

# Study of Augmented Gesture Communication Cues and View Sharing in Remote Collaboration

Seungwon Kim    Gun A. Lee    Nobuchika Sakata    Andreas Dünser    Elina Vartiainen    Mark Billingham  
HIT Lab NZ    HIT Lab NZ    HIT Lab, Osaka University    CCI, CSIRO    ABB Corporate Research    HIT Lab NZ

## ABSTRACT

In this research, we explore how different types of augmented gesture communication cues can be used under different view sharing techniques in a remote collaboration system. In a pilot study, we compared four conditions: (1) Pointers on Still Image, (2) Pointers on Live Video, (3) Annotation on Still Image, and (4) Annotation on Live Video. Through this study, we found three results. First, users collaborate more efficiently using annotation cues than pointer cues for communicating object position and orientation information. Second, live video becomes more important when quick feedback is needed. Third, the type of gesture cue has more influence on performance and user preference than the type of view sharing method.

**Keywords** : Video Conferencing, Augmented Reality.

**Index Terms**: H.4.3 [Information Systems Applications]: Communications Applications - Computer conferencing, teleconferencing, and video conferencing; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities

## 1 Introduction

In face-to-face collaboration with real world tasks, a shared work space, gesture, and speech all combine to create a common understanding [3]. Teleconferencing tries to provide these components so that remote experts can figure out the current state of the task by sharing the local workspace and conveying instructions through speech and gesture communication cues. Most existing video conferencing systems (e.g. Skype) provide good support for verbal and visual communication cues, but limited ability for sharing space and gesture cues. In the past, Augmented Reality (AR) technology has been used to overcome this problem by overlaying virtual cues over a view of a remote workspace. This has been shown to be an effective way to create common understanding [4].

While many types of augmented visual cues have been proposed and investigated, to our best knowledge, there has been limited work to investigate the effect of using different augmented gesture cues combined with various view sharing methods for mobile remote collaboration systems. In this research, we investigate the effect of different gesture cues in two different view sharing techniques for remote collaboration.

## 2 View Sharing And Communication Cues

Fussell et al. [1] studied the role of gesture cues in remote collaboration, including pointing gestures that are used to refer to task objects and locations, and representational/annotation

gestures that are used to represent the form of objects and the actions to be performed with those objects. They conducted experiments comparing different types of gestures in a robot construction task in which the users shared their view through remote video stream from a fixed camera. While their work focused on comparing different gesture cues, they didn't explore the impact of different view sharing techniques. Their work only used a fixed camera for sharing views, compared to our work which focuses on exploring different view sharing techniques using a mobile device.

Gauglitz et al. [2] introduced an unobtrusive mobile remote collaboration system based on augmented world-stabilized annotations with model-free, marker-less visual tracking. They compared different visualization methods for annotations in a user study and showed that user performance can be improved by using world-stabilized visualization. However, the augmented visual communication cue used in the study was limited to placing a simple annotation (an 'x' mark with an identification number) at selected positions.

In our work we investigate different combinations of view sharing methods and augmented gesture cues, to find out how these communication components can complement each other.

## 3 Prototype Implementation

While we are interested in a wide range of view sharing methods and augmented gesture cues, this pilot study compares four different conditions: (PS) Pointers on Still Image, (PV) Pointers on Live Video, (AS) Annotation on Still Image, and (AV) Annotation on Live Video. We developed a prototype system for remote collaboration between a local worker with an Android tablet and a remote user (expert) using a laptop. The system uses touch screen interaction on the tablet and mouse interaction on the laptop. As view sharing methods, in conditions PS and AS, the tablet application transfers still images to the laptop after a picture is taken by the local worker, while in conditions PV and AV, the tablet application streams live video to the laptop. When using pointers as augmented gesture cues (conditions PS and PV), each user controls a colored pointer (red for the local worker; blue for the remote expert) which they can move on top of the still image or live video. With the annotation interface in conditions AS and AV, users can draw or erase annotations on top of the still image or live video. The pointer or drawing annotation is shared between the devices just as the workspace is shared.

## 4 Pilot User Study

In a pilot study, we compared the four conditions with four pairs of participants. Each pair used all four conditions. In the experimental environment (see Figure 1), the subjects were able to talk to (but not directly see) each other, shared view through either live video or still image, and used augmented gesture cues to accomplish a block arranging task. The laptop user (expert) had a set of sequential pictures showing how to construct a block model. The goal was for the laptop user to tell the tablet user how to

---

\* e-mail: seungwon.kim@pg.canterbury.ac.nz

complete the model. Both participants had the same interface except for the size of screen (see Figure 2). For each condition we recorded the performance as task completion time and number of errors, collected subjective data through a questionnaire, and took video recordings of the laptop and Android tablet screens.

