Comparing Pointing and Drawing for Remote Collaboration

Seungwon Kim
Human Interface Technology
Laboratory New Zealand
University of Canterbury
New Zealand
seungwon.kim@pg.canterbury.ac.nz

Elina Vartiainen
ABB Corporate Research
Västerås
Sweden
elina.vartiainen@se.abb.com

Gun A. Lee
Human Interface Technology
Laboratory New Zealand
University of Canterbury
New Zealand
gun.lee@hitlabnz.org

Mark Billinghurst
Human Interface Technology
Laboratory New Zealand
University of Canterbury
New Zealand
mark.billinghurst@hitlabnz.org

Nobuchika Sakata
Human Interface Laboratory
Division of Systems Science
and Applied Informatics
Osaka University, Japan
sakata@sys.es.osaka-u.ac.jp

Abstract

In this research, we explore using pointing and drawing in a remote collaboration system. Our application allows a local user with a tablet to communicate with a remote expert on a desktop computer. We compared performance in four conditions: (1) Pointers on Still Image, (2) Pointers on Live Video, (3) Annotation on Still Image, and (4) Annotation on Live Video. We found that using drawing annotations would require fewer inputs on an expert side, and would require less cognitive load on the local worker side. In a follow-on study we compared the conditions (2) and (4) using a more complicated task. We found that pointing input requires good verbal communication to be effective and that drawing annotations need to be erased after completing each step of a task.

Author Keywords

Video conferencing; Augmented Reality;

ACM Classification Keywords

H.4.3 [Information Systems Applications] Communications Applications - Computer conferencing, teleconferencing, and videoconferencing; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interface - Interaction styles

Introduction

Previous research has shown that gesture communication is important for remote collaboration [5]. Many collaboration tasks include physical objects, and collaborators may communicate with gestures about how to manipulate the objects. The manipulation work often involves changing the spatial state of the objects, and hand gestures are a very effective communication cue for conveying spatial concepts [3].

In our research we are interested in how Augmented Reality (AR) can be used to share gesture cues between remote collaborators. AR is good technology for sharing gesture cues because it provides spatial cues [2], and can convey spatial information with virtual objects. For example, a virtual object could represent a remote hand and the spatial information of the hand gesture could be conveyed by movement of the virtual object.

However there are some important research questions that need to be explored in order to build AR systems for displaying gestures. For example, is showing only a pointing gesture better than showing a drawing annotation? Is it better to show a still image of the remote workspace, or a live video view? In order to explore these and other questions we developed a handheld AR system for remote collaboration.

Related Work

There have been a number of earlier research projects that explored how gesture and annotation can be used to support remote collaboration. For example, Fussell et al. [4] studied the role of gestures in remote collaboration on a desktop interface, comparing them to systems with a video-only connection. They found that pointing gestures did not have any benefit over a video-only connection, but annotations led to significant improvements in user performance over video alone.

Kirk et al. [5] designed a remote gesturing system for comparing gesturing and no gesturing collaboration conditions. A video camera was used to capture images of the remote helper's hands and the gestures made were then projected onto the desk of the local worker. The worker followed the remote helper's gesture and speech instructions and the actions of the local worker was passed back to a monitor situated on the remote helper's desk. They found that using voice and gesture together the task completion time and errors were less than with voice alone.

Alem et al. [1] developed a mobile AR system for remote guiding between an expert and a local worker, where the local worker used a head mounted display (HMD) to see gesture cues provided by the expert. Both users were able to share the local environment through video streaming. The expert's hand gestures were captured on video which was transmitted back to the local worker where they are displayed in the worker's HMD. The helper's gestures were not only used to point at a specific point of an object on the display, but also to demonstrate how to perform a specific procedure. In a user study, the system was found to be quite intuitive and easy to use.

In contrast to these earlier works, our research is focused on collaboration between a local user using a handheld tablet and a remote expert with a traditional computer interface. We are investigating different gesture cues (e.g. pointing and annotation) with different view sharing methods (e.g. still images and live video). This is an important area of exploration as handheld displays are becoming more and more common, and are an ideal platform for remote

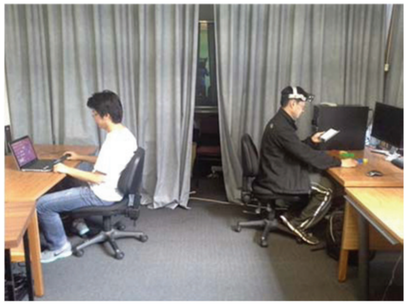

Figure 1. Experimental set up

collaboration.

