

The Combination of Visual Communication Cues in Mixed Reality Remote Collaboration

Seungwon Kim^{1*} Gun Lee¹ Weidong Huang² Mark Billinghurst¹

Received: date / Accepted: date

Abstract Many researchers have studied various visual communication cues (e.g. pointer, sketch, 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, within the two user studies when a dependent view and independent view are supported respectively. In the first user study with the dependent view, the results showed that the hand gesture cue was the main visual communication cue and adding sketch cue to the hand gesture cue helped participants complete the task faster. In the second study with the independent view, the results showed that the hand gesture had an issue of local worker understanding remote expert's hand gesture cue and the main visual communication cue was the pointer cue with fast completion time and high level of co-presence.

Keywords Mixed Reality · Remote Collaboration · communication cue · 3D Scene Reconstruction · hand gesture · sketch · pointer

1 Introduction

In this paper, we explore a Mixed Reality (MR) [1] remote collaboration (the local worker collaborates in Augmented Reality (AR) environment involving a real world task space with virtual communication cues, and the remote expert collaborates in a computer-generated 3D reconstruction of the task space which is similar with Virtual Reality (VR) environment) for completing spatial physical tasks such as fixing a car, maintaining

machinery, or assembling puzzles. Since the remote expert is not in the environment where the task is being performed, there should be a way to provide awareness of the task space. In previous studies, this had typically been done in two ways: using dependent and independent views. A dependent view is where a local worker controls the viewpoint and the remote expert has the same view as the local worker [2–6]. In comparison, an independent view allows the remote expert to look at the task space independently from local worker's viewpoint [7–9]. This paper describes two user studies in dependent and independent views.

Additionally, the remote expert is not in the task space, so it is necessary to virtually represent their ideas or instructions through visual communication cues [10]. The most studied visual communication cues are pointers [11–15], sketches [2, 14–16], and hand gestures [10, 3, 17]. The use of them can be influenced by the view types because the local worker sees the visual cues from the same (with a dependent view, see left of Figure 1) or different viewpoints (with an independent view, see right of Figure 1) from the remote expert's.

In this paper, we explore the use of different combinations of the three visual cues in two user studies when a dependent view or independent view is supported. The combinations of the visual cues are 1) hand only, 2) hand + pointer, 3) hand + sketch, and 4) hand + pointer + sketch. We conduct 2 x 2 factorial design studies 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).

Our studies achieved the following:

1. Implementation of a MR remote collaboration system supporting dependent and independent views when both local and remote collaborators were wearing a head mounted display (HMD) and using a

Address(es) of author(s) should be given

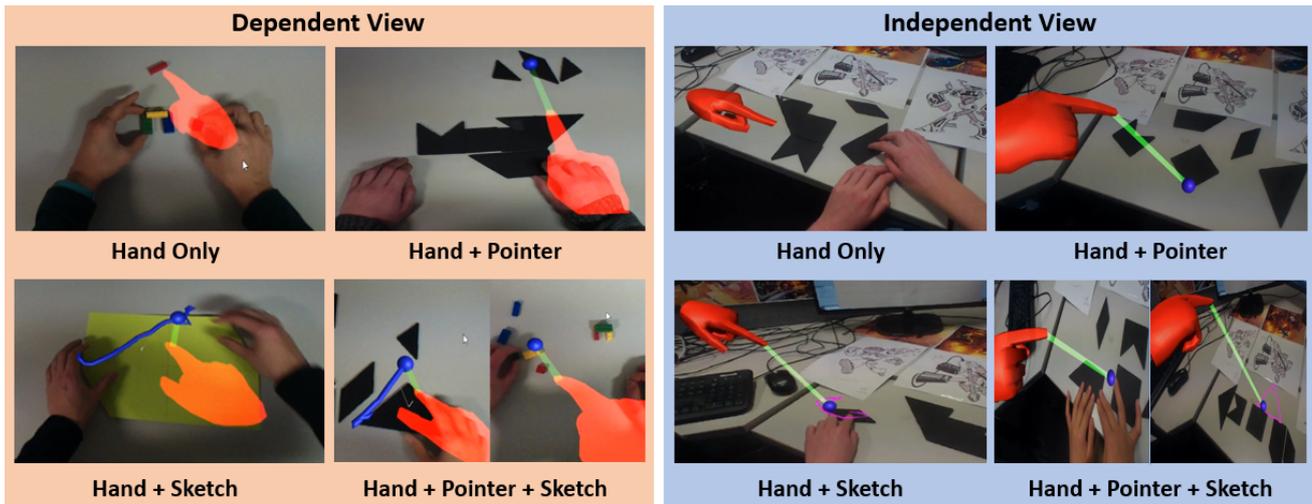


Fig. 1 The combinations of visual communication cues with dependent (left) and independent (right) views in the local worker’s perspective. The pointer is blue in both views, but the sketches are blue and red in the dependent and independent views respectively

combination of visual cues including the pointer, sketch, and virtual hand

2. Comparing the use of four combinations of the visual cues with dependent and independent views
3. Discussion on the different results according to the type of the shared views (dependent or independent view) and suggesting a better interface design

In the following sections, we review related work, describe two user studies including development of prototype interfaces, a user study design, and the study results, then wrap up with a discussion and conclusion.

2 Related Work

Prior remote collaboration studies had two main research topics: 1) sharing a view of the local task space with a remote expert and 2) using visual communication cues. We describe previous studies on both topics in this section.

2.1 Visual Communication Cues

In a remote collaboration study for a real-world physical task, a local worker can share his/her task activities through live video [3, 15] (i.e. showing object manipulation through a live video), but a remote expert cannot. Thus, researcher added visual communication cues for a remote expert to explain and help how the local worker need to manipulate the task objects. In this section, we described previous remote collaboration studies using visual communication cues such as pointers, sketches, and hand gestures.

2.1.1 Pointer Cue

Pointer cues have benefits of being easy, quick, and providing precise pointing interaction for positioning and selecting objects [2, 3]. Kim et al. [2] found that the benefit of easy and quick use encourages active collaboration and increases the level of co-presence between the local worker and remote expert. Gupta et al. [13] also found that both mouse and gaze pointers increased the level of co-presence. Lee et al. [4] and Higuchi et al. [18] found benefits of a gaze pointer in improving the level of co-presence and awareness of where collaborating partners were looking at.

2.1.2 Sketch Cue

While the pointer cue visually shows pointing information, sketch cues can show the shape of objects, manipulation direction, and orientation of an object, in addition to pointing information (e.g. by drawing a small circle or a tick shape) [2, 19]. Once drawn, the sketch remains in the shared task space, so the information is available until it is erased. With these benefits, prior studies found that participants completed assembly tasks faster with sketch cues than with pointer cues [14, 15].

