Evaluating the Combination of Visual Communication Cues for HMD-based Mixed Reality Remote Collaboration

Seungwon Kim
University of South Australia
Adelaide, Australia
Seungwon.Kim@unisa.edu.au

Gun Lee
University of South Australia
Adelaide, Australia
Gun.Lee@unisa.edu.au

Weidong Huang
Swinburne University of Technology
Melbourne, Australia
weidonghuang@swin.edu.au

Hayun Kim
KAIST
Daejeon, Korea
hayunkim@kaist.ac.kr

Woontack Woo
KAIST
Daejeon, Korea
wwoo@kaist.ac.kr

Mark Billinghurst*
University of South Australia
Adelaide, Australia
Mark.Billinghurst@unisa.edu.au

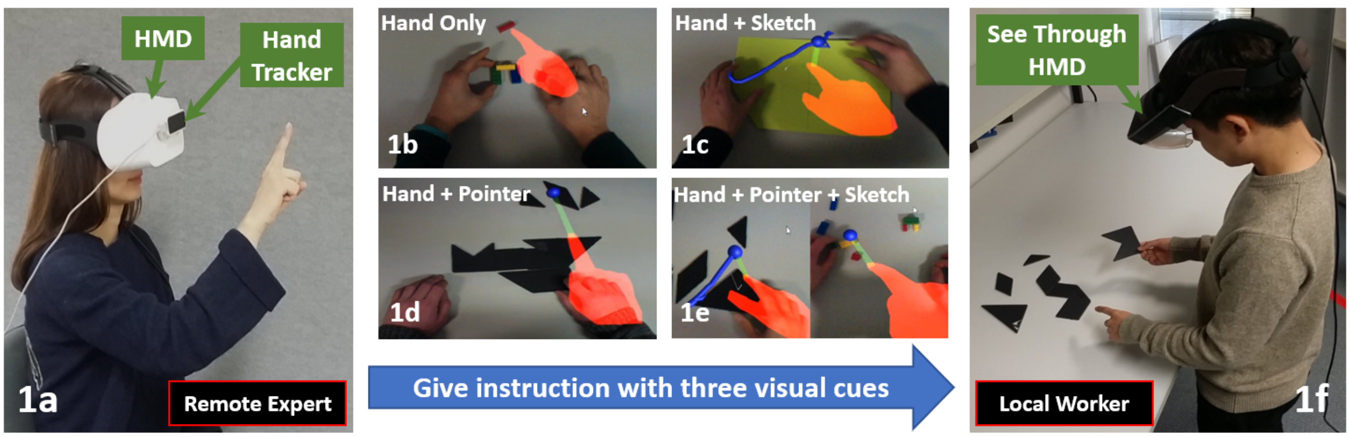

Figure 1: A study setup for remote collaboration between a remote expert (1a) and a local worker (1f) using four combinations of visual cues: hand only (1b), hand + sketch (1c), hand + pointer (1d), and hand + sketch + pointer (1e).

ABSTRACT

Many researchers have studied various visual communication cues (e.g. pointer, sketching, and hand gesture) in Mixed Reality remote collaboration systems for real-world tasks. However, the effect of combining them has not been so well explored. We studied the effect of these cues in four combinations: hand only, hand + pointer, hand + sketch, and hand + pointer + sketch, with three problem tasks: Lego, Tangram, and Origami. The study results showed that the participants completed the task significantly faster and felt a significantly higher level of usability when the sketch cue is added to the hand gesture cue, but not with adding the pointer cue. Participants also preferred the combinations including hand and sketch cues over the other combinations. However, using additional cues (pointer or sketch) increased the perceived mental effort and did not improve the feeling of co-presence. We discuss the implications of these results and future research directions.

*Corresponding Author

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300403

CCS CONCEPTS

• **Human-centered computing** → *Collaborative interaction*; **Mixed / augmented reality**; User studies.

KEYWORDS

Remote collaboration, Visual communication cue, Mixed Reality, Co-presence, Usability

ACM Reference Format:

Seungwon Kim, Gun Lee, Weidong Huang, Hayun Kim, Woontack Woo, and Mark Billinghurst. 2019. Evaluating the Combination of Visual Communication Cues for HMD-based Mixed Reality Remote Collaboration. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300403

1 INTRODUCTION

Compared to co-located collaboration, remote collaboration is more difficult as collaborators are not in the same place [18]. Technology for remote collaboration has been studied for decades to reduce this difficulty [15] by improving user interfaces for better sharing the local task space [45] and supporting better communication between collaborators [10]. In this paper, we mainly focus on the user interface for better communication between collaborators.

In a typical remote collaboration study with a real-world physical task (e.g. fixing a car or making a Lego model), there are two main participants: a local user in the task space and a remote collaborator [10, 26–28]. While the local user directly performs activities in the local task space, the remote collaborator could use technology (such as video conferencing) to see a view of the task space and use visual cues to aid communication. Typical visual communication cues are pointers [29, 35, 42], sketches [12, 13, 26–28], and hands [1, 20, 32, 44], and many researchers have explored the effect of these cues individually in remote collaboration studies.

However, to the best of our knowledge, there is no previous study investigating the effect of combining the common visual cues (i.e. pointer, sketching and hand gestures) together in Mixed Reality (MR) remote collaboration. In this paper, we explore and compare four different combinations: 1) hand only, 2) hand + pointer, 3) hand + sketch, and 4) hand + pointer + sketch - a 2 x 2 factorial design study where the hand gesture cue is the baseline condition with two independent variables (Pointer and Sketch) with each variable having two levels (On or Off).

This research makes the following novel contributions:

(1) Design and implementation of a prototype MR system supporting both local and remote collaborators wearing a head mounted display (HMD) and using combination of visual cues including the pointer, sketch, and virtual hand.
(2) Comparing four combinations of the typical visual communication cues (virtual pointer, sketch, and hand representation) with three physical tasks (Lego blocks, Tangram puzzle, and Origami) in MR remote collaboration.
(3) Measuring the effect of the combinations of visual cues on the co-presence and required mental effort

2 RELATED WORK

Many researchers have investigated the use of visual communication cues in remote collaboration on spatial tasks. In this section, we describe previous studies according to the type of visual cues (pointer, sketch, and hand gesture) used.

The Pointer Cue Using a pointer is a quick and precise way to indicate an object of interest. For example, GestureCam [35] used a pointer in a desktop remote collaboration setup. It allowed a remote user to control a laser pointer in the local task space, so the local user knew which object the remote user was referring to. Sakata et al. [42] put the laser pointer and camera on the local user's shoulder, so the remote user could have an independent view of the task space and point with the laser pointer. Fussell et al. [10] overlaid a virtual pointer on a shared live video with a desktop monitor, and Kim et al. extended it with a tablet interface [29]. Later, Kim et al. [27] compared the pointer and sketch cues and found that the pointer was more precise and quicker to provide a simple pointing information for positioning and selecting objects.

Prior work showed using a pointer can increase users' co-presence by encouraging a remote user to be active in collaboration [27]. Moreover, pointer cues also have been used for displaying collaborators' gaze in the shared task space, so the other user could know where the collaborator was looking at and increase the level of co-presence. Gupta et al. [16] developed a system sharing the local user's gaze pointer and remote user's mouse pointer and found both pointer cues increased participants' feeling of co-presence. Higuchi et al. [19] compared using hand gestures alone and hand gesture + remote user's gaze pointer together and found that adding the gaze pointer helped participants to better understand each other. Lee et al. [39] investigated the effect of sharing both local and remote users' gaze pointers and found that the gaze pointer increased both usersâĂŹ feeling of co-presence and the level of awareness of where the users were looking at.

