Computers, Materials & Continua
DOI: 10.32604/cmc.2023.034209
Article

View Types and Visual Communication Cues for Remote Collaboration

Seungwon Kim[1], Weidong Huang[2], Chi-Min Oh[3], Gun Lee[4], Mark Billinghurst[4] and Sang-Joon Lee[5,*]

[1]AI Convergence College, Chonnam National University, Gwangju, 61186, Korea
[2]University of Technology Sydney, Sydney, 2007, Australia
[3]SafeMotion, Gwangju, 61011, Korea
[4]School of ITMS, University of South Australia, Adelaide, 5001, Australia
[5]Interdisciplinary Program of Digital Future Convergence Service, Chonnam National University, Gwangju, 61186, Korea
*Corresponding Author: Sang-Joon Lee. Email: s-lee@jnu.ac.kr
Received: 09 July 2022; Accepted: 07 September 2022

Abstract: Over the last several years, remote collaboration has been getting more attention in the research community because of the COVID-19 pandemic. In previous studies, researchers have investigated the effect of adding visual communication cues or shared views in collaboration, but there has not been any previous study exploring the influence between them. In this paper, we investigate the influence of view types on the use of visual communication cues. We compared the use of the three visual cues (hand gesture, a pointer with hand gesture, and sketches with hand gesture) across two view types (dependent and independent views), respectively. We conducted a user study, and the results showed that hand gesture and sketches with the hand gesture cues were well matched with the dependent view condition, and using a pointer with the hand gesture cue was suited to the independent view condition. With the dependent view, the hand gesture and sketch cues required less mental effort for collaborative communication, had better usability, provided better message understanding, and increased feeling of co-presence compared to the independent view. Since the dependent view supported the same viewpoint between the remote expert and a local worker, the local worker could easily understand the remote expert's hand gestures. In contrast, in the independent view case, when they had different viewpoints, it was not easy for the local worker to understand the remote expert's hand gestures. The sketch cue had a benefit of showing the final position and orientation of the manipulating objects with the dependent view, but this benefit was less obvious in the independent view case (which provided a further view compared to the dependent view) because precise drawing in the sketches was difficult from a distance. On the contrary, a pointer with the hand gesture cue required less mental effort to collaborate, had better usability, provided better message understanding, and an increased feeling of co-presence in the independent view condition than in the dependent view condition. The pointer cue could be used instead of a hand gesture in the independent view condition because the pointer could still show precise pointing information regardless of the view type.

Keywords: Mixed reality; remote collaboration

1 Introduction

People sometimes use a video call to show a task/play space to a remote collaborator [1–3]. In this case, remote collaboration includes at least two people or groups who are apart, and they play the roles of local and remote collaborators [4,5]. The local collaborator is the person staying with the task objects and directly manipulating them, and the remote collaborator is the person joining the collaboration through the video call [6,7].

Remote collaboration has been studied for more than three decades [8] with a major focus on increasing awareness: understanding what is going on in the task space and understanding collaborators' activities [5]. This is because awareness may be reduced in remote collaboration compared to co-located collaboration [9–11]. To increase the awareness, previous studies mostly focused on two main topics: (1) how to let the remote collaborator (or remote expert) know the current state of the task space while the local collaborator (or local worker) directly looks at the task space [8], and (2) implementing a remote collaboration system for better communication between collaborators [12]. For the first topic, researchers have mainly studied two types of sharing a task space view: dependent and independent views. In the dependent view condition, both the local worker and the remote expert have the same view controlled by the local worker. In the independent view condition, they can individually control the viewpoint, so each has a different view. For the second topic, researchers have focused on adding visual communication cues such as virtual hand gesture, sketches, and pointer cues for better communication [4]. The remote expert shows his/her collaborative activities to the local worker through the visual communication cues while the local worker shows his/her collaborative activities through the live video. For example, the shared live video shows the local worker's hand movements and object manipulations.

Therefore, sharing the task space view and adding visual communication cues are essential components in remote collaboration systems [8], but they were studied individually. In this paper, we investigate how these two characteristics influence on each other. We especially explore the effect of the view types on the use of three visual communication cues (hand gesture, sketch, and pointer) in an Augmented Reality (AR) system for remote collaboration system.

In the following sections, we first review related work, then describe our study methodology including how we developed a prototype and designed a user study. We next present the study results and discussion and then wrap up with a conclusion and directions for future research.

2 Related Work

In this section, we describe previous studies exploring the use of a shared view and visual communication cues in remote collaboration.

2.1 Two View Types

In early studies, researchers shared a screen view between collaborators (e.g., sharing a live video of the screen) rather than a view of the physical space (e.g., sharing a live video of the real world task space) [13,14]. When sharing the screen view, Stefik et al. [15] introduced the What You See Is What I See (WYSIWIS) interface where collaborators have the same screen view, so collaborators do not need to worry about what the other person is looking at, and the interface increases awareness.

From the middle of the 1990's, researchers started to apply the WYSIWIS interface for sharing physical space views. The WYSIWIS interface was a dependent view because the local worker controlled the viewpoint of the shared view, and the remote expert had the same view. Establishing

the dependent view was simple because the system simply shared a live video of the local worker's view with the remote expert. One of the early works demonstrating this was the NaviCam AR system developed in 1995 by Rekimoto et al. [16]. This was a handheld system with a small video camera to detect real-world visual markers and sharing the dependent view. Johnson et al. [6] adopted their idea but used a smart tablet with an embedded camera. Alem et al. [17] also exploited the dependent view with a head mounted display (HMD) based system.

However, this dependent view has an issue in that the remote expert cannot control his/her viewpoint and cannot see what he/she wants to. To solve this issue, researchers developed an independent view that both the remote expert and the local worker can individually control without influencing the other's [18]. To establish the independent view, some researchers developed hardware systems enabling the remote expert to control a camera in the local task space. For example, Kuzuoka et al. [19] developed 'GestureCam' where a remote user could control the orientation of a camera in the local task space. Later, Sakata et al. [20] put the camera on the local worker's shoulder and increased the portability.

Instead of using a hardware system, some researchers used computer vision algorithms to achieve an independent view. For example, Kasahara et al. [18] and Gauglitz et al. [21] stitched images together to create a large view which the remote expert could navigate around. However, this stitching technology had a limitation: the non-updated part of the large view does not show the current state of the task space. Teo et al. [3], Teo et al. [22], and Nagai et al. [23] solved this limitation by sharing 360 degree views and allowing both collaborators to independently control their views. Teo et al. [3] and Teo et al. [22] shared the 360 degree video from a 360 degree camera that was attached to the HMD, above the local worker head.