Early systems often used a fixed camera view to avoid sketches losing the reference frame [14, 20] as simply rendering them at the same position of the fixed camera view accomplished the sketch stabilization. Later, some researchers attached a camera to the local worker’s head [6, 8, 13] or on a handheld device [5, 19]. In this case, the drawn sketches become misaligned to the real world when the camera view moves and shows a different part of the task space. Kato and Billinghurst [21], Kim et al.

[7,22], and Gauglitz et al. [9] solved this issue by using visual tracking systems that track camera motion and enable sketches to be drawn fixed in space.

In addition, using sketch cues increased the feeling of co-presence between the local worker and remote expert by showing the remote expert's messages in the shared view [2,3].

2.1.3 Hand Gesture Cue

Several researchers [18,17,23–26] implemented hand gesture cues for virtually representing the remote expert's hand gesture information in the shared task space. Gestures can express different types of information such as pointing, shapes, and even social cues (e.g. thumbs up) [10,17,27,28].

Hand gesture cues can help participants complete tasks faster [17,23,26]. One of the interesting findings is that participants sometimes showed required hand movement on top of a target object for manipulating it and it looked like the real operation compared to using pointer and sketch cues. Additionally, similar to pointer and sketch cues, using hand gesture cues increased the feeling of co-presence between collaborators [10,17].

2.1.4 Combination of Visual Cues

While most previous studies investigated interfaces supporting individual visual cues (by comparing between them or with and without a visual cue), some others [17,29,30,18] investigated the use of the combinations of visual cues. 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 (hand + sketch). Second, [17] and [30] shared 2D view and investigated the use 2D visual cues while our system supporting 3D visual cues with 3D view. Third, this paper includes two studies exploring the use of the combination of visual cues in two types of the shared views, dependent and independent views, but they studied the combination with one of view types.

2.2 Views for Remote Collaboration

2.2.1 Dependent View

The dependent view in remote collaboration systems means that the local worker controls the view and the remote expert had the same view, [7]. The dependent view has a benefit of WYSIWIS (What You See Is What I See) [22,31]. Since the local worker and remote expert have the same view, they do not worry about

the collaborating partner watching different areas of the task space and missing any shared activities such as a local worker manipulating an object, or a remote expert using and showing visual cues. However, this benefit have trade-offs [32]: the remote expert cannot have different perspective from local worker's.

2.2.2 Independent View

With the independent view, a remote expert can look around the task space regardless of where a local worker is looking at. Some researchers developed a hardware system in which a remote expert controls a camera in the local task space [11,12], so established the independent view.

Others implemented software functions providing an independent view by stitching the images from a live video, constructing a large image covering the overall task space [8,9,33], and supporting remote expert's freely navigating around the scene of the task space. some researchers used a 360-degree panoramic camera and allowed a remote expert to freely look around within the 360-degree independent view [34–36]. However, stitching images and using 360-degree live video for the independent view is still 2D view and limits the depth perception.

Sharing a 3D reconstruction of the local task space is another way of providing an independent view, and supporting depth perception [34,37,38]. Researchers studied the effect of the shared 3D independent view [34], awareness cues in the 3D [38], and the method of representing the remote user (such as avatar) in the 3D reconstructed environment [38]. However, there is no previous study investigating the use of combined visual communication cues within the 3D independent view.

2.3 Visual Cues in 3D Views

When using visual cues in the 3D views, the system mostly supports depth perception. The pointer and sketch cues in previous studies are mostly displayed on the surface of the target objects [2,7,9,16] and this makes them always be aligned with the target object regardless of the perspective difference between the local worker and remote expert (see pointer and sketches in Figure 1).

However, hand gesture cues are displayed in the air based on the hand tracking results [10,39]. If the remote expert performs hand gesture cue in the identical view (a dependent view) with the local worker, the local worker looks at the remote expert's hand gesture in the perspective that the remote expert has when performing the hand gesture (see Hand Only condition with the

dependent view in Figure 1), and the remote expert’s hand gesture is aligned with the target object in local worker’s perspective. However, if they have a different viewpoint with an independent view, the remote expert performs hand gesture in the perspective that is different with local worker’s, so the hand gestures is not align with the target object in the local worker’s perspective (see Hand Only condition with the independent view in Figure 1) and the local worker should guess where the hand gesture refers to.

In this paper, we investigate the use of visual cues in 3D views.

2.4 Hypotheses

From this related work, 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. Based on the findings, we suggest the following three hypotheses for each study with dependent or independent view:

- **H1** Using more visual communication cues in combination results in faster task completion time.
- **H2** Having more visual communication cues in combination leads users to have higher feeling of co-presence.
- **H3** Using more visual communication cues in combination reduces the mental effort for remote collaboration.
- **H4** The effect of using hand gesture cue is different according to the view types

3 Combinations in Dependent View

To investigate the use of four combinations of visual communication cues in the dependent view, we conducted the first user study. This section includes the description about the development of a MR remote collaboration system, user study design, and the results.

3.1 Prototype System

Before describing the system, we note that all software development was done with the Unity game engine (version 2017.3.0f3) for compatibility between the software components.

3.1.1 Hardware Devices

The prototype consists of two units: local worker and remote expert units (see Figure 2). Each is powered



Fig. 2 Devices at local (left) and remote (right) ends

by a computer (Intel Core i7-7700K 4.2GHz quad core CPU, 16 GB RAM, and NVIDIA GeForce GTX 1070 graphics card) and connected to each other by a local area network via Ethernet cable. All devices used in this study are connected to the computers on each end. The local worker uses a Meta2 head mounted display (HMD) [40], an Augmented Reality (AR) optical see-through display with 90-degree field of view (FOV) and 2550 x 1440 resolution at 60 frames per second (fps) (see right of Figure 2). The Meta2 supports 3D mesh reconstruction of the real-world in real time, so a user can place virtual objects on the surface of physical objects with the help of collision detection [2]. It also provides SLAM visual tracking [41], so it can render virtual objects at the proper position reflecting the user’s viewpoint movement.

The remote expert wears a FOVE Virtual Reality (VR) HMD [42] with a Leap Motion hand tracker [43] attached to the front (See left of Figure 2). Both of the devices are connected to another PC at remote expert’s end, 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 [44].

3.1.2 Dependent View and Visual Communication Cues

To develop a dependent view, we used a live video from Meta2’s RGB camera and tracking hand live frames from the Leap Motion. Since the Leap Motion is connected to the PC at remote expert’s end, the tracking hand frames are sent to local worker’s PC. The transferred hand frame is used to represent remote expert’s hand gesture as a live animated 3D virtual hands in the Meta2 for a local worker. This AR scene (virtual hand + live video from Meta2’s RGB camera) is transferred back to the remote expert side and displayed in FOVE for the remote expert. This dependent view with hand gesture is updated in real time at about 20 fps in 1280 x 720 resolution.