The Sketch Cue Sketch cues have been shown to be effective for presenting spatial information for manipulating objects [25]. Fussell et al. [10] and Kim et al. [29, 30] compared pointer and sketch cues and found that participants completed tasks quicker using the sketch cue than with the pointer cue. One important feature of the sketch cue is permanence, with the sketches remaining where they were drawn [25, 27] showing lines or object shapes, while information from the pointer cue is ephemeral.

Early systems often had a fixed camera view to avoid sketches losing the reference frame (e.g. if the view changes, a sketch of an arrow may no longer point at the same object). For example, Tang et al. [47] developed 'Videodraw' sharing a task space with a top-down fixed camera and projecting it on the other end. Ishii et al. [22] developed 'ClearBoard' sharing a synchronized screen view with sketches. Fussell et al. [10] used a fixed camera on a tripod with sketches overlaid on the video.

Using Augmented Reality (AR) tracking techniques, sketches can remain where they are drawn, even when the view changes. For example, Kato and Billinghurst [24] stabilized the user sketches with marker-based tracking that determined the real-world position of the remotely added sketches. Fakourfar et al. [6] used marker-based tracking with a tablet, while Kim et al. [25, 27, 28] and Gauglitz et al. [13] employed SLAM (Simultaneous Localization and Mapping) tracking to stabilize and fix the user sketches to the real-world space.

In previous studies, there were two interesting results about sketch cues. First, participants preferred sketches to be automatically erased after a while when it is drawn because old sketches cluttered the view of the scene as they became no longer relevant [10, 29]. Second, the sketch cue increased the feeling of co-presence between local and remote users as it clearly showed remote user's explanation [27].

The Hand Gesture Cue Hand gestures can express diverse information [31] such as pointing, shapes, and even social cues such as appreciation (e.g. thumbs up) [14, 31] and emotion (e.g. making a heart shape with the hand gesture) [8]. Goldin-meadow [14] mentioned that hand gestures are often used with speech to provide spatial information such as pointing at an object or guiding listeners to focus on a specific area or an object. Researchers found that using hand gesture cues in remote collaboration helped to complete the task faster than without it [32, 33] as it supported easy explanation of object manipulation with pointing and hand shapes, and by showing the required hand movement. The hand gesture cue also increases the users' co-presence [21, 32, 34] as it showed the other user's hand motions.

Tasks Visual communication cues may have different effects according to the task [25–27]. Prior user studies with multiple tasks [6, 19] showed different effects of using visual communication cues. For example, for Origami task which requires 2D information on how to fold a paper, a sketch cue was a good way to show a folding line [6]. However, it was difficult to use sketches to describe how to insert an object at a certain depth in a car repair task [48].

The most relevant work is Chen's study [4] which compared a combination of hand gestures and sketch cues to a video only condition (no additional visual cues) and Higuch's study [19] that compared a hand gesture condition to a hand gesture + gaze pointer condition. However, there are several differences between their studies and ours. First, we investigate the combinations of three typical visual communication cues while they explored the effect of combining two visual cues. Second, in Chen's system [4], the hand capture area and display monitor were not in line with the remote user's perspective, while they are in our system. Third, Higuch et al. used a desktop monitor for remote user's display and studied a user gaze pointer, but we use a HMD and a hand pointer. Most recently, Teo et al. [48] used the three cues in a live 360 video-based MR remote collaboration system. However, their study did not compare the sketch and pointer cues in different combinations, and also did not look into social aspects, such as co-presence.

Overall, the three visual communication cues have different strengths in sharing spatial information and all three cues have been shown to improve the user's feeling of co-presence. However, no prior work further investigated different combinations of these cues. Based on the findings from prior work, we suggest the following five hypotheses:

- **H1** Using more visual communication cues in combination results in faster task completion time.
- **H2** Using more visual communication cues in combination results in better usability.
- **H3** Having more visual communication cues in combination leads users to have higher feeling of co-presence.
- **H4** Using more visual communication cues in combination reduces the mental effort for remote collaboration.
- **H5** The effect of using the combinations of three visual communication cues (hand, sketch, and pointer) would show different trends according to the type of task.

Apart from the visual cues, several researchers have also studied providing an independent view to the remote user [9, 11, 13, 21, 23, 26, 38, 41, 46] and different types of hand gesture cues (virtual hand, 2D hands, and 3D hands) [21, 33]. Moreover, Kim et al. [26] compared two collaboration styles: mutual and remote expert collaborations. In a mutual collaboration, both the local and remote users do not have a solution for the task, so they need to discuss ideas together to find a solution [26–28, 39]. In a remote expert collaboration, a remote expert has the solution for the task and provides instruction to the local worker [10, 13, 20, 21]. In this paper, our system uses dependent view configuration which is known for its benefit of shared focus in remote expert style of collaboration.

3 METHODOLOGY

To investigate the use of combinations of visual communication cues, we developed a prototype MR remote collaboration system and conducted a user study with four combinations of visual communications cues used for three types of physical task in remote expert collaboration. In this section, we describe the prototype system and the user study design.

Prototype System

Devices. The prototype system setup includes two sides: a local worker side and a remote expert side (see Figure 1). Each side has a computer and they are connected to a local area network with Ethernet cable. The local worker uses a Meta2 [40] AR HMD connected to a PC (Intel Core i7-7700K 4.2GHz quad core CPU, 16 GB RAM, and NVIDIA GeForce GTX 1070 graphics card) (See Figure 1e). Meta2 is an optical see-through display with 90-degree field of view (FOV) and supports 2550 x 1440 resolution at 60 frames per second (fps). Meta2 supports SLAM visual tracking, so it can display virtual objects fixed at a position in the real-world even when the user is moving [27]. It also supports mesh reconstruction of the real-world, so a user can place virtual objects on the surface of it with the help of the ray casting method.

The remote expert wears a FOVE Virtual Reality (VR) HMD [7] with a Leap Motion hand tracker [37] attached to the front (See Figure 1a). Both of the devices are connected to another PC with the same specification as the one on the local worker side. The FOVE display has a 100-degree FOV with a resolution of 2560 x 1440 at 70 fps. The Leap Motion tracks hand at 0.7 mm accuracy [49].

User Interface. We implemented the software for AR and VR user interfaces with the Unity game engine (v2017.3.0f3). The local worker shares live video of his/her task environment including hand activities (i.e. hand gesture and object manipulation), while the remote expert can watch the shared lived video and share back the following visual cues:

- **Virtual Hands:** The remote expert's hand gestures in front of the FOVE is captured with the Leap Motion sensor then shared with the local worker by displaying virtual hands overlaid into the shared live video.
- **Pointer:** The remote expert can place a virtual pointer in the real-world by making a pointing gesture (see Figure 2a).
- **Sketch:** The remote expert can sketch in the real-world with a sketching gesture, lifting the thumb while pointing (see Figure 2b).

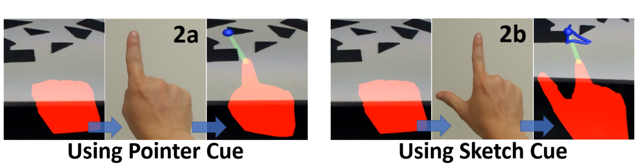

Figure 2: Finger pose to place a pointer (2a) and sketch (2b)

Since our study focuses on the visual communication cue, we did not implement independent view for the remote user. The remote expert is provided with the same view as the local worker, and this is commonly used in prior works [6, 16, 27, 28, 30, 39] for remote expert style collaboration.