Recently, Bai et al. [24] started to reconstruct the 3D environment in real time to provide an independent view. Gao et al. [25] used 3D point-cloud data to share a virtual model of the 3D task space and compared independent and dependent views while supporting 3D hand gestures. Huang et al. [1] introduced a system that reconstructs the task space in real time and renders the remote user's hands in the reconstructed scene. The summary of the previous studies is listed below (see Table 1).

Table 1: Summary of references by the view types

View type	Reference#	Description
Dependent view	[13,14]	Share a live video of the screen
	[15]	Introduce the What You See Is What I See (WYSIWIS) interface for sharing a physical space view
	[6,16]	Establish a dependent view when using a hand-held device
	[17]	Use a head mounted display (HMD) when sharing a physical view
Independent view	[19,20]	Establish an independent view by allowing the remote expert to control a camera in a local space
	[18,21]	Establish an independent view by stitching live video feeds and developed a large view where a remote expert can navigate around
	[3,22,23]	Support a 360-degree independent view with a 360-degree camera
	[24,25]	Establish an independent view by creating a 3D virtual reconstruction of the local space

2.2 Visual Communication Cues

In remote collaboration, verbal communication alone is not enough for presenting spatial information, so researchers have used visual communication cues as well [26,27]. Visual communication cues were used for transferring spatial information from a remote expert to a local worker [12]. The collaborative activities of the local worker are transferred through a live video [8], and the visual communication cues were used as a counterpart to show remote expert's activities to the local worker.

The most studied visual communication cues were hand gestures, sketches, and pointers, and they were overlapped on the shared view [27,28]. The pointer cue is simply showing a pointer to represent point information and there are several types of pointers [29]. For example, one type is a physical laser pointer that was controlled by a remote expert with an actuator. The actuator was synchronized with the one in the local task space and a laser pointer was attached on it [19,20]. Another type of pointer cue is a virtual pointer controlled by a mouse. Fussell et al. [30] used it and found that the mouse pointer cue was not that effective compared to the sketch cues. From the middle of 2010's, researchers started to explore the use of gaze pointers [31–33] and found that it increased co-presence between collaborators.

The sketch cue was mostly studied as a counterpart of the pointer cue in comparison studies between them [30]. The prominent characteristic of the sketch cue is permanence as the sketches remain until being removed [27]. This is both a benefit and a disadvantage. Permanency allows precise communication by drawing the shape of the object at the target position and orientation, so the local worker simply places the object as it was drawn. However, if it is drawn incorrectly or remains after sketches were used, the sketches can cause confusion between collaborators [30]. The first case of unnecessary and incorrect sketches was that sketches lost the real-world reference without the function anchoring them on the real-world space. As a solution, researchers stabilized the sketch with the anchoring the sketches by using computer vision techniques [21,27]. Similarly, when using a sketch cue with a dependent view: the viewpoint can be suddenly changed by a local worker while the remote expert is still drawing sketches and causing incorrect drawing [27]. To solve this issue, researchers implemented an independent view as described in Section 2.1. Kim et al. [34] and Fakourfar et al. [35] explored a freeze function that simply paused the live video of the dependent view and established an instant independent view while drawing sketches.

The hand gesture cue was the most natural communication cue as it is also generally used in the real world [12]. In early studies, systems supporting a hand gesture cue in a 2D view, often had a camera providing a top-down view [7,17,36,37]. With a top-down camera, Alem et al. [17,38] took a live video of the remote user's hands on the screen showing the task space and displayed it on the local user's near eye display. Kirk et al. [36] took the remote user's hands and displayed them on the task space by using a projector. The remote user's hands can also be extracted from a live feed of a top-down camera by using the OpenCV library and rendered on the local user's display [7,37]. The summary of the sutdies is at Table 2.

Table 2: Summary of reference by the visual communication cue

Visual cue	Reference#	Description
Pointer	[19,20]	Introduce a physical laser pointer
	[30]	Introduce a mouse pointer
	[31,32,33]	Introduce a gaze pointer

(Continued)

Table 2: Continued

Visual cue	Reference#	Description
Sketch	[30]	Introduce a sketch cue for remote collaboration, but the sketches were drawn on a screen space rather than a task space
	[21,27]	Introduce a world-stabilized sketch cue for drawing sketches in a task space
	[34,35]	Introduce a freeze view interface for correct drawing sketches
Hand	[17,36]	Simply show hands with a live video
	[7,39]	Extract hands from a live video feed then render it on the 2D shared view
	[4,22]	Support 3D hand gestures in a 3D reconstructed local space

2.3 Combination of Visual Cues

While most previous studies investigated the use of individual visual cues (by comparing between them or with and without a visual cue), some researchers [4] investigated the use of combinations of the visual cues. For example, Chen et al. [7] explored an interface supporting sketch and hand gesture cues and compared it to a voice only condition. Their system included a tablet for the local worker and a laptop for a remote expert when sharing a 2D dependent view. Teo et al. [22] also explored the combination of hand gesture and sketch cues and compared it to the hand gesture condition.

The most relevant work is Kim's study [12] comparing four combinations of the three visual cues in the 3D dependent and independent views. Their four combinations were hand gesture (Hand Only condition in our study), pointer with hand gesture (Hand + Pointer condition in our study), sketch with hand gesture (Hand + Sketch condition in our study), and pointer and sketch with hand gesture. This paper expands Kim's study [12] by adopting the three visual-cue combinations of theirs and using the same display devices: Meta2 [39] and FOVE [40]. The main difference between Kim's [12] and ours is that they mainly focused on comparing the use of the four visual-cue combinations, but we mainly focused on the different use of each combination between the two view types: dependent and independent views.

2.4 Hypothesis

There are many studies exploring the use of visual communication cues and view types individually, but there has been no previous study directly investigating the influence of the view type on the use of visual communication cue. In this paper, we investigate the influence of view type on the use of each combination among Hand Only, Hand + Pointer, Hand + Sketch cues, and make following hypotheses.

(1) The hand gesture cue is a powerful visual cue regardless of the view type.
(2) The additional pointer cue is not very crucial when using it with a hand gesture cue.
(3) The view type does not influence the use of additional sketch cues.

3 Methodology

To investigate the influence of the view types on the use of the visual cue combinations, we developed a prototype system and designed a user study.

3.1 Prototype System

The prototype consists of two units: remote expert and local worker units (see Fig. 1).