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 3). 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 by following previous studies [36,35]. 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 [2,9]. We used a virtual sphere to visualize the pointer and implemented the sketch function based on the Leap Motion’s ‘PinchDrawDemo’ sample code [45]. Based on prior work [14,6,19], 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 3).

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.

3.2 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

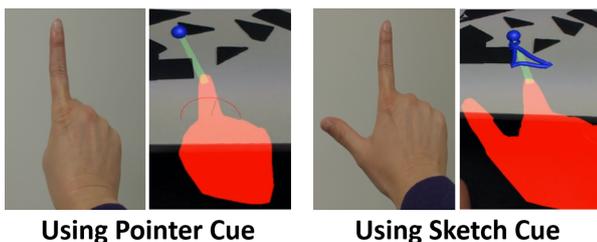


Fig. 3 Finger pose to place a pointer (left) and sketch (right)

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. The four conditions in this study are:

- **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 ‘Hand + Pointer’ at Figure 1).
- **Hands + Sketch (HS):** In addition to the hand gesture, the remote expert can sketch in the real-world view with a sketching gesture (see ‘Hand + Sketch’ at Figure 1).
- **Hands + Pointer + Sketch (HPS):** The remote expert can use all of visual cues including hand gesture, pointing, and sketch (see ‘Hand + Pointer + Sketch’ at Figure 1).

3.2.1 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: Lego, Tangram, and Origami (see Figure 4), and demonstrated how the system works. Then they tried a set of three sample tasks (solving three task in a row) while collaborating face-to-face. This face-to-face collaboration was to ensure that the participants understood the tasks before the experimental trials.

After practicing in face-to-face, the remote expert sat on a chair wearing the FOVE HMD and the local



Fig. 4 Tasks: Lego, Tangram, and Origami.

worker stood in front of another desk wearing the Meta2 HMD. 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 [3, 13].

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 4b). The Origami instruction showed red dotted lines with numbers indicating the folding order (see Figure 4c).

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 finishing all three tasks they answered to a SMEQ (Subjective Mental Effort Questionnaire) [46, 47] and a co-presence questionnaire [48] for overall experience. 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.

3.2.2 Task

Participants tried each condition with three tasks: assembling Lego, Tangram, and Origami (see Figure 4). 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. 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.

3.3 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 between being 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 that they had a 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.

3.4 Results

To analyze the results, we used two-way repeated measures ANOVA ($\alpha = .05$) 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. [49] (ART is a preprocessing step to transform and align nonparametric data before analyzing it using repeated measures ANOVA). The main results are summarized as:

- In general, participants preferred the conditions with the sketch cue among the four conditions.
- The sketch cue with the hand gesture cues significantly improved the performance (task completion time).

- Adding the pointer cue did not have any significant effect on the remote collaboration tasks.
- Adding the pointer and sketch cues did not increase the level of participant’s feeling in co-presence compared to the Hand Only condition.
- There was higher demand of mental effort when using the sketch and/or pointer cue(s) together with the hand gesture cue.

3.4.1 Preference

Participants ranked the four combinations according to their preference from 1 (best) to 4 (worst). Table 1 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 ($\alpha=.05$) 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: $F(1,15)=6.540$, $p=.022$; local worker: $F(1,15)=6.479$, $p=.022$). 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: $F(1,15)=1.924$, $p=.520$).

3.4.2 Task Completion Time

We analyzed the results of the task completion time (aggregating all three tasks) as summarized in Figure 5. Adding the sketch cue to the hand gesture cue enabled users to complete tasks significantly faster than without having it ($F(1,15)=8.452$, $p=.010$), 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, $SD=11.5$; without: $M=288.6$ seconds, $SD=11.1$). No significant interaction effect was found between the pointer and sketch cues ($F(1,15)=0.529$, $p=.477$).

3.4.3 Co-Presence

Figure 6 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 ($F(1,15)=2.230$, $p=.154$).

3.4.4 Mental Effort

Figure 6 shows the results of the level of mental effort participants felt in each condition. 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: $F(1,15)=16.617$, $p<.001$; local worker: $F(1,15)=11.267$, $p=.001$). Adding the sketch cue also increased the remote expert’s feeling of required mental effort but not for the local worker (remote expert: $F(1,15)=11.941$, $p=.001$; 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: $F(1,15)=2.979$, $p=.090$).

Table 1 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)

Task Completion Time (S*)

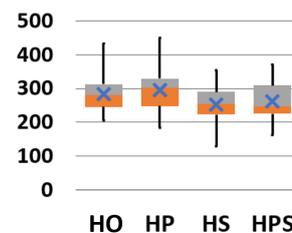


Fig. 5 Average of the task completion time in seconds (x: mean; S*, P*, and I* in red: significant effect of the additional sketch and pointer cues and interaction effect between them, respectively)

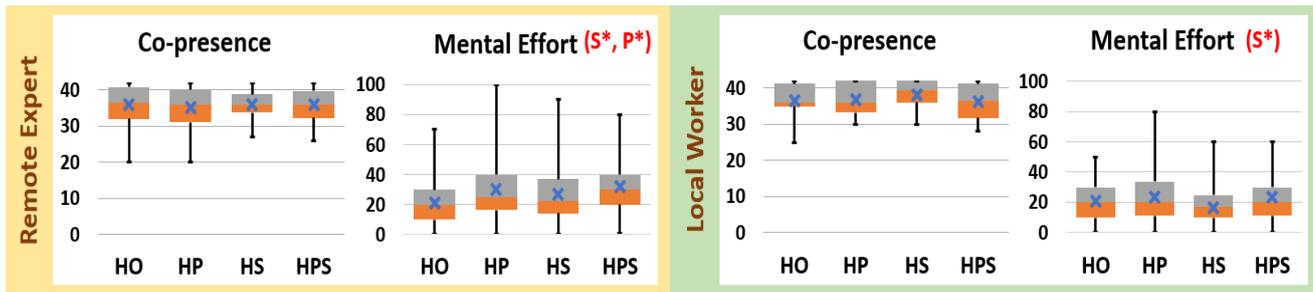


Fig. 6 Participant’s ratings in co-presence and mental effort (SMEQ) (X: mean; S*, P*, and I* in red: significant effect of the additional sketch and pointer cues and interaction effect between them, respectively)

3.4.5 Difference Between Local and Remote

We compared the participants’ ratings between the role of remote expert and local worker using a Wilcoxon Signed Rank Test (see Figure 7). The participants felt higher level of co-presence in the local worker role than in the remote expert role ($Z=-2.396$, $p=.017$, $r=.299$), and significantly more mental effort in the remote expert role than in the local worker role ($Z=-4.714$, $p<.001$, $r=.589$). This difference between the local and remote participants may come from that presenting information is more difficult than reading the information with our interfaces.