The Leap Motion captures real time hand gesture with an infrared camera and the video frame of it is copied and transferred to the local worker side. The transferred hand frame is used to represent remote expert's hand gesture as a live animated 3D virtual hands and the Meta2 shows it in the real-world scene (see Figure 1). The AR scene is not only displayed on the Meta2 for the local worker but also captured and transferred back to the remote expert side and displayed in FOVE, so that the remote expert sees the same live video view of what the local worker sees including the virtual hands. The live video feed is updated in real time at about 20 fps in 1280 x 720 resolution. The 3D virtual hands are drawn semi-transparent to prevent them from occluding the task space.

Implementing the system in this way (i.e. sharing the composited AR video) has a benefit of not needing to synchronize all the data shared between the local and remote users as the shared live video already includes the virtual cues. On the other hand, if the virtual cues (including the virtual hands) are rendered and overlaid into the real-world scene on each side of the system, further calibration is necessary to synchronize the position and size of virtual cues appearing correctly in the real-world scene. Moreover, we note that the implementation is practical as the two sides are connected only through a local area network rather than connecting a remote expert's device to the local worker's PC (e.g. in our system, the Leap Motion capturing the remote expert's hands is connected to remote expert's PC, not to the local worker's PC).

In addition to sharing hand gestures, the remote expert can also use the pointer and/or sketch cues. To make it easy to transition from using the hand gestures to using the pointer or sketch cues, we used common hand gestures (see Figure 2). To use the pointer, users can simply point with an index finger, and to sketch, users simply have to lift their thumb while pointing.

While the virtual hands are visualized in the air and positioned relatively to the HMD, the pointers and sketches are projected onto the reconstructed surface of the real-world. To implement this, our system tracks the index fingertip and uses a ray casting method. The system casts a ray starting from the Meta2's 'Composite Camera' position towards the index fingertip (based on the transferred hand tracking frame from the Leap Motion on the FOVE HMD connected the remote expert's PC) to find the collision point with the reconstructed surface. The collision point is used for both positioning the pointer and drawing a sketch. With the SLAM visual tracking on Meta2, the sketch is shown in a world stabilized manner as in prior work [13, 27, 28]. We used a virtual sphere to visualize the pointer and implemented the sketch function based on the Leap Motion's 'PinchDrawDemo' [36] sample code. Based on prior work [10, 27, 29], we made the sketches to disappear after one second from when it was drawn. To show the linkage between the index fingertip and the visual cues (pointer and sketches), we visualized a ray in semi-transparent green color (see Figure 1c and 1d).

To implement the real-world surface reconstruction, we used the reconstruction example from the Meta2 Unity package and integrated it into our prototype. Since the reconstructed surface is needed for positioning the pointer and drawing the sketches, the system needs to perform reconstruction before using the pointer or sketch cues.

User Study Design

Using the prototype system, we conducted a user study to compare different combinations of visual communication cues. We selected four combinations of pointer and sketch cues (as listed in section 1) with the baseline of showing only the virtual hand. This choice was made following the recent trend of new HMDs (such as Microsoft HoloLens, Meta2, and Magic Leap One) supporting hand gesture interaction by default. Moreover, if excluding the hand gesture while using pointer or sketch, it will require another input method, such as a handheld controller, which could be a confounding factor of different input methods used between conditions. Thus, the study was in a 2x2 factorial within subject design, with the two independent variables being the pointer and sketch cues added on top of the virtual hand cue. This formed the four conditions in the study as following:

- **Hands Only (HO):** The remote expert's hand gesture is shared with the local worker as virtual hands in the shared live video.
- **Hands + Pointer (HP):** In addition to the hand gesture, the remote expert can place a virtual pointer in the real-world by making a pointing gesture (see Figure 2a).
- **Hands + Sketch (HS):** In addition to the hand gesture, the remote expert can sketch in the real-world view with a sketching gesture (see Figure 2b).
- **Hands + Pointer + Sketch (HPS):** The remote expert can use all of visual cues including hand gesture, pointing, and sketch.

Procedure & Data Collection. We recruited participants in pairs and each pair had two rounds of an experiment by swapping their roles between a remote expert and a local worker.

The user study started with a pair of participants filling out a demographic questionnaire asking gender, age, and the level of familiarity with video conferencing and hand gesture interaction. Next, we randomly assigned the role of the local worker and the remote expert to each participant, and the researcher explained the three experimental tasks (see Figure 3) and demonstrated how the system works. Then they tried a set of three sample tasks while collaborating face-to-face. This face-to-face collaboration was to ensure that the participants understood the tasks before the experimental trials.

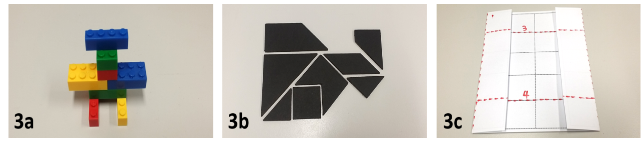

Figure 3: Tasks: Lego, Tangram, and Origami.

After practicing in face-to-face, the remote expert sat on a chair wearing the FOVE HMD (see Figure 1a) and the local worker stood in front of another desk wearing the Meta2 HMD (see Figure 1e). The two participants were back to back in the same room, so they were able to talk but could not see each other, similar to the set-ups used in prior work [29, 30].

There were four different conditions. Each condition included practice of using the interface given in the condition, getting acquainted with the task, performing collaborative task, and answering a questionnaire. During practice, the pair had time to become familiar with the given interface. Next, the participant playing the remote expert got acquainted with the task, learning the solution by completing the task by him/herself using instruction papers. This step was for letting the remote expert become familiar with the task enough to give instructions to the local worker. The Lego instruction included three pictures showing the assembly steps, and the Tangram task instruction showed a completed model with clearly marked border lines between pieces (see Figure 3b). The Origami instruction showed red dotted lines with numbers indicating the folding order (see Figure 3c).

After practicing with the interface in the given condition, the participants completed the three tasks given in random order. To help the remote expert remember the instructions, the instructions (the same as given in the acquaintance step) was also displayed at the top of the remote expert's view in HMD. After completing each task, both participants filled out a SUS [3] questionnaire, this was to see how the user's perceived usability under different tasks. After finishing all three tasks they also answered to a SMEQ [51, 52] and a co-presence questionnaire [17] for overall experience regardless of the task. They repeated this procedure in each of the four conditions, and the order of the conditions was counter balanced using a balanced Latin-square design. After finishing all four conditions, participant ranked the conditions according to their preference and answered to four open questions asking the reason for the ranking and their opinions about the four combinations they used.

After answering the ranking questionnaire, participants changed their role and repeated the experimental procedure with another round of four conditions as described above. The study took about two hours for each pair of participants.

In addition to the subjective measures, we collected objective data from screen video recordings, system log data, and observation.

Task. Participants tried each condition with three tasks: assembling Lego, Tangram, and Origami (see Figure 3). The Lego task was to make a model using eight blocks in various size and colors (three 2x4, three 1x4, and two 2x2 in red, yellow, blue, and green). The blocks are comparatively small as the size of a 2x2 block is about 16 mm x 16 mm with 8.6 mm height. The Tangram task was to make a shape by arranging eight pieces of black cardboard, without any piece overlapping with another. The size of Tangram pieces was comparatively large as the long edge of the smallest triangle piece was 77mm. The Origami task was to fold an A4 size paper for four times to form a given shape. A 5 x 6 grid of lines was printed on the paper to help participants communicate where the paper should be folded.