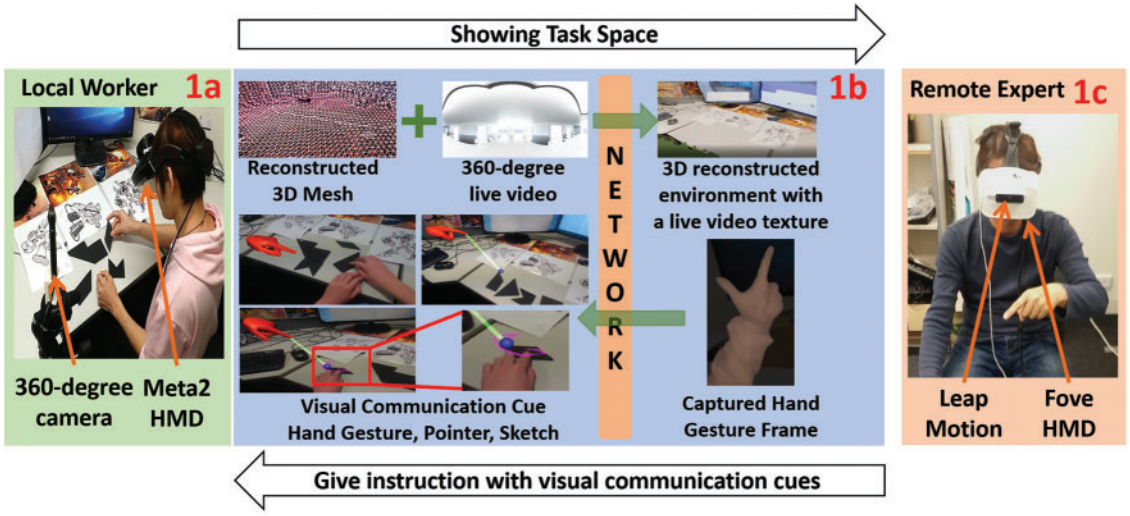

Figure 1: The local worker and remote expert try to solve a puzzle together through a remote collaboration system. The local worker system shares the task space with a mesh reconstruction and a 360-degree live video, and the remote expert gives instruction to a local worker with leap motion capturing hand gesture and providing visual communication cues

Each unit is powered by a personal computer (PC) and is connected with each other through a network connection by an Ethernet cable. All required data-transfer for the collaborative work was proceeded with the network between the two units, and the FOVE and Meta2 HMDs were wired with the PCs for Virtual Reality (VR) and Augmented Reality (AR) views respectively. The local worker uses a Meta2 AR display [39] (see Fig. 1a). Since the Meta2 supports 3D mesh reconstruction of the real-world but not color texture, we used a 360-degree live video camera, the Ricoh Theta V [41], for a live texture image and applying onto the reconstructed mesh. The remote expert wears a FOVE VR display [40] (see Fig. 1c) and a Leap Motion [42] hand tracker is attached on the front of it for the use of visual communication cues. We note that all software development was conducted with Unity 2017.3.0f3 version for compatibility.

3.1.1 View Styles

Our remote collaboration prototype supports 3D views with a 3D reconstructed mesh. The local system first uses the Meta2 Unity SDK to reconstruct a 3D mesh of the task space, which is automatically sent to the remote system. Since the remote system also employs the Meta2 Unity SDK, the position and size of the 3D mesh is equivalent in both local and remote systems. To support live updates of the scene, we map the 360-degree live video texture images onto the shared 3D mesh (see Fig. 1b). For correctly mapping the 360-degree live video texture, the system should know the projection origin (the position and orientation of the 360 camera), so we use a Vive Pro tracker attached to the 360 camera and then used the tracked Vive Pro tracker position to get the projection origin position. Since the Vive tracker coordinate frame is different from Meta2 coordinate frame, we cannot directly use the position data. We attached another tracker to the Meta2 and calculated the relative

transformation between the two trackers and manually calibrated and mapped the two coordinate systems.

To establish a dependent view, we simply shared position and orientation data of the local worker between the local and remote systems. For the independent view, we imitated side-by-side collaboration and placed the 360-degree camera next to the local worker with a tripod similar to Fussell et al. studies [30] (see Fig. 1a). The viewpoint of a remote expert was initially positioned at the 360-degree camera which was 1.5 m away from the center of the task space. Thus, with the independent view, the remote expert and local worker can individually and respectively control the position and orientation of their view.

3.1.2 Visual Communication Cues

Our system supports hand gesture, pointer, and sketch cues, and the design of them follow prior works [4,12]. The details are described below:

Hand Gesture: The representation of the remote expert's hand gesture in the shared space. A hand tracking sensor, the Leap Motion [42], is attached on the front of the remote expert's HMD (see Fig. 1c) to capture the remote expert's hands and share them with the local worker.

Pointer: A virtual blue pointer displayed on the surface of the shared mesh. A pointing gesture (opening the index finger while closing all the others, see left of the Fig. 2) is used to control the position of the pointer.

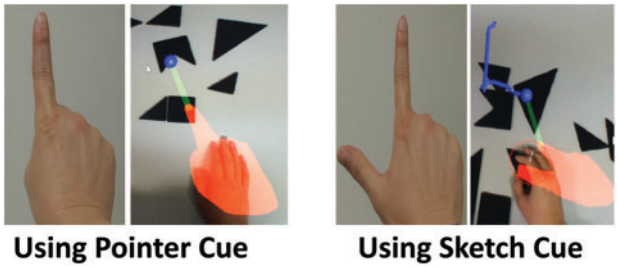

Figure 2: Finger poses for using a pointer (Left) and sketches (Right)

Sketch: Virtual line segments displayed on the surface of the shared mesh. A new sketch is drawn when the remote expert opens their thumb and index fingers while closing the others (see right of the Fig. 2). The blue pointer indicates where the sketch is drawn. By following Fussell's suggestion [30], the sketch is automatically erased after one second from the time it was drawn.

The system exploits a Leap Motion sensor for real-time tracking of hands [42] and shares the tracking frame between the local and remote systems. Thus, the virtual hands representing the remote user's hand gesture are displayed in both the remote expert's (FOVE) and the local worker's (Meta2) displays. The pointer and sketch cues are also implemented in both local and remote systems with the shared hand tracking information from the Leap Motion. Like previous studies [3,4,26], we use specific hand poses for using pointer and sketch cues, hence supporting easy switching between the three visual cues and without requiring any additional devices.

The pointer and sketches are displayed on the surface of the shared mesh while the hand gesture is displayed in the air. The hand gesture in the air is represented with pure hand tracking data, but the pointer and sketch cues need computational power to find the position of them on the shared mesh. We used collision detection between the mesh and a ray from the tip of index finger, and the collision point

becomes the position of the pointer or the drawing point of the sketch (see Fig. 2). The accumulated drawing points form line drawings. The ray direction is decided by the index finger pointing direction, and a green line is drawn between the index finger and the final position of a pointer or sketch.