Prototypes

To better understand how gesture cues can be used in remote collaboration, we compared pointers (representing a hand position) and drawing annotation (representing hand motion) in this study. In addition we were interested in whether shared images or live video of the remote task space would be more useful. Overall the study compared four conditions: (1) Pointers on Still Image (PS), (2) Pointers on Live Video (PV), (3) Annotation on Still Image (AS), and (4) Annotation on Live Video (AV).

We developed a pair of software applications to support remote collaboration; (1) An Android tablet application which uses touch screen interaction to allow a local worker to capture images or video of their local environment and stream it to the remote expert, (2) a PC application for a remote expert that allowed them to view the images sent from the tablet and point or annotate on them with mouse input. Figure 2 and 3 shows the tablet and PC interfaces.

In conditions (1) and (3), the tablet application transfers still images to the laptop after a picture is taken by the local worker, while in conditions (2) and (4), the tablet application streams live video to the laptop after turning on the system. When using pointers as in condition (1) and (2), each user controls a colored pointer (red for local worker and blue for expert) which they can move on top of the still image or live video. With the annotation interface in conditions (3) and (4), users can draw annotation on top of the still image or live video. Additionally, there are ‘Clear’ and ‘Erase’ functions for the annotation interface. The ‘Clear’ function erases any annotations that have been made with the button click interaction and the ‘erase’ function erases the part of annotation that was drawn at the part user touched.

Pilot Study 1: 2D task

In the first pilot study, we compared the four conditions with four pairs of participants. In the experimental environment (see figure 1), each pair sat back to back in the same room. They were able to talk to each other, share views through either live video or still images, and used pointing or annotation gesture cues to complete a block-arranging task (see figure 2 and 3). The laptop users (expert) were given a set of sequential pictures showing how to construct a block model. In this case the blocks were all arranged flat on the table surface so the task was a 2D object arranging task. The goal for the laptop user was to tell the tablet user how to complete the model. For each condition we recorded the performance in terms of task completion time and number of mistakes made, collected subjective measures through questionnaire, and took video recordings of the laptop and Android tablet screens.

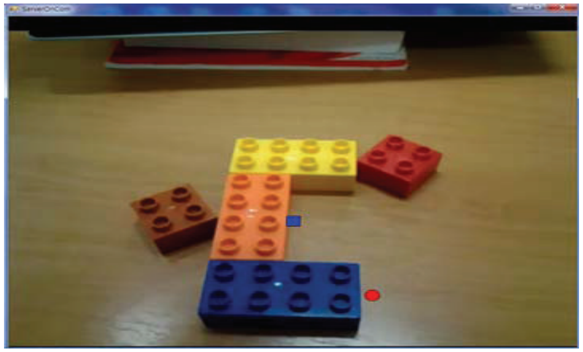

Figure 2. Pointer on a still image or a live video

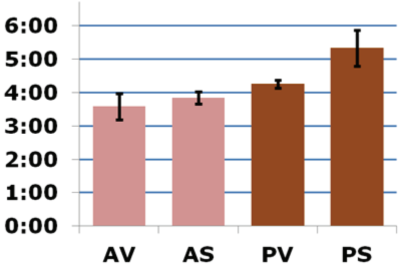

Figure 4. Average task completion time (in mm:ss)

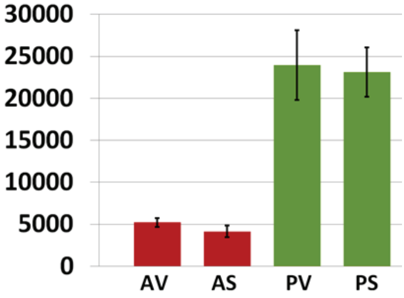

Figure 5. Average mouse cursor movement by the subject playing an expert role (in pixels)