The main reason we prepared these three tasks was to cover wide range of tasks, as conclusions based on the results from a single task can be biased by the task characteristics. The three tasks were chosen based on previous studies (Origami [6], Tangram [6, 27, 39], and Lego [6, 16]). With the three tasks, our study can cover different sizes of objects to manipulate (between small Lego blocks and large Tangram pieces) and different types of spatial information required (pointing and selecting an object, describing shapes of objects, and describing specific lines for folding paper).

Since the study included a face-to-face collaboration and two rounds (with different role allocation) of four experimental conditions, we prepared nine sets of three tasks and balanced the level of difficulty between the sets by constraining each set to include the same number of object manipulations.

Participants. We recruited 16 participants (in pairs) from our university staff and students, and conducted 16 rounds of the user study (by swapping the roles of participants as a local worker and a remote expert). There were 11 males and 5 females with their ages ranging from 22 to 37 years old (M=28.4; SD=4.9). Participants expressed their moderate level of familiarity with VR/AR, hand gesture interaction, and video conferencing system, by rating on a seven-point rating scale (1 = Novice, 7 = Expert) which resulted on average 5.2 (SD=1.9), 4.3 (SD=2.1), and 4.9 (SD=1.9), respectively.

4 RESULTS

In this section, we report the results of task completion time (relating to hypothesis **H1**), usability (**H2**), and co-presence (**H3**), and mental effort (**H4**). We compared each measure between the four conditions with all three tasks aggregated. In addition, to further investigate how the results vary depending on the task (**H5**), we also analyzed the task completion time and usability data in each task. This helps us understand how the visual cues affect task performance and usability differently under each type of task, in addition to giving a general perspective on how the visual cues affect co-presence and level of required mental effort. Moreover, we also split the results in roles of the participants where relevant, but neither the role nor the task was treated as an independent variable in the analysis of the results as our main focus was investigating the effect of visual communication cues rather than exploring the effect of the task or the participant's role (e.g. comparing completion time of different type of tasks is not meaningful).

To analyze the results, we used two-way repeated measures ANOVA (α = .05) with the Pointer cue and the Sketch cue being the two factors, each with two levels (On or Off). In case the data is in ordinal scale (e.g. subjective rating or ranking results), we applied Aligned Rank Transform (ART) as proposed by Wobbrock et al. [50].

The main results are summarized as:

- In general, participants preferred the conditions with the sketch cue among the four conditions.
- Adding the pointer cue did not have any significant effect on the remote collaboration tasks.
- Adding the sketch cue significantly improved the performance (task completion time and usability) in general, but not in certain task, such as Lego, which involved manipulating small-size objects.
- Neither the sketch nor the pointer cues had significant effect on participant's feeling of co-presence
- There was higher demand of mental effort when using the sketch and/or pointer cue(s) together with the hand gesture cue.

Task Completion Time

First, we analyzed the results of the total task completion time (aggregating all three tasks) as summarized in Figure 4.

Adding the sketch cue to the hand gesture cue enabled users to complete tasks significantly faster than without having it ($F(1,15)$= 8.452, p=.010, η_p^2=.332), but adding the pointer cue to the hand gesture cue did not ($F(1,15)$=0.654, p=.430). Descriptive statistics showed that the participants took about 9% less time on average to complete all the tasks when using the sketch cue (with: M=260.9 seconds, SE=11.5;

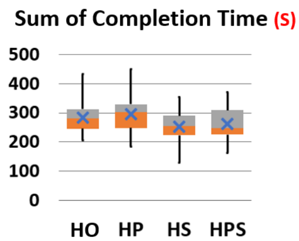

Figure 4: Average of the total task completion time in seconds (x: mean; S and P: significant effect of the sketch and pointer cues, respectively)

without: M=288.6 seconds, SE=11.1). No significant interaction effect was found between the pointer and sketch cues (F(1,15)= 0.529, p=.477).

For further investigation, we analyzed the effect of visual cues with the data from each task (see Figure 5). For the Lego task, there was no significant main effect of the sketch cue (F(1, 15)=.179, p=.678) nor the pointer cue (F(1, 15)=3.231, p=.090). There was no significant interaction effect between the pointer and sketch cues (F(1,15)=1.695, p=.210).

However, with the Tangram task, participants completed the task significantly faster with added sketch cue (F(1,15)= 7.667, p=.013, η_p^2 =.306), but not with the added pointer cue (F(1,15)=1.251, p=.279). Descriptive statics showed that they completed the Tangram task about 12% faster when using the sketch cue (with: M=108.2 seconds, SE=5.4; without: M=122.7 seconds, SE=5.1). No significant interaction effect was found between the two cues (F(1,15)=0.822, p=.377).

The Origami task also showed similar results where the sketch cue had a significant main effect on average task completion time (F(1,15)= 12.019, p=.003, η_p^2= .415), but the pointer cue did not (F(1,15)=2.351, p=.144). On average, participants completed the Origami about 17% faster with additional sketch cue (M=62.2 seconds, SE=5.1) than without it (M=74 seconds, SE=3.2). There was no significant interaction between the two cues (F(1,15)= 0.571, p=.482).

In summary, adding the sketch cue to the hand gesture cue helped participants complete the task faster than using hand gestures only, but adding the pointer cue to the hand gesture did not. The effect of the sketch cue was prominent with the Tangram and Origami tasks, but not with the Lego task.

Usability

Since the user interfaces were different according to the roles, we analyzed the results of the SUS questionnaire [3] separately for each role (see Figure 6). With taking average of all three tasks, the remote experts rated an average of 75.3, 75.2, 79.4, and 81.1 (SD: 14.1, 19.6, 15.7, and 14.6) for HO, HP, HS, and HPS conditions respectively, and these scores are in the range of 'Good' and acceptable usability [2]. The local workers rated 76.2, 78.5, 83.1, and 81.2 (SD: 10.7, 11.6, 12.4, and 11.1) for HO, HP, HS, and HPS conditions, and these scores are also in the range of 'Good' and acceptable.

First, we analyzed the average of the ratings from all three tasks. The remote expert felt higher level of usability when using the additional sketch cue than when without using it (F(1,15)=4.44, p=.05, η_p^2=.207), but they did not feel a significant effect when using the pointer cue with the hand gesture cue (F(1,15)=0.109, p=.745). No significant interaction was found between pointer and sketch cues (F(1,15)=0.762, p=.395). The local worker also felt higher level of usability when the sketch cue is added (F(1,15)=4.689, p=.045, η_p^2= .216). However, they did not feel a significant effect of the pointer (F(1,15)=0.053, p=.820). There was no significant interaction between the two cues (F(1,15)=0.233, p=.636).

For further investigation, we analyzed the usability ratings by each task. Figure 7 shows the SUS rating results of the remote experts for each task. In the Lego task, adding sketch cue or pointer cue to the hand gesture cue did not show significant effect on the usability (sketch: F(1,15) =0.564, p=.463; pointer: F(1,15)= 3.035, p=.100). There was no significant interaction between the sketch and pointer cues (F(1,15)= 0.154, p= 699).

