the 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 reappearing the local user's gestures in the remote user's field of view.

1) Remote Gestures to Local User: As we have introduced in Section III-A, the local user has a 360°independent viewing of the remote world with a perspective by the remote user's

 ISBN: 978-1-61208-686-6

Figure 3. The local user's field of view: the remote user is making gestures

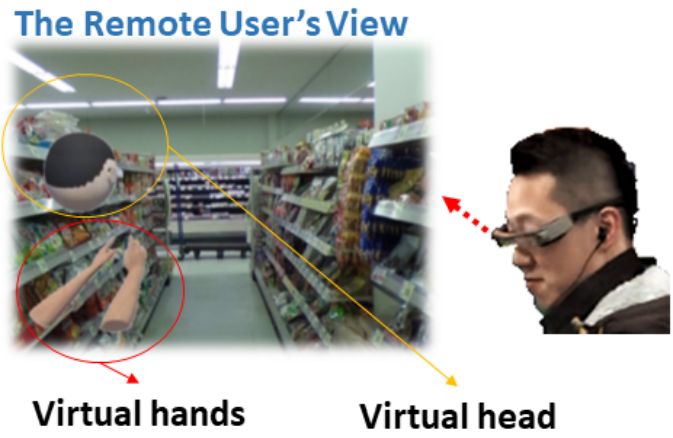

Figure 4. The remote user's field of view: the local user is making gestures. Red circle shows the virtual hands and yellow circle shows the virtual head representing the local user

shoulder. This design allows the local user to see the remote hand gestures, as well as the profile face. As shown in Figure 3, the local user simply looks leftward, and directly see the remote partner performing hand gestures with an object (opening a notebook).

2) Local Gestures to Remote User: One of the important contributions of this system is reappearing the local user's hand gestures in the remote world, as the local user is in a physically separated 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 accuracy and convenient way, depth-based recognition has been used to in current researches for hand motion extraction [14][15]. 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 [16]. We construct a pair of 3D hand models including palms and different finger joints. This pair of 3D hand models is matched with the latest hand structure data. Then, 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 exact same ones, almost simultaneously appearing in the remote user's field of view (Figure 4).

C. 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 easily tell where the partner is looking at. It would improve the efficiency of communication when the user tries to join in the same field of view so as to find out some common interesting points or make some discussion. As we introduced in Section III-A, the local user can easily tell the remote user's viewing direction in the virtual scene. Because 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.

Figure 5. The remote user's view: Pointing cue for instructions

A motion tracking sensor is used to extract the head motion which is used to rotate the virtual head model. The model presents on the left side of the vision, showing the remote user's precise facing direction (see Figure 4).

D. Pointing Assistance

Previous research has shown that utilizing a finger pointing assistance can benefit the cooperation and instruction 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 a pointing assistance with fingers. The user performs a freehand pointing gesture to use a virtual 3D arrow showing specific direction information in the remote user's view. This 3D arrow is treated as a spatial cue assisting a navigation or selection task during the communication (see Figure 5).

Our system uses a heuristic approach for the gesture recognition. Using the depth sensor, our system can keep

 ISBN: 978-1-61208-686-6

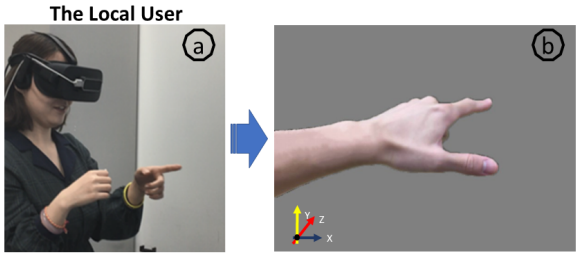

Figure 6. (a): The local user makes a pointing gesture (b): a zoomed in view of the pointing gesture