Since the local and remote systems share the same implementation with the shared hand tracking data, the visualization results are the same on each end. To visualize the hands, a pointer, and sketches, we customized an example Unity scene called 'Pinch Draw' in the Leap Motion SDK [43]. Additionally, our system supports using both hands for hand gesture, pointer, and sketch interfaces.

3.2 User Study Design

3.2.1 Conditions

We conducted a user study to investigate the influence of the view types on the use of visual communication cues. The conditions were two view types: dependent and independent views and three combinations of visual cues (as described below–see Fig. 3). We compared the usage of each combination between the two view types.

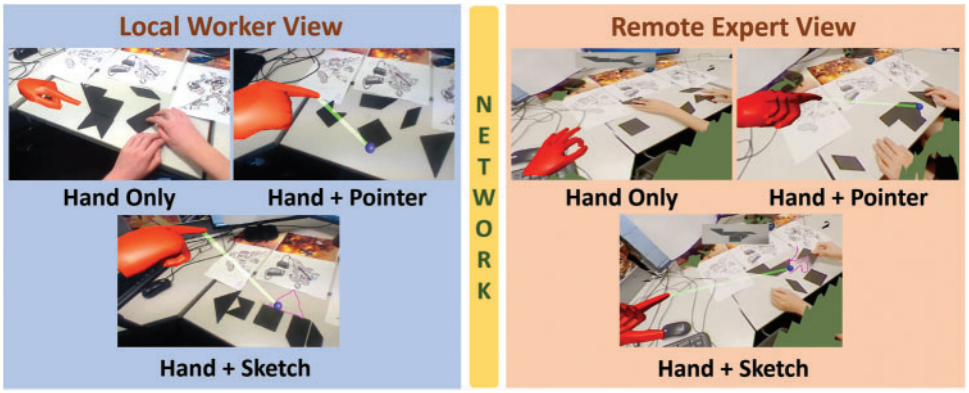

Figure 3: The combinations of visual communication cues with an independent view. The left three pictures show the use of the visual cue combinations in the local worker's perspective, and the right three in the remote expert's perspective

The visual cue conditions were:

- Hand Only: Collaborators communicate with the help of the remote expert's hand gesture shared in the 3D task space.
- Hands + Pointer: The pointer cue is available in addition to the hand gesture cue.
- Hands + Sketch: The sketch cue is available in addition to the hand gesture cue.

We chose the three combinations because recent HMDs such as Microsoft HoloLens and Magic Leap One support hand gesture by default and we use specific hand gesture poses to trigger pointer and sketching cues, so they do not require additional input device, increasing portability.

3.2.2 Procedure & Data Collection

We recruited participants in pairs and each pair had three sessions for three visual cue combinations and each of compared between two view types in each session. The order of the view types was random. For the task, we used Tangram assembly [44] as used in previous studies [2,4] (see Figs. 1–3). The Tangram is a puzzle putting seven flat pieces together to form a shape of a model without

overlapping each other. To minimize the effect of previous experience, we customized the Tangram puzzle to have difference sizes and shapes compared to the original design. We prepared seven Tangram puzzles for a face-to-face sample task and two sessions of three visual cue combinations ($1 + 2 \times 3 = 7$).

The user study started with welcoming the pair of participants and getting them to sign a consent form and fill out a demographic questionnaire asking for their age, gender, and the level of familiarity with AR/VR devices such as HMDs and video conferencing system. We randomly assigned the role of a remote expert and a local worker and let them collaborate face-to-face with a sample task. This face-to-face collaboration was to ensure that the participants understood the task before trying the conditions and remote collaboration. After completing a sample puzzle face-to-face, we explained how the system works and instructed participants about using visual communication cues. Next, we prepared the system for use including share of the mesh reconstruction and live video update.

The participant playing the role of a local worker (local participant afterward) had sat at a desk which was the task space (see Figs. 1 and 3). After sitting, the pair of participants performed two rounds of tasks in two view types with a given visual cue combination. Each round included four steps: 1) remote participants getting acquainted with the task, 2) practicing the use of the given condition, 3) performing the collaboration task with the given condition, and 4) answering questionnaires. In the getting acquainted step, the remote participant (the remote expert) learnt how to solve the Tangram task by completing it by him/herself with an instruction paper, so the remote participant was able to smoothly give instructions to the local participant. In the practice step, both local and remote participants tried the given condition and became familiar with it. Then, they performed the task with the given condition while the instruction paper used in the getting acquainted step was displayed at the top of the remote participants' view to help them to remember the instruction. After completing the task, they answered the questions from several questionnaires; asking the level of understanding messages and co-presence from Harms and Biocca's questionnaire [45], usability from the system usability scale (SUS) questionnaire [46], and the level of mental effort from the subjective mental effort (SMEQ) questionnaire [47]. They repeated these four steps in each round of the six sessions with three visual cue combinations and two view types.

After each session, participants ranked the view types according to their preference with the given visual cue combination and answered an open-ended question asking for the reason of their preference. Overall, the user study took about 80 min per pair of participants.

4 Results

We recruited 24 participants in pairs from the local university. We conducted 24 rounds of the user study by swapping the roles of participants between a local worker and a remote expert. There were 16 males and 8 females, and their ages were from 23 to 37 years old ($M = 29.4$; $SD = 4.9$). Participants reported a moderate level of familiarity with HMDs and video conferencing systems with the average rating 4.3 ($SD = 2.1$) and 5.2 ($SD = 1.9$) on a seven-point rating scale (1 = Novice, 7 = Expert), respectively.

We analyzed the participants' answer to the SUS, co-presence, understanding messages, and SMEQ questionnaires by using Wilcoxon Singed-Rank Test for pair wise comparison ($\alpha = .05$) between dependent and independent view conditions with each combination of visual communication cues. Since the experience according to the role (local worker and remote expert) could be different, we separately analyzed the results by the role. The overall results are shown in Table 3 and Fig. 4.