However, with the Tangram task, participants felt a higher usability with the sketch cue (F(1,15)=6.198, p=.023, η_p^2=.267), but not with the pointer (F(1,15)= 0.02, p=.889). No significant

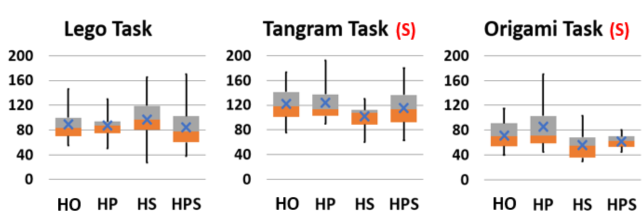

Figure 5: Average task completion time in seconds for Lego (left), Tangram (center), and Origami tasks (right) (x: mean; S and P: significant effect of the sketch and pointer cues, respectively)

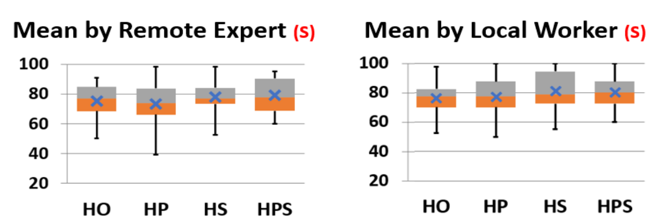

Figure 6: Average of the SUS ratings by remote experts and local workers (x: mean; 0: lowest usability ~ 100: highest usability; S and P: significant effect of the sketch and pointer cues, respectively)

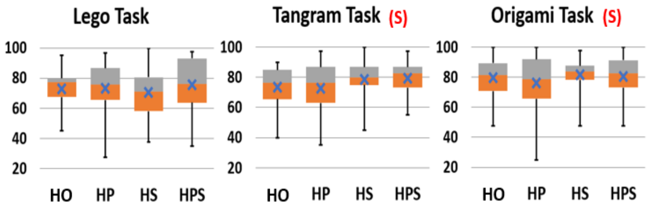

Figure 7: SUS rating results of remote experts for each task (x: mean; 0: lowest usability ~ 100: highest usability; S and P: significant effect of the sketch and pointer cues, respectively)

interaction was found between the two cues ($F(1,15)=0.248$, $p=.625$). Similar results were found with the Origami task where participants felt a significantly higher level of usability when using the sketch cue ($F(1,15)= 4.718$, $p=.044$, $\eta_p^2=.217$) but not with the pointer cue ($F(1,15)=3.078$, $p=.100$). There was no significant interaction between the two cues ($F(1,15)=1.400$, $p=.253$).

We also analyzed the local worker's ratings of the SUS questionnaire for each task as shown in Figure 8.

When assembling a Lego model, the local worker did not feel a significant effect of the pointer ($F(1,15)=3.619$, $p=.074$), or the sketch cue ($F(1,15)=0.038$, $p=.847$). No significant interaction was found ($F(1,15)=.004$, $p=.953$).

With the Tangram task, the local worker felt there was a significantly higher level of usability with added sketch cue than without it ($F(1,15)=10.630$, $p=.005$, $\eta_p^2=.385$), but not with the added pointer cue ($F(1,15)=.985$, $p=.335$). There was no significant interaction between two cues ($F(1,15)=.840$, $p=.372$).

Similarly, with the Origami task, the local worker felt there was a significantly higher level of usability with the additional sketch cue ($F(1,15)=4.680$, $p=.045$, $\eta_p^2=.216$), but there was no significant effect of the pointer cue ($F(1,15)=1.717$, $p=.208$). No significant interaction was found between the sketch and pointer cues ($F(1,15)=2.230$, $p=.154$)

Overall, adding the sketch cue to the hand gesture cue helped the participant to feel a higher level of usability, but

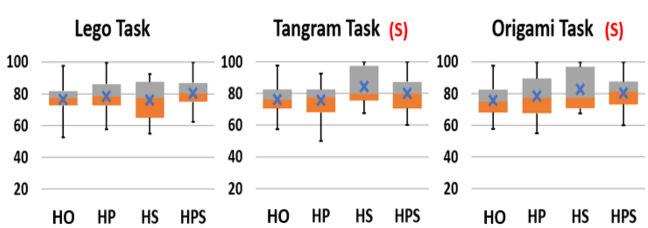

Figure 8: SUS rating results of local workers for each task (x: mean; 0: lowest usability ~ 100: highest usability; S and P: significant effect of the sketch and pointer cues, respectively)

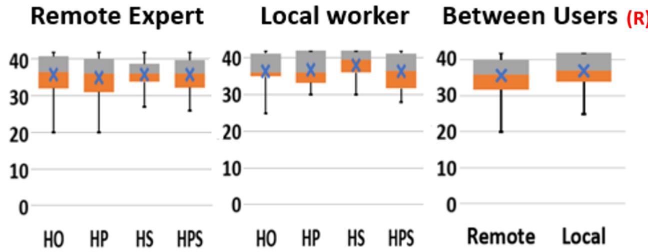

Figure 9: The co-presence rating results for each condition (x: mean; 0: lowest ~ 42: highest co-presence; S, P, and R: significant effect of the sketch, pointer, and participant's role, respectively)

adding the pointer cue did not. While this finding was common between the Tangram and Origami tasks, with the Lego task, there was no significant effect by the sketch cue.

Co-Presence

Figure 9 shows participants' ratings on the level of co-presence they felt for each condition. Adding sketch or pointer cues to the hand gesture cue did not show any significant effect on the remote experts' feeling of co-presence (sketch: $F(1,15)=.270$, $p=.610$; pointer: $F(1,15)=.225$, $p=.641$), and there was no significant interaction between the two cues ($F(1,15)=.625$, $p=.440$). The local worker's ratings also did not show any significant effect (sketch: $F(1,15)=2.908$, $p=.106$; pointer: $F(1,15)=1.717$, $p=.208$) nor significant interaction between the two cues ($F(1,15)=2.230$, $p=.154$)

We compared the ratings between the participants' role of remote expert and local worker with a Wilcoxon Signed Rank Test (see right of the Figure 9), and found that the participants felt higher level of co-presence in the local worker role than in the remote expert role ($Z=-2.396$, $p=.017$, $r=.299$).

Mental Effort

Figure 10 shows the results of the level of mental effort participants felt in each condition.

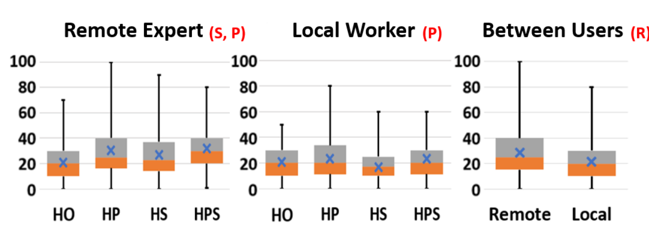

Figure 10: Results of mental effort point (0: lowest ~ 150: highest mental effort) for each condition (x: mean; S, P, and R: significant effect of the sketch, pointer, and participant's role, respectively)

Adding the pointer cue to the hand gesture cue significantly increased both the remote expert and local worker's feeling of required mental effort (remote expert: $F(1,15)$=16.617, p<.001, η_p^2=.603; local worker: $F(1,15)$= 11.267, p=.001, η_p^2=.437). Adding the sketch cue also increased the remote expert's feeling of required mental effort but not for the local worker (remote expert: $F(1,15)$=11.941, p=.001, η_p^2=.447; local worker: $F(1,15)$=.810, p=.372). There was no significant interaction between the cues for either roles (remote expert: $F(1,15)$=1.242, p=.270; local worker: $F(1,15)$= 2.979, p=.090).