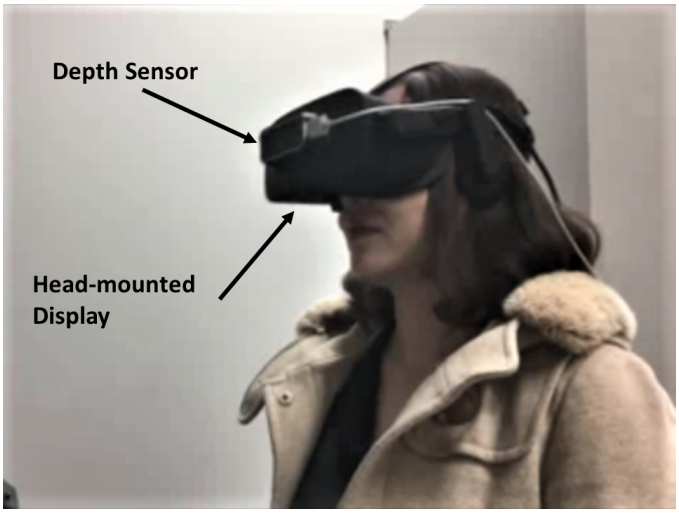

Figure 7. The remote user's wearable device: a head-mounted display with a depth sensor attached to its front side

tracking the 3D structure of the user's hands including different finger joints and extract both the 3D position and orientation of the local users fingers. Our system requires no calibration or precedent training. To activate the pointing technique, the user extends only the thumb and index finger and keeps the angle between them larger than the set threshold (see Figure 6).

IV. Implementation

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

A. Local User's Side

The equipment in local user's side include the wearable devices and a desktop PC (see Figure 7). The desktop PC (Intel Core i5, RX480 Graphics Card, 8GB RAM) placed on the local user side is used to analyze data and engine the core system. We use Unity engine to render and process the incoming data from both remote and local side, as well as to generate GUI for both users. The headset we chose as the local user's head-mounted display uses a pair of low persistence OLED screens, providing a 110 field of view [17]. A point tracking sensor is used to provide a full 6 degree of freedom rotational and positional tracking of the head movements. For hand motion tracking, the depth sensor we used is light enough and introduces a gesture tracking system with sub-millimeter accuracy [18]).

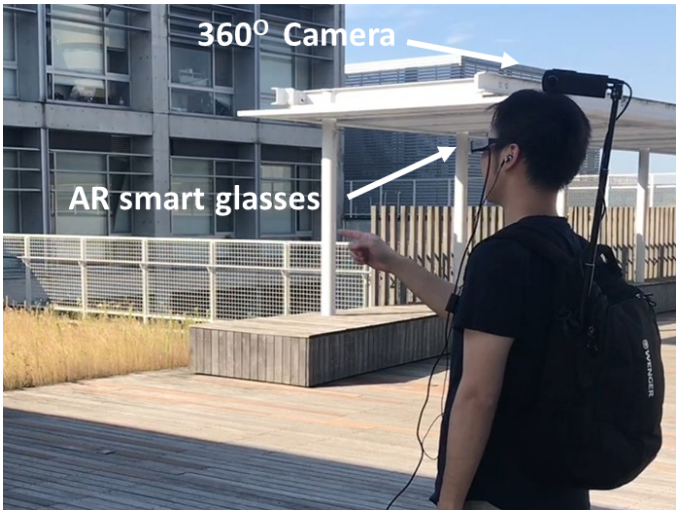

Figure 8. The remote user's field of view: the local user is making gestures. Red circle shows the virtual hands and yellow circle shows the virtual head representing the local user

B. Remote User's Side

The integrated wearable device in remote user's side consists of an AR smart glasses, a 360°camera, and a notebook computer (see Figure 8). The AR glasses presents a semitransparent display on top of the physical world while allows the user to view the physical world clearly. It packs with a motion-tracking sensor to detect the user's facing direction and a wireless module to exchange information with the local user's side via the Internet. It also provides an audio output with an earphone. The camera is connected to a notebook computer over USB (1280x720 15fps) to generate a live stream to send the live video data to the desktop PC on the local user side with Real Time Messaging Protocol (RTMP). The streaming uses H.264 software encoder.

V. Preliminary Evaluation

We carried out a user study for preliminary evaluation. The purpose is to investigate how the shoulder-to-shoulder viewing affects the remote communication experience, especially with hand gesture communication.

A. Participants

In this study, we recruited eight participants in our departments (between 21 an 27 years old). All participants had regular level computer skills. They were divided into four pairs. Each pair had two roles: a local user and a remote user.