Table 3: The averages, standard deviations, and data analyzed results of the participants' Likert scale ratings on SMEQ (required mental effort), SUS (usability), co-presence, and understanding message when having dependent (D) and independent (I) views with Hand only, Hand + Pointer, Hand + Sketch visual cues

			Hand only		Hand + Pointer		Hand + Sketch	
			D	I	D	I	D	I
Local participants	SMEQ	Average	21.89	34.86	29.89	19.83	18.67	32.25
		S.D	14.8	16.49	15.72.	10.71	15.77	20.49
		Statistics	$Z = -2.420, p = .016$		$Z = -2.209, p = .027$		$Z = -1.632, p = .103$	
	SUS	Average	76.39	64.03	71.53	80.14	73.36	66.36
		S.D	12.01	15.10	12.28	12.62	11.77	15.93
		Statistics	$Z = -2.226, p = .026$		$Z = -2.076, p = .038$		$Z = -2.729, p = .006$	
	Co-presence	Average	36.44	30.67	34.56	37.78	37.17	34.39
		S.D	4.91	5.77	4.8	3.85	4.27	4.23
		Statistics	$Z = -2.382, p = .017$		$Z = -1.901, p = .057$		$Z = -1.965, p = .049$	
	Understanding	Average	33.11	28.44	31.28	35.94	34.56	30.50
		S.D	5.94	5.38	5.09	5.18	5.66	5.93
		Statistics	$Z = -2.205, p = .027$		$Z = -2.203, p = .028$		$Z = -2.252, p = .024$	
Remote participants	SMEQ	Average	20.17	29.29	31.33	20.7	28.33	43.19
		S.D	14.41	19.38	17.22	13.76	20.14	22.66
		Statistics	$Z = -2.017, p = .044$		$Z = -2.237, p = .025$		$Z = -2.1, p = .036$	
	SUS	Average	75.14	63.33	68.33	76.94	75.28	58.61
		S.D	13.86	11.11	16.84	13.02	16.17	15.29
		Statistics	$Z = -2.290, p = .022$		$Z = -1.850, p = .064$		$Z = -2.393, p = .017$	
	Co-presence	Average	36.06	30.44	33.39	36.94	34.0	29.67
		S.D	4.84	5.14	6.4	4.42	4.77	4.02
		Statistics	$Z = -2.070, p = .038$		$Z = -1.971, p = .049$		$Z = -2.56, p = .010$	
	Understanding	Average	32.56	30.06	31.39	35.28	33.67	30.61
		S.D	5.42	5.11	5.69	5.47	6.0	6.26
		Statistics	$Z = -1.348, p = .178$		$Z = -2.228, p = .026$		$Z = -1.722, p = .085$	

4.1 Hand Only Visual Cue

The local participants generally gave better points with the dependent view than with the independent view when using a Hand Only visual cue. When using the Hand Only visual cue, they felt requiring less mental effort ($Z = -2.420, p = .016$), better usability ($Z = -2.226, p = .026$), co-presence ($Z = -2.382, p = .017$), and understanding messages ($Z = -2.205, p = .027$) with the dependent view than the independent view. The ratings from the remote participants showed a similar trend as the one from local participants. When using the Hand Only cue, they felt requiring less mental effort ($Z = -2.420, p = .016$), better usability ($Z = -2.226, p = .026$) and co-presence ($Z = -2.382, p = .017$) with the dependent view than the independent view. The ratings of understanding messages were not significantly different between the two views when using the Hand Only cue ($Z = -1.348, p = .178$). This might be because remote participants mostly focused on giving messages (i.e., instructions) rather than understanding message from the local partner.

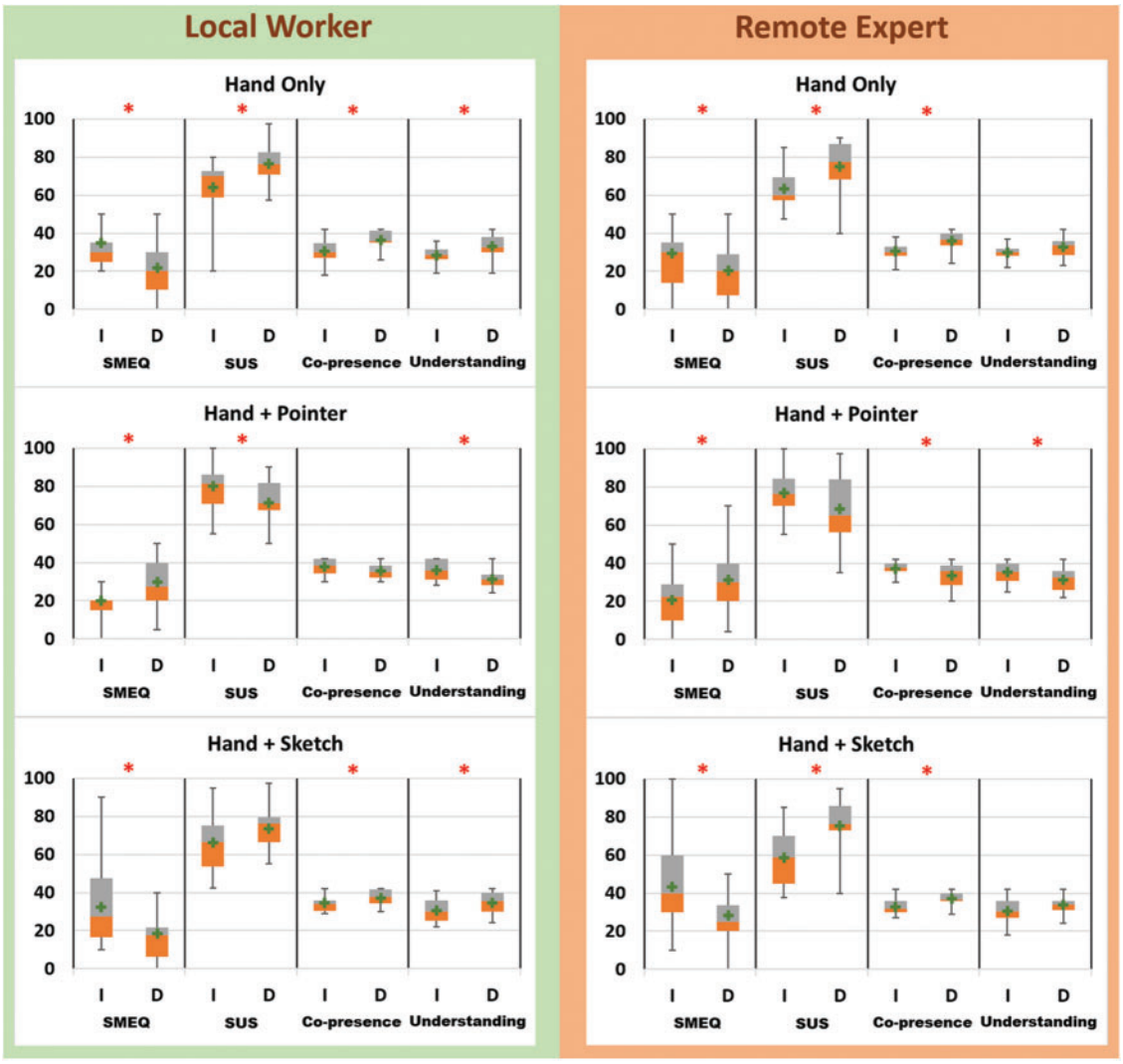

Figure 4: Graphs of the Likert-scale ratings on SMEQ (required mental effort), SUS (usability), co-presence, and understanding message when having dependent (D) and independent (I) views with Hand only, Hand + Pointer, Hand + Sketch visual cues (The significant result is marked with the red star, '∗')

The comments from the participants also showed similar results as the Likert scale ratings. Three participants reported that showing hands and gestures helped them to have better collaboration with the dependent view, but nine participants commented on an issue with the independent view: the virtual hands were not aligned with the target object when remote and local participants had different view positions in the independent view condition. In short, the Hand Only cue was not significantly useful with the independent view while it was in the dependent view condition.