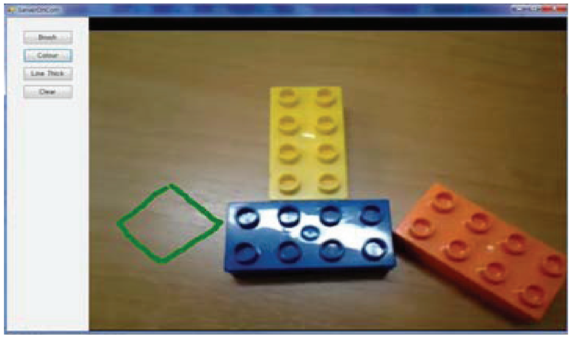

Figure 3. Annotation on a still image or a live video

We found that users were able to complete the task faster with annotation interfaces than with pointer interfaces (see figure 4). Although the number of participants was small, with a two-way repeated measure ANOVA, we found significant main effects for both the augmented gesture cues (annotation and pointer, $F(1,3)=54.7074$, $p=0.0051$) and the view sharing method (video and still image, $F(1,3)=26.0584$, $p=0.0145$). There was no significant interaction between the two factors. The interesting result was that when using annotation there was little difference in performance between live video and still images, while there were big difference between them when using pointer cues.

More mistakes were made when using pointer cues than annotation cues (number of errors in average PV: 5.25, PS: 4.25, AV: 2, AS: 1.75, $F(1,3)=14.8316$, $p=0.031$).

Figure 5 shows the average amount of mouse movement by the subject playing the role of the expert. The result show they had less input with annotation than with pointers ($F(1, 3)=33.3959$, $p=0.010$).

When asked to rank the preferred interface, participants preferred the annotation interfaces over the pointer interfaces, and video over still image (mean rank AV:1.5, AS=1.88, PV=2.88, PS=3.75, Friedman test $\chi^2(3)=14.850$, $p=0.002$).

Observations and Discussion

The results show that for the 2D task using drawing annotation helped users to perform better than when using pointing. This can be explained from annotation providing richer communication cues, such as drawing the 2D shape of a block at the desired position and orientation. In contrast, with pointer cues, the users had to draw the 2D shape inputs repeatedly for several times, in order to convey position and orientation information.

In the case of tablet users, the annotation interface required less cognitive load than the pointer interface. Since the drawn annotation remains on the screen, the tablet users were able to see the whole 2D shape of blocks. In contrast, pointer cues were only able to show a single point of interest, so the tablet user needed to figure out the shape of a block based on the pointer movement (which did not leave a visual trace) and remember the position and orientation information of the shape.

There was little difference in performance between using video or still images in the case of annotation gestures, but there was with pointer input. Moving the live video view while the expert was pointing or annotating on it meant that the virtual cues were no longer aligned with the real blocks. However with the annotations the expert was able to draw the key scene elements and so performance wasn't affected much.

Pilot Study 2: 3D task

From the previous study, we found that drawing annotations would be better than using pointer. However, when drawing on the live video, the local worker could be confused as new drawings and old drawings are drawn over each other. To explore this, we had another experiment with a 3D task that requires depth information for the object manipulation and leads to overlapping annotations.

The 3D task was also completing a block arrangement. However in this case the task was not only placing the blocks on the table, but also placing blocks on top of each other and rotating them about any axis (see Figure 6). For this pilot study, we had the same measurement and experiment environment as the first study. The only difference from the previous study was that we compared only two conditions, annotation on video (AV) and pointers on video (PV).

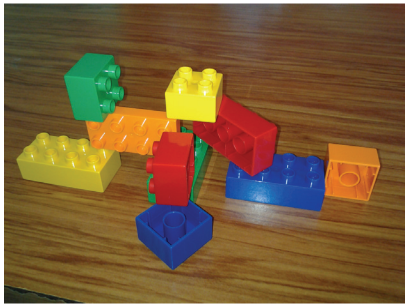

Figure 6. An example of 3D block arrangement task.

	AV(M)	PV(M)	Clear
Group 1	4:34(3)	6:10(9)	20
Group 2	4:58(4)	5:58(10)	18
Group 3	6:13(7)	4:02(2)	1
Group 4	4:55(5)	4:35(7)	8

Table 1. Average task completion time and the number of times the 'Clear' function was used (M – the number of mistakes the groups made with the interface).