3.4.6 Observation and Interview Comments

During the experiment we found three interesting observation. First, the hand gesture cue was the main visual communication cues. According to observation, participants used the hand gesture cue more frequently than the pointer and sketch cues.

Second, 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*”.

Third, 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*”.

4 Combinations in Independent View

For better understanding the use of four combinations of visual communication cues, we conducted the second user study with the independent view. As a following study, we updated the prototype system from the one in the first study and made some changes in the user study accordingly. This section includes the description about the changes in the prototype system and the user study, then the results will be followed.

4.1 Prototype System

Same as the prototype system in the first study with a dependent view, all software development was done with the unity game engine (version 2017.3.0f3) for compatibility between the software components. The prototype consists of two units: local worker and remote expert units (see Figure 8a and 8f), and used the same computers, network between two units, Meta2 [40], and FOVE HMDs [42] from the first study.

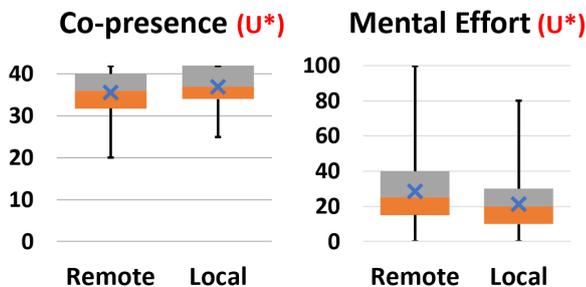


Fig. 7 Participants’ ratings according to their roles in co-presence and mental effort (SMEQ) (x: mean; Local: local participant; Remote: remote participant; U* in red: significant effect between the roles)

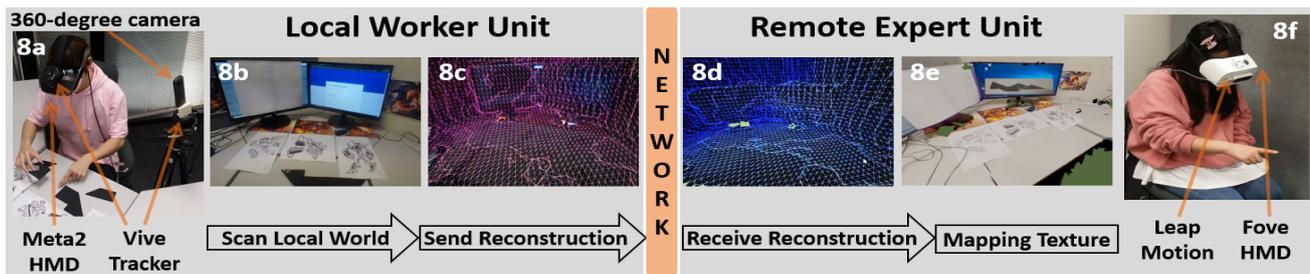


Fig. 8 Devices at local (8a) and remote (8f) ends. A local worker scans the task space (8b), the system automatically send and receive the reconstruction (8c and 8d), then maps the live video on the reconstruction (8e).

The main update is supporting independent view. To establish a 3D independent view, the local system reconstructs the local task space by using the Meta2 3D mesh reconstruction in the Meta2 Unity SDK 2.6 [50] (see Figure 8b). When completing the reconstruction, the system automatically transfers it to remote end system (see Figure 8c). Then the remote system displays it with the same Meta2 Unity SDK [50] which is integrated into the remote end system (see Figure 8d). One limitation of using the Meta2 3D mesh reconstruction is that it does not include image textures. To acquire live textures, we used a Richo Theta VR 360-degree video camera [51] (see Figure 8a). For correctly mapping the 360-degree live video onto the reconstructed mesh the system sets the projection origin as the position and orientation of the 360-degree camera in the shared 3D space, found from the Vive Pro tracker [52] (see Figure 8a). Since the Vive tracker coordinate frame is different from the one we use (Meta2 coordinate frame), we cannot directly use the position data. We attached another tracker on the Meta2 and then calculated the relative transformation between the two trackers in the Vive Pro coordinate frame and performed manual calibration to map the two coordinate systems. Since our system supports saving the Meta2 coordinate frame (i.e. the SLAM map) and the calibration information, this calibration does not require if it set once or the camera is moved. Unlike previous work [34, 53], this reconstruction can be performed even during the collaboration, so collaborators do not need to prepare the reconstruction before starting the collaboration. Additionally, our system also supports manually updating of the scene reconstruction during the collaboration when needed.

The FOVE HMD (see Figure 8f) tracks its movement with inertial measurement units (IMUs) consisting of tri-axis gyroscopes and accelerometers for tracking orientation, and an infrared camera for tracking position. With the tracking function, the view is updated according to the remote expert movement so the remote expert can freely navigate around the space and have an independent view.

4.2 User Study Design

For consistency with our first study, we prepared the same study conditions: Hand Only (HO), Hand + Pointer (HP), hand + sketch (HS), and hand + pointer + sketch (HPS). Finger poses to trigger the pointer and sketch cues were same with the one in the first study (see Figure 3), and the technology used for positioning them on the surface of the shared task space was also same (a ray casting and detecting collision between the ray and the surface). We also employed the same study measurement (task completion time, co-presence, required mental effort, and preference) with the same set of questionnaires (demographic, co-presence [48] and mental effort [46, 47], and preference questionnaires).

We employed the same user study procedure with the one in the first study. 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, then practiced in face-to-face collaboration. After the practicing in face-to-face, participants started testing the given conditions in counter balanced order using a balanced Latin-square design. For each condition, participants conducted a practice of using the interface given in the condition, getting acquainted with the task, performing collaborative task, and answering a questionnaire. After all four conditions, participant ranked the conditions according to their preference and answered for the questions asking the reason for the rank.

For the task, we used Tangram assembly used in our first study (see Figure 4b) but with different target models. We prepared nine Tangram puzzles for a face-to-face sample task and two sessions of four conditions ($1+2 \times 4 = 9$). Overall, the user study took about 90 minutes per pair of participants.