We compared the scores between the participant's role of remote expert and local worker using a Wilcoxon Signed Rank Test and found that the participants felt significantly more mental effort in the remote expert role than in the local worker role (Z=-4.714, p<.001, r=.589).

Frequency of Using Cues

We analyzed the log data to compare frequency of using each visual cue, and we found that even when the sketch cue was available in HS condition, the hand gesture cue was still dominant visual cue as shown in the table 1.

Table 1: Average frequency of using hand gesture and sketch cue in the HS condition

	Mean (SD)		
	Lego	Tangram	Origami
Hand Gesture	18.6 (5.1)	18.3 (5.01)	10.3 (3.59)
Sketch	1.9 (1.18)	5.3 (2.24)	4.8 (1.87)

We compared the frequency of using hand gesture and sketch cues with a paired t-tests, and found that participants used the hand gesture cue significantly more than the sketch cue (Lego: $t(15)$ = 13.467, p <.001, d=3.18; Tangram: $t(15)$ = 8.432, p <.001, d=2.23; Origami: $t(15)$ = 5.307, p <.001, d=1.68) in HS condition.

Preference

After completing all four conditions, participants ranked them according to their preference from 1 (best) to 4 (worst). Table 2 shows the average rank.

Majority of remote experts and local workers chose the HPS conditions as the best (6 and 7 respectively) and the HO and HP conditions as the worst (7 and 7). Majority of the remote experts chose the HS condition as the second best (6) and the HO and HP conditions as the third best (5). Majority of the local workers chose the HPS condition as the second best (6) and the HO condition as the third best (6).

We analyzed the ranking results by a two-way repeated measures ANOVA (α=.05) with Aligned Rank Transform (ART) [48]. Results showed that the pointer cue added to the hand gesture cue did not produce a significant effect on the remote expert's nor local worker's preference (remote expert: $F(1,15)$=0.310, p=.586, local worker: $F(1,15)$=0.212, p=.652). However, the sketch cue added to the hand gesture cue had a significant effect on their preference (remote expert: $F(1,15)$= 6.540, p=.022, η_p^2=.304; local worker: $F(1,15)$= 6.479, p=.022, η_p^2=.302). There was no significant interaction effect on both remote experts' and local workers' preference (remote expert: $F(1,15)$=0.001, p=.977, local worker: $F(1,15)$= 1.924, p=.520).

Table 2: The results of ranking based on participants' preference (1:best 4:worst)

	Mean (SD)			
	HO	HP	HS	HPS
Remote Expert	3.00 (1.15)	2.93 (1.06)	2.12 (1.02)	1.93 (0.93)
Local Worker	2.88 (1.02)	3.06 (1.06)	2.25 (1.13)	1.81 (0.91)

Observation and Interview Comments

During the experiment the researcher closely observed participants' behavior and found five interesting themes. First, the sketch cue was important for solving the issues of misunderstandings during collaboration. For example, the remote expert could correct the local worker's mistakes by drawing a shape of the Tangram piece at the right orientation and by drawing a folding line in the Origami task. However, in the Lego task it was difficult to correct mistakes by drawing sketches because the Lego blocks were too small to draw and sometimes required 3D sketch at certain depth (e.g. when describing the shape of the block whose right end is on top of another, but left end is in the air). Some of the comments from the participants about the sketch cue were"*sketch allowed drawings for accuracy and hand for general use*", and "*sketch is pretty useful for describing actions that was difficult by verbal words and could express more details*".

Second, the benefit of the additional pointer cue was not clear. The hand gesture cue could show the same information that the pointer cue could convey (e.g. selecting and positioning an object with a pointer). Participants said, "*The pointer cue was quite similar to the hand experience*" and "*The pointer could show position information, but the hand gesture could provide wider range of options to place a piece*".

Third, mapping pointer and sketch cues to hand poses reduced the use of pure hand gesture. With a certain hand pose triggering the pointer and sketch cues, the use of hand gesture was limited. This is especially true with the pointer cue because the hand pose for triggering a pointer cue was the same as the pointing hand gesture (see Figure 2a), so the local worker could be confused whether the remote expert is pointing at the tip of the index finger or at the position of the pointer. One of the related comments from participants was "*the pointer cue was confusing because it showed the index finger and pointer together*".

Fourth, the local workers sometimes had prediction of the next manipulation based on the characteristics and rules of the task. For example, assembling a Lego model usually start from the bottom to the top and the goal model was usually bilaterally symmetric (at least in our user study tasks). With these in mind, after assembling a Lego block on the left side, when the remote expert asked to hold the same size block as the previous one, the local worker could easily predict that the next block would be placed on the right side, symmetrically. In case of a Tangram task, a new piece was usually bordering with the piece placed in the previous step, so the local worker could predict that it would be placed next to the previous piece. With an Origami task, the next folding manipulation was sometimes symmetric with the previous one.

Fifth, the collaboration between the remote expert and the local worker was smooth, clear, and fast when the spatial information from the remote expert was aligned with the local worker's prediction. However, there were several occurrences when they conflicted with each other. In such cases, the local worker sometimes misunderstood the remote expert's messages and made mistakes.

5 DISCUSSION

In this section, we discuss the results and compare them to prior works. We also suggest a future design of the system.

Discussion on study results In the results, the additional sketch cue contributed to a faster task completion time and higher feeling of usability. Sketch cue was able to present spatial information that was difficult to express only using the hand gesture cue (such as showing the shape of a target object at the right position and orientation). However, the additional pointer cue did not provide as much benefit because the information that it presented could also be given by the accompanied hand gesture cue. Thus, our hypotheses **H1** and **H2** are supported by the results only for the sketch cue but not pointer.

The results of perceived mental effort showed that using additional pointer and sketch cues requiring higher mental effort, so the hypothesis **H4** is not supported. Using the additional pointer cue was confusing on where the remote worker is actually pointing (either at the tip of the index finger or at the pointer) and this required extra communication and collaboration, so it might have demanded more mental effort. This would be also true on the local worker's end as the local worker felt higher demand of mental effort with using the additional pointer cue. However, the additional sketch cue demanded higher mental effort only for the remote expert but not for the local worker. This might be because of the difficulty of sketching in the air and drawing sketches onto roughly reconstructed surfaces requiring more effort.

Another interesting result was on co-presence. Both remote experts and local workers did not feel that the additional pointer and sketch cues improved co-presence, so the hypothesis H3 is not supported. This is conflicting to the results of the previous works which showed increased co-presence when using a mouse pointer [16, 27] controlled by a remote expert to give instruction and also when using sketch cues [27]. We postulate that this conflict is due to the fact that their systems [16, 27] did not support the hand gesture cue. Since our system supported sharing hand gestures, the need for additional pointer and sketch cues were limited. Moreover, we note that showing virtual hands (part of user's body) would have had higher impact on sense of co-presence than showing artificial pointer and lines that may not have comparable amount of effect on co-presence.

We also note that the results on co-presence in our study might be short for comparing it to prior work on gaze pointer [19, 39]. The pointer cue in our study was for indicating a point of interest using hands which is voluntary but only available when use it. However, the gaze pointer is always available and reveals remote partner's gaze behavior, i.e. how he/she is looking around, in addition to indicating a point of interest. Therefore, results from prior work on gaze pointers may not be comparable to our result of using a simple pointing indicator controlled by user's hand gestures.