In the experiment, we did not find a major difference on the user preference and task completion time between the two interfaces. Among the subjects playing the expert, two subjects preferred the annotation interface and the other two preferred the pointer interface. Among the subjects playing the local worker, three subjects preferred the annotation interface and one preferred the pointer interface. As shown in table 1, first two groups finished the task faster with the annotation interface, but the other two groups finished the task quicker with the pointer interface. All the mistakes the groups made were caused by erroneous orientation information conveyed between the two subjects. These results imply that we need a follow-on study with more participants.

Observations and Discussion

In case of the pointer interface, verbal communication was observed as the key factor for conveying the orientation information. We found that the words indicating the orientation of the blocks were used more frequently with the pointer interface than the annotation interface. This is due to the fact that with pointers it could be hard to represent the orientation information In case of group 1 and 2, the participants with the role of local worker had problems with understanding the words from their partners, time to time. Hence they made more mistakes, and needed more time to finish the task while using the pointer interface.

With the annotation interface, erasing the previous annotation could help reducing erroneous communication for orientation information (see Table 1). The two groups (group 1 and 2) that used the clear function more frequently had better performance with the annotation interface than the pointers interface, while the other two groups (group 3 and 4) completed the task faster with the pointer interface.

In the case of group 1 and 2, the participants with the role of expert usually drew a sketch for a block and cleared the annotation after their partner placed the block in correct orientation. The other two groups overlapped their drawings and needed more time to complete the task (see figure 7). Since the annotations were overlapped, the local worker was confused and found it difficult to figure out the proper position and orientation of the block from the new drawing.

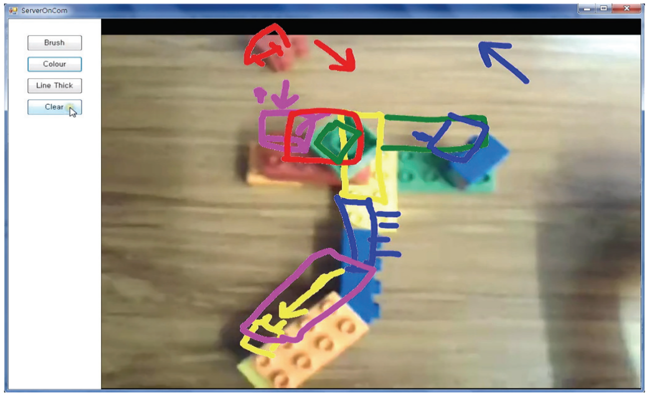

Figure 7. Stacked annotations with the 3D task.

While the annotation interface was very powerful in the 2D task, in the 3D task it was only beneficial when the subjects cleared the previous annotations. This is because annotation could not support multiple levels of depth with a 3D environment. Clearing the previous annotations would erase the confusing cues from the 3D task.

Conclusion and future work

In this paper we have presented work in progress exploring how pointing and drawing annotations with a handheld tablet can be useful to support remote collaboration. In the first pilot study, we found that allowing users to share drawing annotations provides better user performance than only showing pointers. The drawing annotation requires fewer inputs on the expert side, and less cognitive load on the local worker side to understand the communication cues. In second study, we found that the pointers interface could be completed quickly with clear verbal communication and enabling people to easily clear the shared drawings in the annotation interface was a key factor for improving performance.

In the future we will explore the benefit of providing richer gesture cues with a 3D virtual hand. Using a depth sensing camera (e.g. Kinect) the users' real hand could be captured and shared with the remote collaborator, allowing users to perform natural hand gestures. The pilot studies showed the annotation interface which has richer information performed better. Therefore, we expect that using a virtual hand could lead to better communication between the remote users.

References

[1] Alem, L., Tecchia, F. and Huang, W. Remote Tele-assistance System for Maintenance Operators in Mines. In *Proc*. COAL(2011), 171-177.

[2] Billinghurst, M. and Kato, H. Collaborative augmented reality., In *Proc*. CACM, (2002), 64-70

[3] Emmorey, K. and Casey, S. Gesture, Thought, and Spatial Language, In *Proc*. Gest (2001), 35-50.

[4] Fussell S, Setlock L, Yang J, Ou J, Mauer E, Kramer A. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. Human-Computer Interaction Inst, vol. 19, pp. 273-309, 2004

[5] Kirk, D. S., Rodden, T. and Fraser, S. D. Turn It This Way: Grounding Collaborative Action with Remote Gestures. In *Proc*. CHI (2007)1039-1048.