Overall, sixteen participants preferred using the Hand Only cue with the dependent view over the independent view while the other two were opposing.

4.2 Hand + Pointer Visual Cue

The participants mostly gave better scores in the independent view condition than in the dependent view condition when using the Hand + Pointer visual cue. When using the Hand + Pointer visual cue, the local participants felt less required mental effort ($Z = -2.209$, $p = .027$), better usability ($Z = -2.076$, $p = .038$) and understanding message ($Z = -2.203$, $p = .028$) in the independent view than in the dependent view condition. The ratings of co-presence did not show significance between dependent and independent views ($Z = -1.901$, $p = .057$), but the p-value is close to the significant level (.05). When using the Hand + Pointer cue, the remote participants felt less required mental effort ($Z = -2.237$, $p = .025$), better co-presence ($Z = -1.971$, $p = .049$) and understanding messages ($Z = -2.228$, $p = .026$) in the independent view than in the dependent view condition. The ratings of usability did not show the significance between the two views ($Z = -1.850$, $p = .064$), but the p-value is close to the significant level (.05).

The participants comments were also aligned with the questionnaire results. Four participants mentioned that the pointer was not useful in the dependent view because the Hand + Only cue could also show the pointing information with hand gestures. Ten participants reported on the usefulness of the pointer cue in the independent view. Since the hand gesture was not aligned with the target object in the independent view, the pointer cue that could point at the target object was essential for providing pointing spatial information.

Overall, fourteen participants preferred using the Hand + Pointer cue in the independent view condition than in the dependent view condition, and the other four had the opposite view.

4.3 Hand + Sketch Visual Cue

The participants felt that they had better collaboration in the dependent view condition than in the independent view condition when using the Hand + Sketch visual cue. When using the Hand + Sketch visual cue, the local participants felt better usability ($Z = -2.729$, $p = .006$), co-presence ($Z = -1.965$, $p = .049$), and understanding messages in the dependent view condition than in the independent view condition. The ratings of required mental effort for using Hand + Sketch cue did not show significance between the two views ($Z = -1.632$, $p = .103$). When using the Hand + Sketch cue, the remote participants felt less required mental effort ($Z = -2.1$, $p = .036$), better usability ($Z = -2.393$, $p = .017$), and co-presence ($Z = -2.56$, $p = .010$) in the independent view condition than in the dependent view condition. The ratings of understanding messages did not show the significance between the two views ($Z = -1.722$, $p = .085$).

The questionnaire results were also supported by the participants' comments. Three participants reported that they could draw instructions by showing the position and orientation of the Tangram with sketches and it was easy to do this in the dependent view condition. Seven of them reported that the sketch cue was not easy to draw in the independent view condition. In our user study, the independent view mostly had further distance to the objects than the dependent view, so the sketches were drawn much bigger in the independent view condition than in the dependent view condition and participants had to be much more careful to correctly draw the sketch in the independent view condition than in the dependent view condition. In other words, the drawing in the independent view condition was much more difficult than in the dependent view condition and required precise controls.

Overall, thirteen participants preferred using the Hand + Sketch cue in the dependent view condition than in the independent view condition, with the other five having opposing views.

To sum up, the dependent view was well-matched with the Hand Only and Hand + Sketch cues, and the independent view was suited for the Hand + Pointer cue. When using Hand only and Hand + Sketch cues, the communication between a remote expert and a local worker required less mental effort, had higher levels of usability and co-presence, and supported better understanding of messages in the dependent view condition compared to the independent view condition. In contrast, when using the Hand + Pointer cue, these benefits were found in the independent view condition rather than in the dependent view condition.

5 Discussion

In this paper, we explored the interaction between view type and visual communication cues in AR remote collaboration. There were two view types: dependent and independent views, and three combinations of visual cues: Hand Only, Hand + Pointer, Hand + Sketch. There is recent trend in current HMDs towards supporting hand gesture interaction, so we used the hand gesture (hand only) as a baseline, and specific hand poses for triggering the pointer and sketching cues. These designs may not be available in other remote collaboration systems, for example using mouse or gaze for pointer and sketch cues, so our study results would not be valid for the systems. However, we want to report that portability is one of the key characteristics for these days' mixed reality systems and the hand gesture cue already becomes essential and based interaction method for them.

Our results showed that the use of visual communication cues could be influenced by the view types. Hand Only was powerful in the dependent view condition, but not in the independent view condition because it was not aligned on the target Tangram. The hand gesture cue was rendered at a position relative to the viewpoint and people interpreted the hand gesture messages in the task space reference. However, the viewpoint in the independent view condition could be different between the local worker and the remote expert, so interpreting hand gesture messages in the reference task space when having an independent view could be meaningless from the local worker's viewpoint. Therefore, our first hypothesis (The hand gesture cue is a powerful visual cue regardless of the view type) was not supported.

The Hand + Pointer cues was a powerful alternative when the hand gesture cue had an issue in the independent view condition. Since the hand gesture was not aligned with the target Tangram pieces from the local worker's viewpoint in the independent view, participants used the pointer cue. In contrast, while the hand gesture cue was useful in the dependent view condition, the pointer cue was not useful because the hand gesture could also show pointing information. Therefore, our second hypothesis (The additional pointer cue is not very crucial when using it with a hand gesture cue) was partly valid.

The use of Hand + Sketch cues was significantly impacted by the distance to the target object. When drawing sketches, the distance was the key characteristics influencing their size. Since our target task object was smaller than the hand palm (Tangram pieces) and the independent viewpoint was further away from the Tangram pieces than the dependent view, a small incorrect drawing or small handwringing might have much more impact in the independent view condition than in the dependent view condition. Therefore, the Hand + Sketch cues was less useful in the independent view condition than in the dependent view condition and our third hypothesis (The view type does not influence the use of additional sketch cues) was not supported. The remote participants in our study sat on a chair so they did not come closer to the target Tangram pieces and kept a distant view which might have affected the results.