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

	Mean (SD)			
	HO	HP	HS	HPS
Remote Expert	3.0 (1.2)	2.0 (1.0)	3.0 (0.9)	1.9 (1.0)
Local Worker	3.2 (1.1)	2.0 (0.8)	2.8 (1.1)	2 (1.1)

4.3 Participant

We recruited 24 participants (in pairs) from our university staff and students and conducted 24 sessions of the user study by swapping their roles between a local worker and a remote expert. There were 19 males and 5 females with ages ranging from 22 to 62 ($M=30.2$, $SD=9.3$). They reported having moderate levels of experience in using AR/VR systems ($M=4.7$, $SD=2.0$), hand gesture interaction ($M=3.8$, $SD=1.9$), and video conferencing system ($M=4.3$, $SD=2.1$), by rating on a seven-point rating scale (1=Novice, 7=Expert). We note that eight participants participated in both the first and second studies, but the second study was conducted a year after the first study.

4.4 Results

To analyze the results, we used two-way repeated measures ANOVA ($\alpha=.05$). 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. [49].

The main results are summarized as:

- With the hand gesture cue, it was difficult to understand what it referred to in local worker’s perspective.
- The pointer cue led to significantly faster task completion, required significantly less mental effort in using it, and had significantly higher level of co-presence.
- The sketch cue did not have any significant benefit compared to the condition without it because of the difficulty in drawing from the side and at a distance

4.4.1 Preference

Participants ranked the four combinations according to their preference from 1 (best) to 4 (worst). Table 2 shows the average rank. We analyzed the ranking results by a two-way repeated measures ANOVA ($\alpha=.05$) with Aligned Rank Transform (ART).

Local and remote participants significantly preferred the conditions with the pointer cue than the conditions without it (see Table 2) (remote expert: $F(1,23)=11.725$, $p=.002$, local worker: $F(1,23)=27.407$, $p<.001$). The sketch cue did not give significant effect on their preference (remote expert: $F(1,23)=0.263$, $p=.613$, local worker: $F(1,23)=0.570$, $p=.458$). Ten (41.7%) remote participants chose the HPS condition as the most preferred and another nine (37.5%) chose the HP condition, while eleven (45.8%) local participants most preferred the HPS and six (25%) preferred the HP most. Only one remote participant chose the HS as the most preferred condition and five chose the HO condition. Among local participants, four preferred the HS condition most and another three preferred the HO condition most. 19 remote participants (79.2%) and 22 local participants (91.7%) least preferred the condition without the pointer cue (HO: 11 and 14 by remote and local participants; HS: 8 by each remote and local participants).

4.4.2 Task Completion Time

we analysed the task completion time (see Figure 9). Adding the pointer cue to the hand gesture cue helped participants to complete the tasks significantly faster than without having it ($F(1,23)=5.711$, $p=.025$), but adding the sketch cue to the hand gesture cue did not ($F(1,23)=1.169$, $p=.291$). On average, participants took about 19% less time to complete the tasks with the additional pointer cue (with: $M=87.917$ seconds, $SD=5.885$; without: $M=108.875$ seconds, $SD=9.241$). There was no interaction effect between the additional pointer and sketch cues ($F(1,23)=.118$, $p=.734$).

4.4.3 Co-Presence

Figure 10 shows participants’ ratings on the level of co-presence they felt for each condition. Participants felt higher co-presence with the condition using the pointer

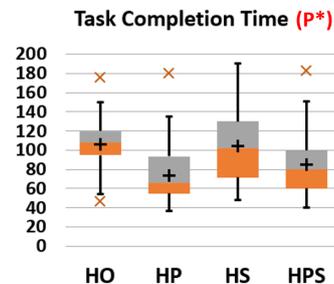


Fig. 9 Average of the task completion time in seconds (+: mean; x: outlier; S*, P*, and I* in red: significant effect of the additional sketch and pointer cues and interaction effect between them, respectively)

cue compared to the conditions without it (remote expert: $F(1,23)=5.855$, $p=.001$; local worker: $F(1,23)=25.804$, $p<.001$). However, adding sketch cue to the hand gesture did not show any significant effect on participants' feeling of co-presence (remote expert: $F(1,23)=.372$, $p=.548$; local worker: $F(1,23)=3.932$, $p=.059$).

The interaction between the effects of pointer and sketch cues was not revealed in remote expert's feeling in co-presence ($F(1,23)=1.683$, $p=.207$), but it was in local worker's ($F(1,23)=22.550$, $p<.001$). In the interaction of the local participants' co-presence, the average of the local participants' ratings was increased when providing the sketch cue (HO: 31.0 -> HS: 35.8) on top of the HO condition, but it was not with the HP condition (HP: 38 -> HPS: 38, HPS's interquartile range is inside of HP's). Since their third quartiles (41) are close to highest value (42), the reason of this interaction could be a ceiling effect.

4.4.4 Mental Effort

Participants also felt that they spent less mental effort when using the pointer cue than without it (remote expert: $F(1,23)=9.397$, $p=.005$; local worker: $F(1,23)=25.804$, $p<.001$). On average, remote and local participants gave 39.4% and 40.9% lower scores in the spent mental effort with the pointer cue than without it (With the pointer cue - Remote: $M=21.7$, $SD=14.7$; Local: $M=15.0$, $SD=8.8$; Without it - Remote: $M=35.9$, $SD=21.6$; Local: $M=25.4$, $SD=13.6$).

However, there was not significant effect on adding sketch cue to the hand gesture cue for both remote and local ends (remote expert: $F(1,23)=.193$, $p=.664$; local worker: $F(1,23)<0.001$, $p=.993$). There is no significant interaction between the effects of additional pointer and sketch cues (remote expert: $F(1,23)=1.290$, $p=.268$; local worker: $F(1,23)=0.008$, $p=.930$).

4.4.5 Difference Between Local and Remote

We compared the participants' subjective ratings according to their roles: local worker and remote expert, by using a Wilcoxon signed rank test ($\alpha=.05$). Figure 11 show the results.

Local participants had less mental effort in using the interfaces with the visual cues compared to the remote participants. The local participants also gave significantly higher points in co-presence compared to the remote participants. This difference between the local and remote participants may come from that presenting information is more difficult than reading the information with our interfaces.

