Supporting Hand Gestures in Mobile Remote Collaboration: A Usability Evaluation

Weidong Huang, Leila Alem
CSIRO ICT Centre
PO Box 76
Epping NSW 1710 Australia
{Tony.Huang, Leila.Alem}@csiro.au

Rapid advances in networking and hardware have made it possible for remotely located individuals to perform physical tasks together. Although a range of systems have been developed for remote collaboration, how to support the richness of hand gestures for an expert guiding a mobile worker located in a non-traditional-desktop environment has not been fully explored. HandsOnVideo is a system developed to fill this gap. The system uses a near-eye display to support mobility and unmediated representations of hands to support remote gestures. A usability evaluation has been conducted to gain in-depth understanding of the usefulness and usability of HandsOnVideo and the study yields positive results. In this paper, we describe the evaluation method, report the experimental results, discuss the findings and envision possible future improvements.

Remote guiding; mobile collaboration; hand gesture, usability evaluation; user interface.

1. INTRODUCTION

There are a range of real world situations in which remote expert guidance is required for a local novice to complete physical tasks. Typical examples of such situations include delivering healthcare services through telehealth: a specialist doctor guiding remotely a non-specialist doctor or nurse performing surgery on a patient. Or providing technical support for maintenance: a remote expert guiding a local technician into repairing a piece of equipment. The past decade has seen a fast growing interest among researchers and engineers in supporting remote collaborations between helper and worker (e.g., Alem and Li, 2011; Huang and Alem, 2011; Stevenson et al., 2008).

In general, it is often challenging to support interactions when collaborations take place over a distance, purely depending on computer mediated communications. This can be more difficult when remote collaboration occurs between a helper and a worker. Different interfaces and functions are needed to support the specific actions taken by the worker and the helper and to facilitate multi-modal communication and interactions between them (Fussell et al., 2004). Prior research has demonstrated that providing access to a shared visual space and supporting gesturing in the space are critical to the success of remote guiding (e.g., Kirk et al., 2007). Recently, a number of systems have been developed to achieve these using different technologies. For example, Ou et al. (2003) developed a DOVE system in which gestural sketches are integrated into the live video of the working environment and presented via a monitor in the local space. Sakata et al. (2003) developed a WACL system in which the worker wears a steerable camera/laser head and the helper can independently set his own viewpoint and point to real objects in the task space with the laser spot. Kuzuoka et al. (2004) developed GestureMan systems in which remote gestures are conveyed by a mobile robot through the use of a laser pointer. Kirk and Fraser (2005, 2006) presented a mixed ecology system. In this system, the helper's hands are captured by a video camera and the gestures he made are directly projected onto the desk of the worker to avoid fractured ecologies.

In comparing remote gesture technologies for supporting collaborative physical tasks, Kirk and Fraser (2005, 2006) conducted two studies to investigate the effects of different gesture formats on both immediate task performance and longer-term knowledge development (learning). It was found that using unmediated representations of hands significantly improved collaborative performance. However, the existing systems either assume that the workspace of the worker is confined in a fixed desktop setting, or support only limited gestures such as pointing. How to support the richness of hand gestures for an expert guiding a mobile worker located in a non-traditional-desktop environment has not been fully explored. In an effort to fill this gap, we have developed a system as part of our Human System Integration project within the CSIRO's Minerals Down Under, a National Research Flagship. The system is called HandsOnVideo, in which a near-eye display is used to support worker's mobility and unmediated representations of helper's hands are used to support remote gestures. For more details on technical specifications, see Alem et al. (2011).

In this paper, we describe a usability evaluation on HandsOnVideo. The study was conducted to gain in-depth understanding of the usefulness and usability of the system. In the next sections we give the overview of the system first. Then we describe the main experiment and a follow-up study. Finally we conclude the paper with a discussion, in which we discuss the results, limitations, implications and future work.

2. SYSTEM OVERVIEW

HandsOnVideo was developed with the following specific objectives in mind:

- Support richness of hand gestures of the helper.
- Support mobility of the worker.
- Support usability of the system in non-traditional-desktop environments in general, and in mining environments in particular.

This system includes two parts: a fixed helper station and a mobile worker end. The two sides are connected through a wireless network. We introduce them below.