6 Conclusion

In this paper, we explored the relationship between visual communication cues and view types in remote collaboration. Our remote collaboration system supported three visual communication cues (Hand Only, Hand + Pointer, Hand + Sketch) and two view types (dependent and independent views), and we conducted a user study to investigate the influence of the view types on the use of visual communication cues. From the user study, we found that the hand gesture cue was the most useful in the dependent view condition, so the additional pointer cue was not crucial when using it together with the hand gesture. However, unlike the pointer cue, the additional sketch cue with the hand gesture was still important in the dependent view condition. This is because the sketch cue could precisely show the final position and orientation of any object manipulation. However, using visual cues in the independent view condition showed conflicting results. The hand gesture in the independent view condition had an issue of not being aligned well with the target object because the local worker and the remote expert had different viewpoint positions. In this case, the additional pointer cue was a useful alternative. The additional sketch cue was not easy to draw in the independent view condition because the drawing activity required very precise finger movements to correctly draw sketches.

In the future, we plan to explore the use of the combination of gaze and gesture cues for enhancing communication. They each have different characteristics, so the combination should make collaboration more effective. In addition, we plan to add virtual avatars during the remote collaboration and explore the effect of different avatar representations. We are especially interested in the effect of the avatar size, which may influence gaze movement. When a collaborator switches his/her gaze point between their partner and the target object, if the avatar is small, the amount of gaze movement may be much less than with a normal sized or large avatar.

Acknowledgement: This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by Korea government Ministry of Science, ICT (MSIT) (No. 2019-0-01343, convergence security core talent training business) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT). (NRF-2020R1A4A1019191).

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. Huang, L. Alem, F. Tecchia and H. B. L. Duh, "Augmented 3D hands: A gesture-based mixed reality system for distributed collaboration," *Journal on Multimodal User Interfaces*, vol. 12, no. 2, pp. 77–89, 2018.

[2] A. Jing, G. Lee and M. Billinghurst, "Using speech to visualise shared gaze cues in MR remote collaboration," in *Proc. 2022 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR),* Christchurch, New Zealand, pp. 250–259, 2022.

[3] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst and M. Adcock, "A technique for mixed reality remote collaboration using 360 panoramas in 3D reconstructed scenes," in *Proc. 25th ACM Symp. on Virtual Reality Software and Technology*, Parramatta, Australia, pp. 1–11, 2019.

[4] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo *et al.,* "Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration," in *Proc. 2019 CHI Conf. on Human Factors in Computing Systems*, Glasgow, Scotland, UK, paper no. 173, pp. 1–13, 2019.

[5] K. Gupta, G. A. Lee and M. Billinghurst, "Do you see what I see? The effect of gaze tracking on task space remote collaboration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2413–2422, 2015.
[6] S. Johnson, M. Gibson and B. Mutlu, "Handheld or handsfree? Remote collaboration via lightweight head-mounted displays and handheld devices," in *Proc. of the 18th ACM Conf. on Computer Supported Cooperative Work & Social Computing*, Vancouver BC Canada, pp. 1825–1836, March 2015.
[7] S. Chen, M. Chen, A. Kunz, A. E. Yantaç, M. Bergmark *et al.,* "SEMarbeta: Mobile sketch-gesture-video remote support for car drivers," in *Proc. 4th Augmented Human Int. Conf. (AH '13)*, Stuttgart, Germany, pp. 69–76, 2013.
[8] S. Kim, M. Billinghurst and K. Kim, "Multimodal interfaces and communication cues for remote collaboration," *Journal on Multimodal User Interfaces*, vol. 14, no. 4, pp. 313–319, 2020.
[9] C. Gutwin and S. Greenberg, "The effects of workspace awareness support on the usability of real-time distributed groupware," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 6, no. 3, pp. 243–281, 1999.
[10] K. Schmidt, "The problem with 'awareness': Introductory remarks on 'awareness in CSCW'," *Computer Supported Cooperative Work (CSCW)*, vol. 11, no. 3, pp. 285–298, 2002.
[11] S. Greenberg and C. Gutwin, "Implications of we-awareness to the design of distributed groupware tools," *Computer Supported Cooperative Work (CSCW)*, vol. 25, no. 4–5, pp. 279–293, 2016.
[12] S. Kim, G. Lee, M. Billinghurst and W. Huang, "The combination of visual communication cues in mixed reality remote collaboration," *Journal on Multimodal User Interfaces*, vol. 14, no. 4, pp. 321–335, 2020.
[13] H. Ishii, M. Kobayashi and K. Arita, "Iterative design of seamless collaboration media," *Communications of the ACM*, vol. 37, no. 8, pp. 83–97, 1994.
[14] J. C. Tang and S. L. Minneman, "Videodraw: A video interface for collaborative drawing," *ACM Transactions on Information Systems*, vol. 9, no. 2, pp. 170–184, 1991.
[15] M. Stefik, D. G. Bobrow, G. Foster, S. Lanning and D. Tatar, "WYSIWIS revised: Early experiences with multiuser interfaces," *ACM Transactions on Information Systems (TOIS)*, vol. 5, no. 2, pp. 147–167, 1987.
[16] J. Rekimoto and K. Nagao, "The world through the computer: Computer augmented interaction with real world environments," in *Proc. UIST '95: Proc. of the 8th Annual ACM Symp. on User Interface and Software Technology*, Pittsburgh, Pennsylvania, USA, pp. 29–36, 1995.
[17] L. Alem, F. Tecchia and W. Huang, "Remote tele-assistance system for maintenance operators in mines," in *Proc. 11th Underground Coal Operators' Conf.*, University of Wollongong & the Australasian Institute of Mining and Metallurgy, Australia, pp. 171–177, 2011.
[18] S. Kasahara and J. Rekimoto, "Jackin: Integrating first-person view with out-of-body vision generation for human-human augmentation," in *Proc. 5th Augmented Human Int. Conf.*, Kobe, Japan, pp. 1–8, 2014.
[19] H. Kuzuoka, T. Kosuge and M. Tanaka, "Gesturecam: A video communication system for sympathetic remote collaboration," in *Proc. 1994 ACM Conf. on Computer Supported Cooperative Work*, Chapel Hill, North Carolina, USA, pp. 35–43, 1994.
[20] N. Sakata, T. Kurata, T. Kato, M. Kourogi and H. Kuzuoka, "WACL: Supporting telecommunications using wearable active camera with laser pointer," in *Proc. Seventh IEEE Int. Symp. on Wearable Computers*, White Plains, NY, USA, pp. 53–56, 2003.
[21] S. Gauglitz, B. Nuernberger, M. Turk and T. Höllerer, "World-stabilized annotations and virtual scene navigation for remote collaboration," in *Proc. 27th Annual ACM Symp. on User Interface Software and Technology*, Honolulu, Hawaii, USA, pp. 449–459, 2014.
[22] T. Teo, L. Lawrence, G. Lee, M. Billinghurst and M. Adcock, "Mixed reality remote collaboration combining 360 video and 3D reconstruction," in *Proc. 2019 CHI Conf. on Human Factors in Computing Systems*, Glasgow, Scotland, UK, paper no. 201, pp. 1–14, 2019.
[23] S. Nagai, S. Kasahara and J. Rekimoto, "Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration," in *Proc. Ninth Int. Conf. on Tangible, Embedded, and Embodied Interaction*, Stanford, California, USA, pp. 113–116, 2015.