4.4.6 Observation and Interview Comments

The participants' comments were aligned with our questionnaire results. The hand gesture cue was the baseline visual cue in the user study and available in all conditions, but more than half of the participants reported the difficulty of communication because the virtual hands were in the air and not aligned with the target in the local participant's perspective. Remote participants performed hand gestures from his/her perspective, but local participants saw the gesture from a different perspective,

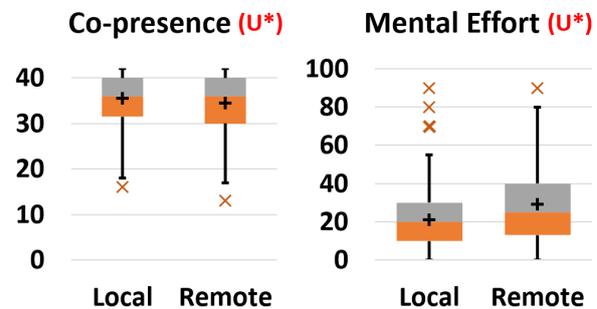


Fig. 11 Participants' ratings according to their roles in co-presence and mental effort (SMEQ) (+: mean; x: outlier; Local: local participant; Remote: remote participant; U* in red: significant effect between the roles)

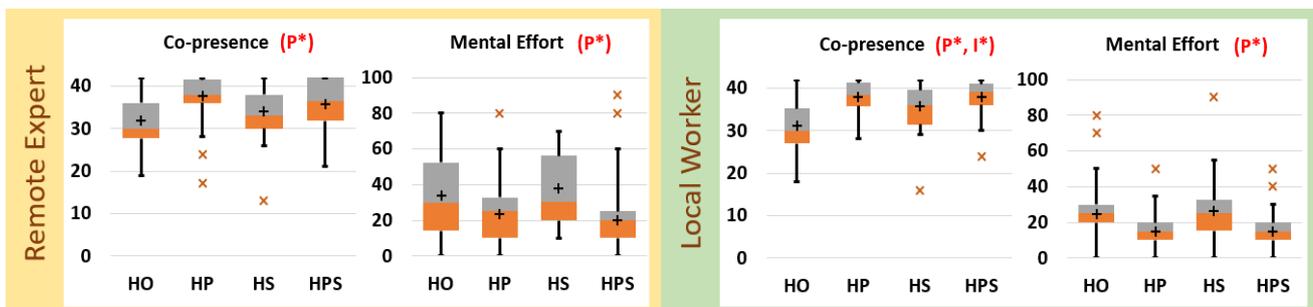


Fig. 10 Participant's ratings in co-presence and mental effort (SMEQ) (+: mean; x: outlier; S*, P*, and I* in red: significant effect of the additional sketch and pointer cues and interaction effect between them, respectively)

local participants' perspective ("*Hard to exactly point to unless the hand was on the shapes*", "*Partner's hand gesture didn't come closer than I expected*", "*The hand gestures was very difficult to tell where it was pointing to*").

According to participants' comments, the additional pointer cue was easy to control and required a low level of mental effort to use. Eight remote participants reported that it was easy to control ("*Pointer made it easier to position and select objects*" and "*easy to select the piece and move it to the right spot*") and five local participants commented that it was easy to understand ("*could see explicitly what piece or location they were pointing to*" and "*I can know which object I need to pick up*").

In our observation, the pointer cue was a powerful tool to represent the position and select an object by simply positioning the pointer. Interestingly, with the pointer cue, remote participants developed methods for sharing orientation of target objects. First, they used the shape of the objects and used verbal words (e.g. the local worker holds a right-angled triangle, then say "*the long side next to it*" when pointing an already assembled block). Second, they used the pointer to tell the edges of the objects ("*the edges come here and here*").

The sketch cue had an issue of being difficult to draw shapes. Some remote participants mentioned the difficulty ("*Drawing is difficult*" and "*It is a little difficult to draw lines*"), and some local participants sometimes could not understand the sketches ("*Drawings were confusing*").

5 General Observation

In this section, we described our observation from both studies.

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.

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.

In addition, the remote experts and the local workers showed different behaviour between the two views. We observed that when using the dependent view, both remote experts and the local workers did not need to perform any additional effort or activity to have better hand gesture communication. However, they did need to perform additional actions to overcome the limitations of the hand gesture cues with the independent view. For example, the remote experts stretched their arm as much as they could do or tried to move to the task object to make their hand gesture cue closer to the task objects. The local workers moved to see the hand gesture closer or have an aligned viewpoint with the remote expert.

6 Discussion

In this paper, we explored the use of four combinations of three visual cues: hand gesture, pointer, and sketch, with dependent and independent views through two user studies.

6.1 Contrasting Results between Two Studies

Hand Gesture The first and second studies explored the combination of visual cues, but the results were different. In the first study with a dependent view, the major visual communication cue was the hand gesture and it was easy to use. On the contrary, in the second study, the hand gesture cue had the issue of not being easy to understand where/what it referred to with the independent view. This difference in using the hand gesture cue was from how the local worker looked at the hand gesture cue. With dependent view in the first study, the local worker understood the remote expert's hand gesture in the identical view in which the remote expert made the hand gesture. So, the hand gesture appeared close and aligned with the target task object in the local worker's view. However, with the independent view in the second study, the viewpoint where a remote expert made a hand gesture was different with local worker's viewpoint, so the hand gesture was not aligned with the target object in the local worker's viewpoint and it was difficult to understand. Since the hand gesture cue was available in all conditions and had different

effects between the two views, the alternative visual cues, pointer and sketch cues, were used differently to supplement the hand gesture cue between the views.

Pointer As an alternative to the hand gesture cue, the pointer cue was not useful in the dependent view because the hand pointing gesture was easy to understand, but it was because the hand pointing gesture was difficult to understand. In a dependent view, the pointing hand by the remote expert was still in a line with the target object in the local worker’s perspective as they had the same perspective. So, the additional pointer cue was not essential for sharing pointing information. However, in the independent view, the hand pointing gesture was not in line with the target object in the local worker’s perspective and the pointer was displayed on the target object with a ray, so it was useful to share the pointing information.

Sketch The additional sketch cue was useful with a dependent view, but it was not with an independent view. This difference may come from different levels of difficulty in drawing the sketches with the dependent and independent views. The dependent view mostly provided close and front views near the task objects because the local worker controlling the dependent view should stay within an arm reachable area to manipulate the task object (in study [7], the average view distance when solving the Tangram puzzle was 38.69 cm). In an independent view allowing the remote expert to navigate around, the remote expert may have a more distant view than the dependent view and a side view when the local worker places the task object in front of him/her (this is especially true in our study because our design of the independent view was side by side remote collaboration and the initial position of the remote expert view was at a distance of about 70 cm).

Drawing sketches from the side and from a distance within an independent view could be more difficult than drawing them in front of the task space with a close dependent view. Drawing from the side has a smaller inner angle (the angle between the line from the index finger to the drawing point and the surface of reconstruction where the sketch was drawn on) compared to drawing from the front, and the same drawing activity results in a longer sketch if the inner angle is smaller. In this sense, the same incorrect (or wrong) drawing activity results in longer incorrect sketches with smaller inner angles when drawing from the side. Similarly, small mistakes in drawing can result in longer incorrect sketches having greater distance to the target object. Therefore, because of the distance and angle, drawing in an independent view could be more difficult to be accurate compared to the drawing in a dependent view.