The effect of the sketch cue was different depending on the task it was used for. The sketch cue appeared to be more useful in the Tangram and Origami tasks than the Lego task, supporting the hypothesis **H5**. Since the size of the Lego block was too small to draw, the effect of the sketch cue was not prominent. We also postulate that the three-dimensional structure of Lego may have also limited the use of sketch, as well.

Our results are similar to those of Fussell et al. [10] which compared the benefits of using sketch and pointer cues compared to a video only condition. However, they did not investigate the combination of visual communication cues with a baseline hand gesture cue, but simply compared the sketch, pointer, and only video condition. Moreover, our paper investigates two more factors: co-presence and mental effort, not explored in Fussell's study.

Design Suggestions In designing a remote collaboration system with multiple visual communication cues, we simply used two hand poses to trigger additional pointer and sketch cues, rather than using a button press interaction. However, this may have reduced the effectiveness of using the hand gesture cue, as some hand poses were reserved for triggering other visual cues. Therefore, we suggest not to employ frequently used hand poses to trigger additional visual cues when using the hand gesture cue.

Generally, combining various functions into an interface might increase the complexity. For effective use, the user should know which function would be good for a certain task and how to use the function effectively. These capacities could be achieved by training, so we recommend that the users would better be trained well when using interfaces including several functions.

Limitations In remote collaboration study, there are two important topics: providing a better view to the remote expert [26, 45] and providing better visual communication cue(s) between the local and remote users [13, 27]. Our study did not investigate providing a better view but focused on providing better visual cues. Moreover, there are two typical types of remote collaboration: remote expert and mutual collaborations [26]. The type of remote collaboration in this study was the remote expert case, and the results might be different in mutual collaboration.

Accurate reconstruction of the real-world surface is important for better placement of virtual pointer and sketching cues. However, we found that with Meta2 the accuracy of the reconstruction reduces as more objects are reconstructed. This could lead to an increasing inconsistency in placing the pointer and sketches on the reconstructed surface. We minimized this by using a simple empty workspace in the user study, while this issue should be overcome by using better quality 3D scanning and future HMDs for practical use.

One of the main benefits of using a VR HMD is supporting immersive experience with a large FOV. However, our system had limited support for immersive experience with a VR HMD (FOVE) as the view in FOVE was dependent to the video captured by the AR HMD (Meta2) which supported only 90-degree FOV. Future systems could support a full 360-degree independent view for the remote expert so that the user could have a more immersive experience of remote collaboration.

Our study results mostly depend on the subjective user data rather than objective data except the task completion time and frequency of the using additional sketch and hand gesture cues in the HS condition. This could be solved in the future through measuring the user's microsaccades eye movement that indicates mental fatigue [5] and feeling of task difficulty [43] which could be helpful for measuring mental effort during collaboration.

6 CONCLUSIONS

In this paper, we compared four combinations of the visual cues (hand only, hand + pointer, hand + sketch, and hand + pointer + sketch) in the MR remote collaboration system with three task scenarios (assembling Lego, assembling a Tangram, and origami tasks). The results showed that by adding sketch cue to the hand gesture cue participants completed the task faster and felt a higher level of usability. On the other hand, adding pointer cue did not provide any significant benefit on the performance or usability. There was no significant effect of the two additional cues on the level of co-presence, and participants felt higher level of required mental effort when using the additional visual cues. With the result, we suggest not to use the frequently used hand poses (e.g. pointing pose with index finger) to trigger additional pointer or sketch cue.

In the future, we will extend our study by comparing independent and dependent views while using the three visual communication cues.

ACKNOWLEDGMENTS

This work was partly supported by Next-Generation Information Computing Development Program from the National Research Foundation of Korea (NRF) (Grant Number: NRF-2017M3C4A7066316).

REFERENCES

[1] Leila Alem, Franco Tecchia, and Weidong Huang. 2011. Remote teleassistance system for maintenance operators in mines *(11th Underground Coal Operators' Conference)*. University of Wollongong.

[2] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies* 4, 3 (May 2009), 114–123. http://dl.acm.org/citation.cfm?id=2835587.2835589

[3] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[4] Sicheng Chen, Miao Chen, Andreas Kunz, Asim Evren Yantaç, Mathias Bergmark, Anders Sundin, and Morten Fjeld. 2013. SEMarbeta: Mobile Sketch-gesture-video Remote Support for Car Drivers. In *Proceedings of the 4th Augmented Human International Conference (AH '13)*. ACM, New York, NY, USA, 69–76. https://doi.org/10.1145/2459236.2459249

[5] Leandro L Di Stasi, Michael B McCamy, Andrés Catena, Stephen L Macknik, José J Canas, and Susana Martinez-Conde. 2013. Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience* 38, 3 (2013), 2389–2398.

[6] Omid Fakourfar, Kevin Ta, Richard Tang, Scott Bateman, and Anthony Tang. 2016. Stabilized Annotations for Mobile Remote Assistance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1548–1560. https://doi.org/10.1145/2858036.2858171

[7] Fove. 2018. FOVE head mounted display website. https://www.getfove.com/. Accessed: 2018-12-28.

[8] HS Friedman. 1979. The concept of skill in nonverbal communication: Implications for understanding social interaction. *Skill in nonverbal communication* (1979), 2–27.

[9] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. 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 (CHI '03)*. ACM, New York, NY, USA, 513–520. https://doi.org/10.1145/642611.642701

[10] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *HumanâĂŞComputer Interaction* 19, 3 (2004), 273–309. https://doi.org/10.1207/s15327051hci1903_3

[11] Lei Gao, Huidong Bai, Gun Lee, and Mark Billinghurst. 2016. An Oriented Point-cloud View for MR Remote Collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications (SA '16)*. ACM, New York, NY, USA, Article 8, 4 pages. https://doi.org/10.1145/2999508.2999531

[12] Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. 2012. Integrating the Physical Environment into Mobile Remote Collaboration. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 241–250. https://doi.org/10.1145/2371574.2371610

[13] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized Annotations and Virtual Scene Navigation for Remote Collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 449–459. https://doi.org/10.1145/2642918.2647372

[14] Susan Goldin-Meadow. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences* 3, 11 (1999), 419 – 429. https://doi.org/10.1016/S1364-6613(99)01397-2

[15] Saul Greenberg and Carl Gutwin. 2016. Implications of We-Awareness to the Design of Distributed Groupware Tools. *Computer Supported Cooperative Work (CSCW)* 25, 4 (01 Oct 2016), 279–293. https://doi.org/10.1007/s10606-016-9244-y

[16] Kunal Gupta, Gun A. Lee, and Mark Billinghurst. 2016. Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 22, 11 (Nov 2016), 2413–2422. https://doi.org/10.1109/TVCG.2016.2593778

[17] Professor Chad Harms and Professor Frank Biocca. 2004. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence, Mariano Alcaniz and Beatriz Rey (Eds.).

[18] Joerg Hauber. 2008. *Understanding Remote Collaboration in Video Collaborative Virtual Environments*. University of Canterbury Press, Christchurch New Zealand. Doctoral dissertation from University of Canterbury.

[19] Keita Higuch, Ryo Yonetani, and Yoichi Sato. 2016. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5180–5190. https://doi.org/10.1145/2858036.2858438

[20] Weidong Huang and Leila Alem. 2013. Gesturing in the air: supporting full mobility in remote collaboration on physical tasks. *Journal of Universal Computer Science* 19, 8 (2013), 1158–1174.