[24] H. Bai, P. Sasikumar, J. Yang and M. Billinghurst, "A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing," in *Proc. 2020 CHI Conf. on Human Factors in Computing Systems*, Honolulu HI, USA, pp. 1–13, 2020.

[25] L. Gao, H. Bai, G. Lee and M. Billinghurst, "An oriented point-cloud view for mr remote collaboration," in *Proc. SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, Macau, paper no. 8, pp. 1–4, 2016.

[26] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst and M. Adcock, "A technique for mixed reality remote collaboration using 360 panoramas in 3D reconstructed scenes," in *Proc. 25th ACM Symp. on Virtual Reality Software and Technology*, Parramatta, NSW, Australia, article no. 23, pp. 1–11, 2019.

[27] S. Kim, G. Lee, N. Sakata and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," in *Proc. 2014 IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*, Munich, Germany, pp. 83–92, 2014.

[28] K. Higuch, R. Yonetani and Y. Sato, "Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks," in *Proc. 2016 CHI Conf. on Human Factors in Computing Systems*, San Jose, California, USA, pp. 5180–5190, 2016.

[29] W. Huang, M. Wakefield, T. A. Rasmussen, S. Kim and M. Billinghurst, "A review on communication cues for augmented reality based remote guidance," *Journal on Multimodal User Interfaces*, vol. 16, no. 2, pp. 239–256, 2022.

[30] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer *et al.,* "Gestures over video streams to support remote collaboration on physical tasks," *Human-Computer Interaction*, vol. 19, no. 3, pp. 273–309, 2004.

[31] K. Gupta, G. A. Lee and M. Billinghurst, "Do you see what I see? The effect of gaze tracking on task space remote collaboration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2413–2422, 2016.

[32] G. Lee, S. Kim, Y. Lee, A. Dey, T. Piumsomboon *et al.,* "Improving collaboration in augmented video conference using mutually shared gaze," in *Proc. ICAT-EGVE '17: Proc. of the 27th Int. Conf. on Artificial Reality and Telexistence and 22nd Eurographics Symp. on Virtual Environments*, Adelaide, Australia, pp. 197–204, 2017.

[33] Y. Lee, C. Shin, A. Plopski, Y. Itoh, T. Piumsomboon *et al.,* "Estimating gaze depth using multi-layer perceptron," in *Proc. 2017 Int. Symp. on Ubiquitous Virtual Reality (ISUVR)*, Nara, Japan, pp. 26–29, 2017.

[34] S. Kim, G. A. Lee, S. Ha, N. Sakata and M. Billinghurst, "Automatically freezing live video for annotation during remote collaboration," in *Proc. 33rd Annual ACM Conf. Extended Abstracts on Human Factors in Computing Systems*, Seoul, Republic of Korea, pp. 1669–1674, 2015.

[35] O. Fakourfar, K. Ta, R. Tang, S. Bateman and A. Tang, "Stabilized annotations for mobile remote assistance," in *Proc. 2016 CHI Conf. on Human Factors in Computing Systems*, San Jose, California, USA, pp. 1548–1560, 2016.

[36] D. Kirk, T. Rodden and D. S. Fraser, "Turn it this way: Grounding collaborative action with remote gestures," in *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, San Jose, California, USA, pp. 1039–1048, 2007.

[37] W. Huang and L. Alem, "Gesturing in the air: Supporting full mobility in remote collaboration on physical tasks," *Journal of Universal Computer Science*, vol. 19, no. 8, pp. 1158–1174, 2013.

[38] L. Alem, F. Tecchia and W. Huang, "Handsonvideo: Towards a gesture based mobile AR system for remote collaboration," in *Proc. Recent Trends of Mobile Collaborative Augmented Reality Systems*, Springer, New York, NY, pp. 135–148, 2011.

[39] J. Odom, *Up Close & Personal with the Meta 2 Head-Mounted Display*, Santa Monica, CA, USA: Wonder How To Inc., [Online]. Available: https://meta.reality.news/news/hands-on-up-close-personal-with-meta-2-head-mounted-display-0178607/ (accessed on 15 June 2022).

[40] FOVE Inc., *FOVE Head Mounted Display*, Tokyo, Japan: FOVE Inc., [Online]. Available: https://fove-inc.com/ (accessed on 15 June 2022).

[41] Ricoh Company, Ltd., *Ricoh Theta V*, Tokyo, Japan: Ricoh Company, Ltd., [Online]. Available: https://theta360.com/ko/about/theta/v.html (accessed on 15 June 2022).
[42] Leap Motion, Inc., *LeapMotion Hand Tracking*, San Francisco, CA, USA: Leap Motion Inc., 3 March 2016. [Online]. Available: https://developer.leapmotion.com/ (accessed on 15 June 2022).
[43] Leap Motion, Inc., *Pinch Draw Example*, California, SF, USA: Leap Motion Inc., 3 March 2016. [Online]. Available: https://gallery.leapmotion.com/pinch-draw/ (accessed on 15 June 2022).
[44] Wikipedia contributors, *Tangram*, Wikimedia Foundation, 24 February 2022. [Online]. Available: https://en.wikipedia.org/wiki/Tangram (accessed on 15 June 2022).
[45] C. Harms and F. Biocca, "Internal consistency and reliability of the networked minds measure of social presence," in *Proc. Annual Int. Presence Workshop*, Valencia, pp. 246–251, 2004.
[46] J. Brooke, "SUS-A quick and dirty usability scale," in *Usability Evaluation in Industry*, 1st ed., Boca Raton, Florida, USA: CRC Press, pp. 189–194, 1996.
[47] F. R. H. Zijlstra, "Efficiency in work behavior: A design approach for modern tools," Doctoral Dissertation, Delft University of Technology, Delft, The Netherlands, 1993.