Additionally, using the tip of the index finger may cause wrong sketches. A small change in finger or hand movement changed the direction of the ray which could cause a big movement of the drawing point (that results in drawing an incorrect big line).

6.2 Examine Hypotheses

As described previous section (5.1), the effect of using hand gesture cue in combination with pointer and sketch cues is different according to the view types. Therefore, **H4** is supported.

However, **H1**, **H2**, and **H3** were not confirmed from our studies. The three hypotheses were based on the results of previous studies showing the benefits of individual visual cue (more details in section 2.1), but the results in our studies were not sum of the benefits. Rather than the sum of the benefits, the harmony between the three visual communication cue and between the cues and view types were the main factors impacting our results. For example, the effect of additional pointer cue relied on the use of hand gesture cue. When using the hand gesture was easy and effective with the dependent view, the effect of the additional pointer was minor. When the other way around with the independent view, the effect was significant. The effect of the additional sketch cue, similar to the hand gesture cue, was influenced by the view types. Using the sketch cue was comparatively easier with the dependent view, so the effect of it was significant with the dependent view but was not with the independent view.

6.3 System Recommendation

We recommend a system supporting both dependent and independent views and allowing participants to switch between them. The independent view would be better for looking at the task space from a distance, so the remote expert can see a broad area and look around the task space. With a wide view and by looking around, the remote expert can have an overall understanding of the task space and find an object for next step. After finding an object, the remote user can use the pointer cue for attracting the local worker’s attention to the object. When the local worker comes close for manipulating the object, the view can be switched to a dependent view for easily using hand gesture and sketch cues for complex object manipulation. The issue is who or how to decide when to switch between two views (is it close enough to switch to the dependent view?).

7 Limitation

In the first study, participants solved the three tasks (Lego assembly, Tangram, and Origami) in sequence, and it was to cover wide range of tasks. However, we prepared only Tangram task in the second study. For better consistency, we could have prepared the same three tasks in the second study. We note that the results of the second study may not get affected by this, but by the difficulty in understanding hand gesture cue and drawing sketches.

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.

Our system can be installed and used at any place if there is no moving object when reconstructing the real world. One issue of our system is reduced portability from using the Vive tracker to get the position of the 360-degree camera in the Meta2 coordinate frame. This issue can be solved by tracking the 360-camera using the Meta2 library (e.g. integrating Vuforia tracking library with Meta2 library) or attaching it on the Meta2.

Our study results mostly depend on the subjective user data rather than objective data except the task completion time. This could be solved in the future through measuring the user's microsaccades eye movement that indicates mental fatigue [54] and feeling of task difficulty [55] which could be helpful for measuring mental effort during collaboration.

8 Conclusion and Future Work

In this paper, we studied the use of four combinations of the visual cues (hand only, hand + pointer, hand + sketch, and hand + pointer + sketch) in the MR remote collaboration system within two user studies when a dependent view and independent view is supported respectively. In the first user study with the dependent view, the results showed that the hand gesture cue was the main visual communication cue and adding sketch cue to the hand gesture cue helped participants complete the task faster. In the second study with the independent view, the results showed that the hand gesture had an issue of local worker understanding remote expert's hand gesture cue and the main visual communication cue was the pointer cue with fast completion time and

higher level of co-presence. The sketch cue in the independent view was difficult to use because of drawing sketches at a distance.

The reason that the hand gesture cue was not easy to understand with the independent view was because it was not aligned with the target object in the local worker's perspective. In the future, we will study how to address this issue. To solve the issue of the hand gesture cue, we recommend putting the virtual hands on the surface of the target object like the pointer and sketch cues. Further, to mitigate the difficulty of drawing sketches from the distance and side view, we can develop auto zooming function for having close and front view when drawing sketches.

Acknowledgements

References

1. P. Milgram, F. Kishino, *IEICE TRANSACTIONS on Information and Systems* **77**(12), 1321 (1994)
2. S. Kim, G. Lee, N. Sakata, M. Billinghurst, in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2014), pp. 83–92. DOI 10.1109/ISMAR.2014.6948412
3. S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, M. Billinghurst, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–13
4. G. Lee, S. Kim, Y. Lee, A. Dey, T. Piumsomboon, M. Norman, M. Billinghurst, in *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2017* (Institute of Electrical and Electronics Engineers Inc., 2017), pp. 79–80
5. S. Kim, G.A. Lee, N. Sakata, in *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2013), pp. 1–6. DOI 10.1109/ISMAR.2013.6671833
6. S. Kim, G.A. Lee, S. Ha, N. Sakata, M. Billinghurst, in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (ACM, New York, NY, USA, 2015), CHI EA '15, pp. 1669–1674. DOI 10.1145/2702613.2732838. URL <http://doi.acm.org/10.1145/2702613.2732838>
7. S. Kim, M. Billinghurst, G. Lee, *Computer Supported Cooperative Work (CSCW)* **27**(3), 569 (2018). DOI 10.1007/s10606-018-9324-2. URL <https://doi.org/10.1007/s10606-018-9324-2>
8. S. Kasahara, J. Rekimoto, in *Proceedings of the 5th Augmented Human International Conference* (ACM, New York, NY, USA, 2014), AH '14, pp. 46:1–46:8. DOI 10.1145/2582051.2582097. URL <http://doi.acm.org/10.1145/2582051.2582097>
9. S. Gauglitz, B. Nuernberger, M. Turk, T. Höllerer, in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (ACM, New York, NY, USA, 2014), UIST '14, pp. 449–459. DOI 10.1145/2642918.2647372. URL <http://doi.acm.org/10.1145/2642918.2647372>
10. W. Huang, L. Alem, F. Tecchia, H.B.L. Duh, *Journal on Multimodal User Interfaces* **12**(2), 77 (2018). DOI 10.1007/s12193-017-0250-2. URL <https://doi.org/10.1007/s12193-017-0250-2>