2.1. Worker interface

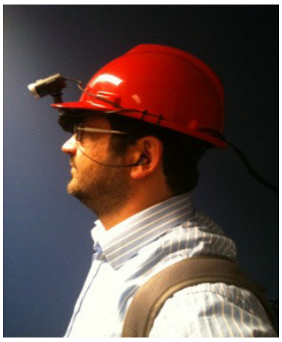

Figure 1: The worker interface

In mining sites, the environment is often noisy and risky with an unpredictable condition. Traditional stationary desktop setup on the worker side is no longer feasible. In addition, the worker is often required to walk around to fetch tools and inspect equipments during the task. We therefore made use of the helmet worn by mining workers by attaching a near-eye display (a small device with two screens) under the peak and a scene camera on top of it as shown in Figure 1. The worker can easily look up and see video instructions shown on the two small screens, and at the same time he can see the workspace around with little constraint.

2.2. Helper interface

The helper interface is a large touch-enabled display. The display has three main components: 1) a shared visual space that shows video streams captured by the scene camera on the worker side; 2) a panoramic view of the worker's workspace which the helper can use to maintain the situation awareness; and 3) four storage areas with two on each side of the shared visual space, which allows the helper to save and revisit specific scenes of the workspace.

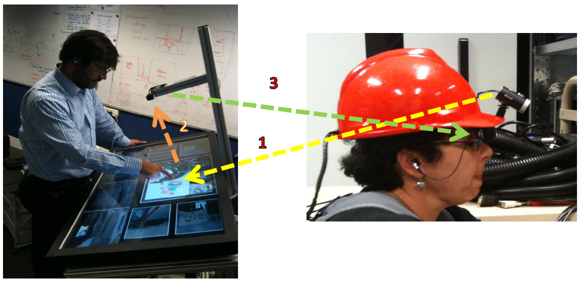

Figure 2: Capture and display of gestures.

As shown in Figure 2 (the left image), there is also a camera mounted on top of the display, which is to capture helper's hand movements within the area of the shared space.

2.3. How the system works

The worker interface is powered by a wearable laptop, while the helper side is powered by a high-end desktop computer. There is also an audio connection between both sides. How the whole system works can be seen from Figure 2. First, the videos of the scene camera on the worker side are sent to the helper side and displayed on the shared visual space (arrow 1). The helper performs gesturing over the shared visual space. The gestures together with background scenes are captured by the display camera (arrow 2), which are then compressed and streamed to the worker side and displayed on the near-eye display (arrow 3). During the task performance, the worker can see the hand gestures on the screens while hearing instructions from the helper.

3. METHOD

In this section, we present a user study we conducted with representative end users.

3.1. Design

The helper station of the system was located in a room, while the worker station was in a workshop room where the experimental environment was set similar to that of mining sites. Both rooms were about 20 meters away from each other. The helper and worker could talk to each other through a headphone. Users who had experience with remote collaboration were recruited to perform two physical tasks of different types: one representative task and one real world task. The participants were randomly grouped in pairs with one playing the role of helper and the other playing worker. Each pair had to perform the two tasks. After the first task, the two participants switched the roles for the second task. For each task, the whole process was video recorded on both helper and worker sides for further analysis.

There were also a questionnaire session after each task and a discussion session in the end for each pair. There were two questionnaires with one for helper and the other for worker. The two questionnaires included the same Likert style questions about ease of learning, ease of use, environment awareness of the work space, sense of co-presence, perceived task performance and interaction. Open questions specific to the role played in the task and associated interfaces were also included.

3.2. Participants

Six staff members volunteered to participate with the study. Two of them were workshop engineers maintaining equipment on daily bases. Another two were software engineers who had been working on remote collaboration projects for a number of years. And the rest of the participants were managers supervising maintenance and collaboration projects.

3.3. Tasks

Two types of tasks were used. One is the assembly task using Lego toy blocks. This task has been used in previous research for similar purposes (e.g., Fussell et al., 2004; Ou et al., 2003). This task is considered representative because it has a number of components that can be found in a range of real world physical tasks such as assemble, disassemble, select, move, attach and rotate (Kirk et al., 2005). During the task, the worker was asked to assemble the Lego toys into a pre-specified complex model under the instruction of the helper.

The other task is a repair task. This is a real task that may occur in mining sites. Since we did not have access to mining equipment, we used the repair of a PC as our second task instead. During this task, the worker was asked to take the cover of the PC off, replace one part inside the PC with another and put the cover back in place, under the guidance of the remote helper.

At the start of each task, the manual on how to construct the Lego model or how to fix the PC was provided to the helper. The helper was instructed that he could provide verbal and gestural instructions to the worker at any time, but not allowed to show any part of the manual to the worker. The worker, on the other hand, had no idea about what steps were needed to complete the tasks.