[21] Weidong Huang, Leila Alem, Franco Tecchia, and Henry Been-Lirn Duh. 2018. Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration. *Journal on Multimodal User Interfaces* 12, 2 (01 Jun 2018), 77–89. https://doi.org/10.1007/s12193-017-0250-2

[22] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. 1993. Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments. *ACM Trans. Inf. Syst.* 11, 4 (Oct. 1993), 349–375. https://doi.org/10.1145/159764.159762

[23] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: Integrating First-person View with Out-of-body Vision Generation for Human-human Augmentation. In *Proceedings of the 5th Augmented Human International Conference (AH '14)*. ACM, New York, NY, USA, Article 46, 8 pages. https://doi.org/10.1145/2582051.2582097

[24] Hirokazu Kato and Mark Billinghurst. 1999. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. 85–94. https://doi.org/10.1109/IWAR.1999.803809

[25] Seungwon Kim, Mark Billinghurst, Chilwoo Lee, and Gun Lee. 2018. Using Freeze Frame and Visual Notifications in an Annotation Drawing Interface for Remote Collaboration. *KSII Transactions on Internet and Information Systems* 12, 12 (Dec 2018), 6034–6056. https://doi.org/10.3837/tiis.2018.12.023

[26] Seungwon Kim, Mark Billinghurst, and Gun Lee. 2018. The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration. *Computer Supported Cooperative Work (CSCW)* 27, 3 (01 Dec 2018), 569–607. https://doi.org/10.1007/s10606-018-9324-2

[27] Seungwon Kim, Gun Lee, Nobuchika Sakata, and Mark Billinghurst. 2014. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 83–92. https://doi.org/10.1109/ISMAR.2014.6948412

[28] Seungwon Kim, Gun A. Lee, Sangtae Ha, Nobuchika Sakata, and Mark Billinghurst. 2015. Automatically Freezing Live Video for Annotation During Remote Collaboration. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1669–1674. https://doi.org/10.1145/2702613.2732838

[29] Seungwon Kim, Gun A. Lee, and Nobuchika Sakata. 2013. Comparing pointing and drawing for remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1–6. https://doi.org/10.1109/ISMAR.2013.6671833

[30] Seungwon Kim, Gun A. Lee, Nobuchika Sakata, Andreas DÃijnser, Elina Vartiainen, and Mark Billinghurst. 2013. Study of augmented gesture communication cues and view sharing in remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 261–262. https://doi.org/10.1109/ISMAR.2013.6671795

[31] David Kirk, Andy Crabtree, and Tom Rodden. 2005. Ways of the Hands. In *ECSCW 2005*, Hans Gellersen, Kjeld Schmidt, Michel Beaudouin-Lafon, and Wendy Mackay (Eds.). Springer Netherlands, Dordrecht, 1–21.

[32] David Kirk, Tom Rodden, and Danaë Stanton Fraser. 2007. Turn It This Way: Grounding Collaborative Action with Remote Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1039–1048. https://doi.org/10.1145/1240624.1240782

[33] David Kirk and Danae Stanton Fraser. 2006. Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1191–1200. https://doi.org/10.1145/1124772.1124951

[34] Adam D. I. Kramer, Lui Min Oh, and Susan R. Fussell. 2006. Using Linguistic Features to Measure Presence in Computer-mediated Communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 913–916. https://doi.org/10.1145/1124772.1124907

[35] Hideaki Kuzuoka, Gen Ishimoda, Yushi Nishimura, Ryutaro Suzuki, and Kimio Kondo. 1995. Can the GestureCam be a Surrogate?. In *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work ECSCW'95*. Springer, 181–196.

[36] LeapMotion. 2014. Leap Motion Pinch Draw Demo. https://gallery.leapmotion.com/pinch-draw/. Accessed: 2018-12-28.

[37] LeapMotion. 2018. Leap Motion hand tracking website. https://www.leapmotion.com/. Accessed: 2018-12-28.

[38] Gun Lee, Theophilus Hua Lid Teo, Seungwon Kim, and Mark Billinghurst. 2018. A User Study on MR Remote Collaboration using Live 360 Video. In *Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018*. 153–164. https://doi.org/10.1109/ISMAR.2018.00051

[39] Gun A. Lee, Seungwon Kim, Youngho Lee, Arindam Dey, Thammathip Piumsomboon, Mitchell Norman, and Mark Billinghurst. 2017. Improving Collaboration in Augmented Video Conference using Mutually Shared Gaze. In *ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, Robert W. Lindeman, Gerd Bruder, and Daisuke Iwai (Eds.). The Eurographics Association. https://doi.org/10.2312/egve.20171359

[40] Meta2. 2018. Meta2 augmented reality head mounted display website. https://www.metavision.com/. Accessed: 2018-12-28.

[41] Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM, New York, NY, USA, 113–116. https://doi.org/10.1145/2677199.2680549

[42] Nobuchika Sakata, Takeshi Kurata, Takekazu Kato, Masakatsu Kourogi, and Hideaki Kuzuoka. 2003. WACL: Supporting telecommunications using wearable active camera with laser pointer. In *null*. IEEE, 53.

[43] Eva Siegenthaler, Francisco M Costela, Michael B McCamy, Leandro L Di Stasi, Jorge Otero-Millan, Andreas Sonderegger, Rudolf Groner, Stephen Macknik, and Susana Martinez-Conde. 2014. Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience* 39, 2 (2014), 287–294.

[44] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. 2013. BeThere: 3D Mobile Collaboration with Spatial Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 179–188. https://doi.org/10.1145/2470654.2470679

[45] M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar. 1987. WYSIWIS Revised: Early Experiences with Multiuser Interfaces. *ACM Trans. Inf. Syst.* 5, 2 (April 1987), 147–167. https://doi.org/10.1145/27636.28056

[46] Matthew Tait and Mark Billinghurst. 2015. The Effect of View Independence in a Collaborative AR System. *Computer Supported Cooperative Work (CSCW)* 24, 6 (01 Dec 2015), 563–589. https://doi.org/10.1007/s10606-015-9231-8

[47] John C. Tang and Scott L. Minneman. 1990. VideoDraw: A Video Interface for Collaborative Drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 313–320. https://doi.org/10.1145/97243.97302

[48] Theophilus Teo, Gun Lee, Mark Billinghurst, and Matt Adcock. 2018. Hand Gestures and Visual Annotation in Live 360 Panorama-based Mixed Reality Remote Collaboration. In *in Proceedings of the 30th Australian Computer-Human Interaction Conference (OzCHI'18)*. ACM, New York, NY, USA, 406–410.

[49] Frank Weichert, Daniel Bachmann, BartholomÃďus Rudak, and Denis Fisseler. 2013. Analysis of the Accuracy and Robustness of the Leap Motion Controller. *Sensors* 13, 5 (2013), 6380–6393. https://doi.org/10.3390/s130506380

[50] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[51] Ferdinand Rudolf Hendrikus Zijlstra. 1993. *Efficiency in work behavior. A design approach for modern tools*. Delft University of Technology, Delft, The Netherlands. Doctoral dissertation from Delft University of Technology.

[52] Ferdinand Rudolf Hendrikus Zijlstra and L van Doorn. 1985. The construction of a scale to measure subjective effort. *Delft, Netherlands* (1985), 43.