11. H. Kuzuoka, G. Ishimoda, Y. Nishimura, R. Suzuki, K. Kondo, in *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work EC-SCW'95* (Springer, 1995), pp. 181–196
12. N. Sakata, T. Kurata, T. Kato, M. Kourogi, H. Kuzuoka, in *ISWC*, vol. 2003 (Citeseer, 2003), vol. 2003, p. 7th
13. K. Gupta, G.A. Lee, M. Billinghurst, *IEEE Transactions on Visualization and Computer Graphics* **22**(11), 2413 (2016). DOI 10.1109/TVCG.2016.2593778
14. S.R. Fussell, L.D. Setlock, J. Yang, J. Ou, E. Mauer, A.D.I. Kramer, *Human-Computer Interaction* **19**(3), 273 (2004). DOI 10.1207/s15327051hci1903_3
15. S. Kim, G.A. Lee, N. Sakata, A. Dünser, E. Vartiainen, M. Billinghurst, in *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2013), pp. 261–262. DOI 10.1109/ISMAR.2013.6671795
16. S. Gauglitz, B. Nuernberger, M. Turk, T. Höllerer, in *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (2014), pp. 197–205
17. W. Huang, S. Kim, M. Billinghurst, L. Alem, *Journal of Visual Communication and Image Representation* **58**, 428 (2019)
18. K. Higuch, R. Yonetani, Y. Sato, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (ACM, New York, NY, USA, 2016), CHI '16, pp. 5180–5190. DOI 10.1145/2858036.2858438. URL <http://doi.acm.org/10.1145/2858036.2858438>
19. O. Fakourfar, K. Ta, R. Tang, S. Bateman, A. Tang, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (ACM, New York, NY, USA, 2016), CHI '16, pp. 1548–1560. DOI 10.1145/2858036.2858171. URL <http://doi.acm.org/10.1145/2858036.2858171>
20. J.C. Tang, S.L. Minneman, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, NY, USA, 1990), CHI '90, pp. 313–320. DOI 10.1145/97243.97302. URL <http://doi.acm.org/10.1145/97243.97302>
21. H. Kato, M. Billinghurst, in *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)* (1999), pp. 85–94. DOI 10.1109/IWAR.1999.803809
22. S. Kim, M. Billinghurst, C. Lee, G. Lee, *KSII Transactions on Internet and Information Systems* **12**(12), 6034 (2018). DOI 10.3837/tiis.2018.12.023
23. L. Alem, F. Tecchia, W. Huang, (University of Wollongong, 2011), 11th Underground Coal Operators' Conference
24. M. Adcock, S. Anderson, B. Thomas, in *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (2013), pp. 235–242
25. D. Kirk, D. Stanton Fraser, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, NY, USA, 2006), CHI '06, pp. 1191–1200. DOI 10.1145/1124772.1124951. URL <http://doi.acm.org/10.1145/1124772.1124951>
26. D. Kirk, T. Rodden, D.S. Fraser, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, NY, USA, 2007), CHI '07, pp. 1039–1048. DOI 10.1145/1240624.1240782. URL <http://doi.acm.org/10.1145/1240624.1240782>
27. H. Friedman, Skill in nonverbal communication pp. 2–27 (1979)
28. S. Goldin-Meadow, *Trends in Cognitive Sciences* **3**(11), 419 (1999). DOI [https://doi.org/10.1016/S1364-6613\(99\)01397-2](https://doi.org/10.1016/S1364-6613(99)01397-2)
29. T. Teo, G.A. Lee, M. Billinghurst, M. Adcock, in *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (2018), pp. 406–410
30. S. Chen, M. Chen, A. Kunz, A.E. Yantaç, M. Bergmark, A. Sundin, M. Fjeld, in *Proceedings of the 4th Augmented Human International Conference* (ACM, New York, NY, USA, 2013), AH '13, pp. 69–76. DOI 10.1145/2459236.2459249. URL <http://doi.acm.org/10.1145/2459236.2459249>
31. M. Stefik, D.G. Bobrow, G. Foster, S. Lanning, D. Tatar, *ACM Transactions on Information Systems (TOIS)* **5**(2), 147 (1987)
32. C. Gutwin, S. Greenberg, in *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (1998), pp. 207–216
33. J. Young, T. Langlotz, M. Cook, S. Mills, H. Regenbrecht, *IEEE transactions on visualization and computer graphics* **25**(5), 1908 (2019)
34. T. Teo, L. Lawrence, G.A. Lee, M. Billinghurst, M. Adcock, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–14
35. G. Lee, T.H.L. Teo, S. Kim, M. Billinghurst, in *Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018* (2018), pp. 153–164. DOI 10.1109/ISMAR.2018.00051
36. G.A. Lee, T. Teo, S. Kim, M. Billinghurst, in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications* (2017), pp. 1–4
37. T. Piumsomboon, G.A. Lee, J.D. Hart, B. Ens, R.W. Lindeman, B.H. Thomas, M. Billinghurst, in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, Montreal QC, Canada, 2018), CHI '18, pp. 1–13. DOI 10.1145/3173574.3173620. URL <https://doi.org/10.1145/3173574.3173620>
38. T. Piumsomboon, A. Day, B. Ens, Y. Lee, G.A. Lee, M. Billinghurst, in *SA '17* (2017). DOI 10.1145/3132787.3139200
39. L. Gao, H. Bai, T. Piumsomboon, G. Lee, R.W. Lindeman, M. Billinghurst, (2017)
40. Meta2 website. URL <https://www.metavision.com/>
41. H. Durrant-Whyte, T. Bailey, *IEEE robotics & automation magazine* **13**(2), 99 (2006)
42. Fove head mounted display website. URL <https://www.getfove.com/>
43. Leap motion hand tracking website. URL <https://www.leapmotion.com/>
44. F. Weichert, D. Bachmann, B. Rudak, D. Fisseler, *Sensors* **13**(5), 6380 (2013)
45. Leap motion pinch example. URL <https://gallery.leapmotion.com/pinch-draw/>
46. F.R.H. Zijlstra, *Efficiency in work behavior. A design approach for modern tools* (Delft University of Technology, Delft, The Netherlands, 1993). Doctoral dissertation from Delft University of Technology
47. F.R.H. Zijlstra, L.v. Doorn, Delft, Netherlands p. 43 (1985)
48. P.C. Harms, P.F. Biocca, (2004)
49. J.O. Wobbrock, L. Findlater, D. Gergle, J.J. Higgins, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, NY, USA, 2011), CHI '11, pp. 143–146. DOI 10.1145/1978942.1978963. URL <http://doi.acm.org/10.1145/1978942.1978963>
50. Meta2 unity. URL https://docs.metavision.com/external/doc/latest/unity_sdk_release_notes260.html
51. Ricoh theta v. URL <https://theta360.com/en/about/theta/v.html>

52. Htc vive. URL <https://www.vive.com/au/>
53. T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, M. Adcock, in *25th ACM Symposium on Virtual Reality Software and Technology* (2019), pp. 1–11
54. L.L. Di Stasi, M.B. McCamy, A. Catena, S.L. Macknik, J.J. Canas, S. Martinez-Conde, *European Journal of Neuroscience* **38**(3), 2389 (2013)
55. E. Siegenthaler, F.M. Costela, M.B. McCamy, L.L. Di Stasi, J. Otero-Millan, A. Sonderegger, R. Groner, S. Macknik, S. Martinez-Conde, *European Journal of Neuroscience* **39**(2), 287 (2014)