During the experiment, the toy blocks and the PC parts were placed in different locations of the workspace; the worker had to move around the workspace to collect them and get the task done. Also, there were also obstacles being deliberately placed between the locations; the worker had to avoid them while moving around. This was to test whether the worker was able to be aware of the environment while he walked with a near-eye display.

3.4. Procedure

The study was conducted in pairs. First two participants were gathered in the meeting room of the helper station. They were informed about the procedure of the study. The helper interface and the worker interface were introduced. They were also given chance to get familiar with the system and try out the equipment. During the introduction, the participants could ask questions and answers were provided by two experimenters.

When ready, the two participants were randomly assigned roles. Then they went to the corresponding rooms where the helper or worker station was located. On each site, there was also an experimenter providing further assistance to the participant, recording videos, observing and taking notes of the communication behavior.

The participants performed the Lego task first. After the first task, each participant was asked to fill the helper or worker questionnaire depending on his role. Then the participants switched roles, went to the corresponding rooms and proceeded to perform the second task: repair of a PC, followed by the questionnaires.

After finishing the two tasks and the two questionnaires, the participants went to the meeting room where they were debriefed about the purposes of the study first. Then a semi-structured interview followed. They were encouraged to ask questions, propose ideas and further improvements, debate on the issues and comment on the system. The whole session for the two tasks for each pair took about one hour.

3.5. Results

3.5.1. Observations

All pairs of the participants were able to complete their assigned tasks within reasonable periods of time. The main components of the helper interface: the shared visual space and the panoramic view were frequently used during the guidance. The helpers were able to perform a range of gestural actions over the shared space, such as those described by Kirk et al. (2005) and Fussell et al. (2004), while giving verbal instructions. It was also seen that the helpers were able to identify the locations of PC parts and toy blocks and guide their collaboration partners to the specific locations using the panoramic view of the workspace.

On the other hand, the workers were able to walk around the workspace without apparent difficulties. This demonstrates that the participants were able to be aware of the environment with the near-eye display. It can be seen that the workers were able to seek visual instructions while communicating with the helpers verbally. The communications between the collaboration partners seemed smooth and effective.

During the experiments, a few usability issues were also observed. We occasionally saw confusions on workers' faces when they looked at the near-eye display. Participants explained that the scenes showed on the near-eye display did not match what was mentioned by the helper. Later they realized that it was because the display was updated slowly. This was largely due to the network delay as complained by participants: “video lag is annoying”, “video delay makes it harder to use”, “video lag is not good”.

We also observed from three of the participants that there could be an issue of spatial awareness with the use of the near-eye display:

- One of the participants seemed to have difficulties locating a computer that was next to him. This participant used the near-eye display as his main source of information. He hardly used the natural and unmediated view of his workspace. This participant did not feel confident moving around his workspace.
- Another participant adjusted the display frequently. It is likely that he did not notice that he can switch between the views of the display and workspace simply by looking up the near-eye display, and without having to adjusting the display.
- One participant wore the near eye display very low and hence the focus of his attention was more on the instruction than on the task space. This also resulted in limited spatial awareness.

This issue seems to indicate that the near-eye display if not worn properly, may lead to a focus of the attention on the help provided, rather than conducting the task while referring to the help being displayed. A further exploration on how the near-eye display should be configured in the worker interface is needed in order to prevent such issues from happening again.

3.5.2. Questionnaire results

Six participants filled two questionnaires each: the helper questionnaire and the worker questionnaire. We had 12 responses in total. The detailed responses from the participants were presented as follows.

First, both helper and worker questionnaires included 6 usability questions to be answered in a Likert scale fashion, from 1: being strongly negative to 7: being strongly positive, with 4 being neutral. The overall rating statistics for the usability measures are shown in Table 1, while average ratings for helper and worker are illustrated in Figure 3:

	Mean	StDev
ease of learning	5.50	1.06
ease of use	5.25	1.21
task satisfaction	5.58	0.66
co-presence	4.33	1.15
awareness of environment	4.83	1.38
perception of interaction	5.25	0.96

Table 1: Means of overall ratings of the usability measures with standard deviations

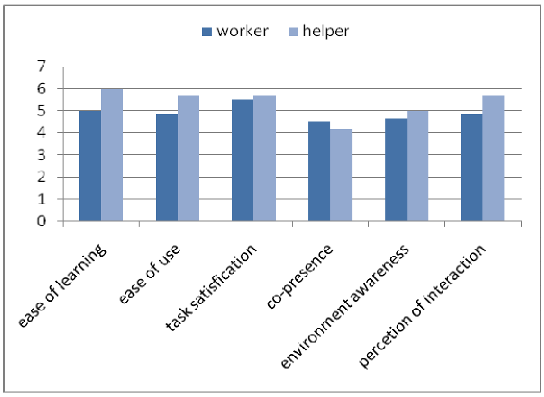

Figure 3: Mean ratings of the usability measures for worker and helper

As can be seen from Table 1, the participants generally thought that HandsOnVideo is relatively easy to learn and use. Co-presence and environmental awareness were rated just above being neutral, while perceived task performance and interaction were rated relatively high.