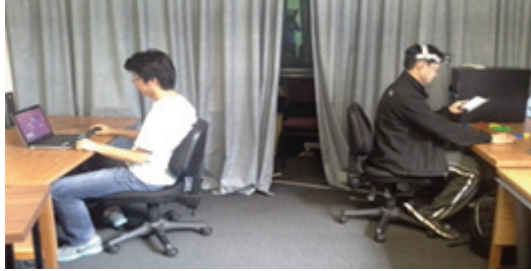


Figure 1: Experiment set up

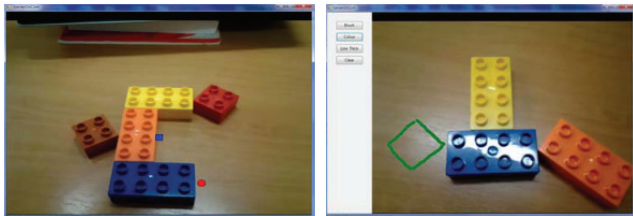


Figure 2: Pointers (Left) and annotation (Right) on live video

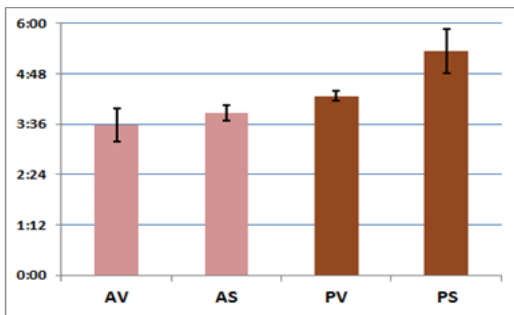


Figure 3: Average performance time

We found that users were able to complete the task faster with annotation interfaces than with pointer interfaces (see figure 3). More mistakes were made when using pointer cues (average, 12) than annotation cues (average, 3.75) and all subjects preferred the annotation interfaces over the pointer interfaces. When using annotation we found little difference in task completion time between live video and still images, while there were big differences in task completion time when using pointer cues. Analyzing the task completion time with a two-way repeated measure ANOVA, we found significant main effects for both the augmented gesture cues (annotation and pointer,  $F(1,3) = 54.71$ ,  $p > 0.01$ ) and the view sharing method (video and still image,  $F(1,3) = 26.06$ ,  $p = 0.01$ ). There was no significant interaction between the two factors ( $p = 0.42$ ).

## 5 Discussion

In our pilot study, annotation gestures provided richer communication cues for 2D tasks than pointing cues. With annotation, the process of communicating the position and orientation of a block was very simple. Since the drawn annotation remained on screen, the tablet users were able to see the whole 2D shape of blocks. In contrast, pointer cues only showed a single point of interest, so the tablet user needed to

figure out the shape of a block based on the movement of the pointer (which does not leave a visual trace) and remember the position and orientation of its shape. In short, a gesture cue which shows more information about how to manipulate objects, and which requires less cognitive load resulted in better performance.

We observed a difference in task completion time between the view sharing methods, and also a larger difference when using pointer cues compared to using annotation gestures. From the video recordings, we found that local users obtained feedback from the remote user immediately when they made mistakes with the live video interface. However, receiving feedback from the collaborator with a still image took several seconds. With users rarely making mistakes with annotation, slow feedback from the collaborator by sharing a still image affected the overall task performance much less.

The type of gesture cue appeared to be a more important factor for better performance and user preference than the type of view sharing method. Irrespective of the view sharing technique, the interfaces with annotation cues showed better performance and were more preferred than the interfaces with pointer cues. This also could be due to the task we have used in this study in which the laptop user only needed to check the before and after states rather than the whole progress. Since the tablet user knew how to pick up, hold, and put down the block, the laptop user did not need to lead the tablet user during this process. So sharing still images was good enough for showing the current task state to the laptop users. For dynamic tasks in which local users need help from remote experts all the time, sharing views through live video might become more important.

## 6 Conclusion & Future work

We presented a study exploring how different augmented gesture cues and view sharing techniques can impact a remote collaboration task between a tablet and laptop user. In our pilot experiment, we found that using annotation cues is more effective than using pointer cues because it can help in reducing task load for the local worker. Moreover, the live video can be beneficial when quick feedback is needed.

In the future we will study hand gesture cues with live video. As shown by our results, using annotations in a 2D task is better than using simple pointer cues. Since hand gestures can convey even richer communication cues, we expect that they should lead to even better collaboration performance. We plan to compare two types of hand gestures: 1) showing the hand movement on top of the shared view, 2) showing a trace of hand positions (similar to an annotation).

## REFERENCES

- [1] S. Fussell, L. Setlock, J. Yang, J. Ou, E. Mauer, and A. Kramer. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction Inst*, vol. 19, pp. 273-309, 2004.
- [2] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of MobileHCI*, pp. 241-250, 2012.
- [3] E. Hutchins. and P. Leysia. Constructing Meaning from Space, Gesture, and Speech. In *Discourse, Tools and Reasoning - Essays on Situated Cognition*, Ed. by Lauren Resnick, Roger Saljo, Clotilde Pontecorvo, and Barbara Burge, eds., Springer-Verlag, pp. 23-40, 1997.
- [4] J. Krimm and K. Nagao. The World through the Computer: Computer Augmented Interaction with Real World Environments. In *Proceedings of the 8th annual ACM symposium on User interface and software technology (UIST '95)*, pp. 29-36, 1995.