B. Task and Procedure

In each pair, one participant played the role of the local user, while the other one played the role of the remote user. Before the experiment, our researchers explained how to use the system and the participants were allowed to practice for 10 minutes. The whole experiment took about 40 minutes for each group.

The environment of user study involved an indoor workspace for the local user and a department store where the remote user stayed.

 ISBN: 978-1-61208-686-6

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 together during the test?

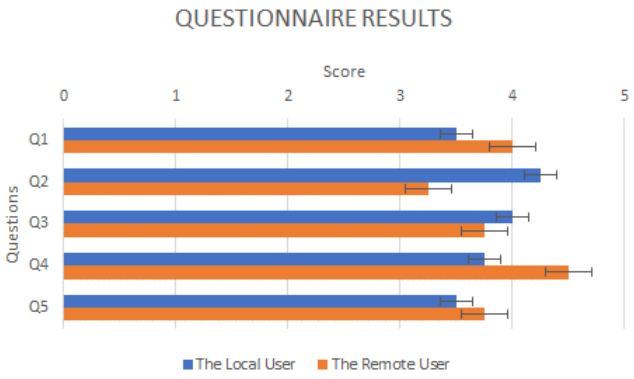

Figure 9. Questionnaire results

The study task was joint shopping in a department store to find out a product that could interest both participants (such as a pencil box). 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 participated in the shopping via remote communication. The subsystem in local user's part was connected to the cabled Internet, and the remote user's subsystem used a wireless connection (LTE).

After each experiment, all four pairs of participants were asked to fill out a questionnaire including to get the user feedback. The participants graded each question with 5-point Likert Scale (1 = very negative, 5 = very positive).

C. Results

Table 1 shows the questions of our questionnaires. We calculated the average score of each question in each group. Figure 9 shows the results. The results were divided into two groups–the local user's group and the remote user's group.

Question 1 – *Did you observe interesting thing independently?* was used to test whether our system could provide the users with viewing independence. According to the results, it was clear that both users could have independent control of viewpoint in the remote view sharing.

Question 2 –*Did you find it easy to tell your partner's viewing direction?* indicated that the users could be aware of the partner's attention condition easily which provides the possibility to join in the same scenery for further communication.

Question 3 – *Did you felt gestural communication useful?* was 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 was useful.

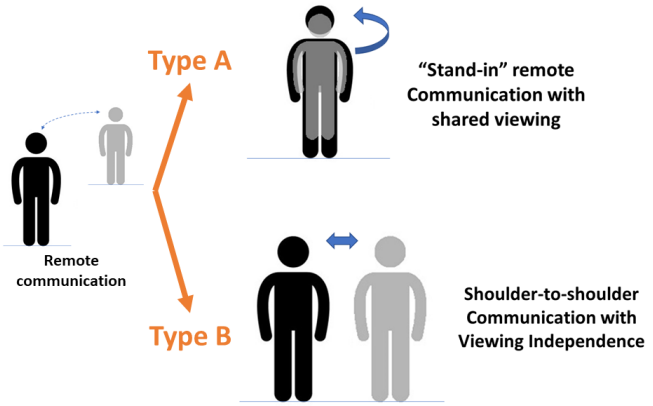

Figure 10. Comparison between two types of remote communication

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

In Question 5 – *How much did you feel co-located with your partner together during the test?* we aimed to investigate the overall performance and user experience. It demonstrates that, during the remote communication, both users perceived a certain extent of co-located sensation.

In the results of Q3, both users gave positive scores. So, we confirmed that users could perform gestures to transmit their intentions and achieve a mutual smooth communication. During the communication, users used mutual gesture interaction as a nonverbal body cue.

From the results of Question 3, 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 means an incomplete equivalence of the gesture communication that benefits the local users more. After further communication with the participants in some post-task interviews, we found it was probably because the remote users could use hand gestures (such as touching, squeezing or grasping) to actually interact with physical objects.

In this evaluation, all 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 were aware of their partners during the task, which provides users with a close connection. We confirmed that both users could enjoy the communication experience and generally receive a certain level of co-located feeling.

VI. DISCUSSION

In this section, we discuss the difference between our should-to-shoulder communication design and traditional remote communication design. We also describe some potential applications.