Specifically as shown in Figure 3, the participants perceived the system more useful when they played helper than when they played worker in terms of the usability measures except co-presence. Although t tests indicated that these differences were not statistically significant, the higher ratings with the helper role suggest that the participants were more comfortable with the helper interface and that on the other hand, they might need more time to get used to the worker interface.

In regard to co-presence, co-presence was rated relatively low compared to other measurements, it is reasonable since we did not expect that the system would present a sense of “being together” as strong as virtual environments would do. More specifically, as can be seen from Figure 3, co-presence was rated higher when participants played worker. This indicated that the worker had a greater sense of co-presence than the helper did. According to Kirk et al. (2005) and Alem and Li (2011), this difference was likely due to one of the key features that our HandsOnVideo system offered: the worker being able to see gesturing hands of the helper.

Second in regard to open questions, the participants were generally positive about the system. They appreciated being able to perform hand gestures and see the helper’s hands via the near-eye display. Examples of user comments include “the system should be useful in many situations”; “I looked at it (near-eye display) all the time”; “I find it very intuitive to use”; “It was kind of fun to receive instructions while walking around”; “Using hands was helpful when it was difficult to describe”; “Using hands made me feel like we were talking face-to-face”; “The near-eye display helped me to see what my partner could see”.

However, gains came at a price. The participants mentioned some features on the worker side might have negative impact on them. For example: “Video lag was annoying”; “The display blocked some of my field of view”; “Jerkiness of the images caused me a little bit of a headache”. “You had to look away from what your hands were doing to see what the instructor was doing”.

4. A FOLLOW-UP STUDY

The above experiment revealed that the participants generally considered the HandsOnVideo system useful and usable for remote guiding. However, based on the comments made by them, one issue that stood out was network latency. The latency caused a slow update on the near-eye display. While receiving verbal instructions from the helper, the worker had to wait to see what the helper meant.

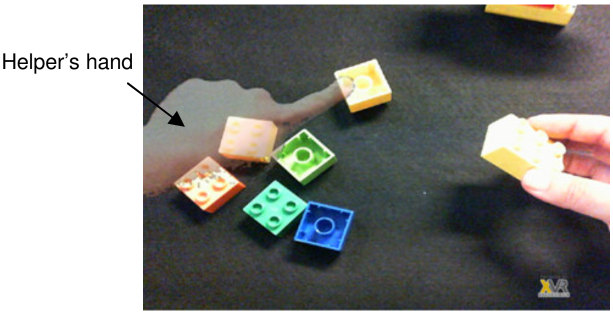

Figure 4: A local view augmented with helper’s hand

The network latency was mainly caused by the video data sent between two ends of the system. To reduce the system bandwidth requirement, we developed a hand-extraction algorithm with the help of an optical filter. The algorithm captures and extracts only the helper’s hands, without background information. Such hand gesturing is then compressed and streamed to the worker side and overlapped with the local copy of the scene videos (see Figure 4).

Two new participants were asked to evaluate the systems following the same protocol described in section 3. The user feedback indicated that the network latency was within the acceptable level (generally less than one second) and was no longer an issue. It was further commented that once the latency was noticed, the users were able to quickly adapt their own behaviour to cope with it during the task. The average ratings of the usability measures were generally higher than those in the main experiment.

5. DISCUSSION AND CONCLUSION

Our usability evaluation confirmed the usability and usefulness of HandsOnVideo for supporting real world scenarios in which a remote expert guides a mobile worker performing physical tasks in a non-traditional-desktop environment. The users were able to complete assigned tasks with quality and satisfaction in a reasonable time. The rating results of the usability measures indicated that users were generally positive with the system.

As far as we are aware, HandsOnVideo is the first system that uses the near-eye display to support mobile remote guiding. Although the display may partially block the local view, our usability study demonstrated that the use of it in such types of systems is promising. First it is small and light and requires little hardware and environment support. This is ideal for supporting mobility of the worker in non-traditional-desktop environments. Second, on the worker side, both the near-eye display and the scene camera are attached to the peak of a helmet. Therefore, the two devices move with the worker at the same time. This ensures that the view of camera and the view of worker are consistent. Although it is always desirable to make the two views the same, they are still different in the current setting. However, we were satisfied that none of our users had raised issues in relation to reference and orientation mapping, which is usually an issue when different viewpoints are used. This indicated that the view difference in the worker interface is small enough to avoid any noticeable negative consequences.

It is worth noting that although both HandsOnVideo and the mixed ecology system of Kirk and Fraser (2005, 2006) use unmediated hand representations for remote gestures, different approaches are used to represent hands in these two systems. The former combined the hands with the live video of the workspace and shown on the near-eye display, while the latter directly projects the helper hands into the workspace. At the first glance, presenting hand gestures on external monitors seems to require extra effort on shifting attention between workspace and monitor. This perception is also reflected in the comments of our users. However, prior empirical research has shown that the location of gesture output, no matter whether it is on an external monitor or it is on the surface of the workspace, does not make any significant differences in performance of collaborative physical tasks (Kirk and Fraser, 2006). In addition, in our system, effort on attention shift has been reduced to minimum: the near-eye display is located above the eyes of the worker; seeing hand gestures is just a matter of an eyelid lift.

Most of real world remote guiding scenarios do not happen daily. However, when remote expertise is required, it is often urgent. This is particularly true for telehealth and equipment maintenance. An urgent surgery that requires special expertise may be needed when a small medical team try to recover a patient in a rural and remote area. When a machine on a production line stops functioning, a quick fix is needed to avoid more serious consequences. Therefore it is important that such systems are lightweight, easy to set up and requires minimum training. However, our usability test indicated that for a wider application, HandsOnVideo needs more fine-tuning in this regard. For example, during the evaluation, we noticed that the storage area on the helper side was hardly used. Sometimes the helper performed gestures in a wrong area. All this indicated that the current helper interface was not intuitive enough. The heavy touch display also reduces the portability of the system when an expert is on the move. We are currently simplifying the helper interface and attempting to turn the whole system into a fully mobile and wearable one.

6. REFERENCES

Alem, L., Tecchia, F. and Huang, W. (2011) HandsOnVideo: Towards a gesture based mobile AR system for remote collaboration. In Alem, L. and Huang, W. (eds), Mobile collaborative augmented reality: Recent trends. 127-138. Springer, NY, USA.

Alem, L. and Li, J. (2011) A Study of Gestures in a Video-Mediated Collaborative Assembly Task. Advances in Human-Computer Interaction, Article ID 987830, 7 pages.

Fussell, S. R., et al. (2004) Gestures over video streams to support remote collaboration on physical tasks. Hum.-Comput. Interact., 19:273-309.

Huang, W. and Alem, L. (2011) HandsInAir: A Wearable System for Remote Collaboration. In Proc. 2nd Asia-Pacific Conference on Wearable Computing Systems, 171-174.

Kirk, D. S., Crabtree, A. and Rodden, T. (2005) Ways of the Hand. In Proc. 9th conference on European Conference on Computer Supported Cooperative Work, 1-21.

Kirk, D. S. and Fraser, S. D. (2005) The Impact of Remote Gesturing on Distance Instruction. In Proc. the International Conference on Computer Supported Collaborative Learning, 301-310.

Kirk, D. S., Rodden, T. and Fraser, S. D. (2007) Turn It This Way: Grounding Collaborative Action with Remote Gestures. In Proc. CHI Conference on Human Factors in Computing Systems, 1039-1048.

Kirk, D. S. and Fraser, S. D. (2006) Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. In Proc. CHI Conference on Human Factors in Computing Systems, 1191-1200.

Kuzuoka, H., et al. (2004) Mediating dual ecologies. In Proc. the ACM conference on Computer supported cooperative work, 477-486.

Sakata, N., Kurata, T., Kato, T., Kourogi, M. and Kuzuoka, H. (2003) WACL: supporting telecommunications using - wearable active camera with laser pointer. In Proc. 7th IEEE International Symposium on Wearable Computers, 53-56.

Stevenson, D., Li, J., Smith, J. and Hutchins, M. (2008) A Collaborative Guidance Case Study. In Proc. 9th Australasian User Interface Conference, 33-42.

Ou, J., Fussell, S. R., Chen, X., Setlock, L. D. and Yang, J. (2003) Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In Proc. 5th international conference on Multimodal interfaces, 242-249.