 ISBN: 978-1-61208-686-6

A. 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 of the remote user. With such sharing of first-person perspective (FPP) of the content, the remote user acts more like a "stand-in" of the local user rather than a communicating partner (see Figure 10-Type A). It might lead to misunderstanding and limits the natural communication between users. By contrast, our shoulder-to-shoulder communication simulates a shoulder-to-shoulder togetherness, which provides both users with more independence and let them could focus more mutual interaction (see Figure 10-Type B). This could enhance a co-located sensation, which is also supported by our user study results.

B. Possible Applications

Our shoulder-to-shoulder communication design can be used in a variety of applications where remote collaboration is useful. For example, in the use of remote maintenance or remote instructions of industrial operations, the local users would be an expert to guide a worker who would be the remote user in a shared workspace. Or, the local users would be people with physical inconveniences who have to stay in the hospital or other comfort environments try to have virtual sightseeing with a remote user who might be friends or relatives.

VII. Conclusion

In this paper, we introduced our design and implementation of a shoulder-to-shoulder communication prototype which aimed to enhance a co-located sensation between two users in separated environments. This prototype supported users with viewing independence and bidirectional gesture communication. We also described our user study to investigate the system usability and user performance. The results demonstrated both 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. In the future work, we plan to the apply our prototype to different scenarios and perform further evaluations.

References

[1] S. Hunter, P. Maes, A. Tang, K. Inkpen, and S. Hessey, "WaaZam ! Supporting Creative Play at a Distance in Customized Video Environments," Conference on Human Factors in Computing Systems, 2014, p. 146.

[2] K. Tajimi, N. Sakata, K. Uemura, and S. Nishida, "Remote collaboration using real-world projection interface." 2010, pp. 3008–3013.

[3] P. Gurevich, J. Lanir, B. Cohen, and R. Stone, "Teleadvisor: a versatile augmented reality tool for remote assistance," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012, pp. 619–622.

[4] S. Kasahara and J. 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, vol. 23, no. 3, 2015, pp. 217–225.

[5] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "BeThere: 3D Mobile Collaboration with Spatial Input," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13, 2013, pp. 179–188.

[6] C. Goodwin, "Gestures as a resource for the organization of mutual orientation," Semiotica, vol. 62, no. 1-2, 1986, pp. 29–50.

[7] S. W. Cook and M. K. Tanenhaus, "Embodied communication: Speakers gestures affect listeners actions," Cognition, vol. 113, no. 1, 2009, pp. 98–104.

[8] J. C. Tang, "Findings from observational studies of collaborative work," International Journal of Man-machine studies, vol. 34, no. 2, 1991, pp. 143–160.

[9] D. S. Kirk and D. S. 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! International Society of the Learning Sciences, 2005, pp. 301–310.

[10] D. Gergle, R. E. Kraut, and S. R. Fussell, "Action as language in a shared visual space," in Proceedings of the 2004 ACM conference on Computer supported cooperative work. ACM, 2004, pp. 487–496.

[11] M. Cai and J. Tanaka, "Trip together: A remote pair sightseeing system supporting gestural communication," Proceedings of the 5th International Conference on Human Agent Interaction, 2017, pp. 317–324.

[12] M. Cai, S. Masuko, and J. Tanaka, "Gesture-based mobile communication system providing side-by-side shopping feeling," Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion, 2018, pp. 2:1–2:2.

[13] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "Mixed reality collaboration through sharing a live panorama," in SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications. ACM, 2017, p. 14.

[14] H. Karam and J. Tanaka, "Finger click detection using a depth camera," Procedia Manufacturing, vol. 3, 2015, pp. 5381–5388.

[15] J. Amores, X. Benavides, and P. 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, 2015, pp. 1343–1348.

[16] Leap Motion, "Leap Motion's SDK," Retrieved: January 2019. [Online]. Available: https://developer.leapmotion.com/documentation/

[17] Oculus, "Oculus Rift," Retrieved: January 2019. [Online]. Available: https://www.oculus.com/

[18] Leap Motion, "LEAP MOTION," Retrieved: January 2019. [Online]. Available: https://www.leapmotion.com/

 ISBN: 978-1-61208-686-6