# Supporting the Changing Roles of Maintenance Operators in Mining: A Human Factors Perspective

Leila Alem[*,1], Weidong Huang[1] and Franco Tecchia[2]

[1]*CSIRO ICT Centre, Australia*

[2]*Scuola Superiore Sant'Anna, Italy*

**Abstract:** Rapid advances in networking and hardware have made it possible for remotely located individuals to perform physical tasks. Recently a number of systems have been developed with each supporting a different scenario of remote collaboration. Of particular interest to us is to explore the value of these technologies in the context of mining. As automation is being introduced in mines, more and more skilled operators are operating remotely. As a result, onsite operators' job becomes more complex and requires input from remote skilled operators. The productivity of the future mine relies on the effective delivery, just in time, of remote guidance. This paper presents a remote guiding system called HandsOnVideo, which is developed as part of our Human System Integration project within the CSIRO's Minerals Down Under (MDU), a National Research Flagship. Our aim is to design and develop a system that supports a remote helper guiding a mobile local worker in maintaining complex equipment in mine sites. The system is developed following a participatory design approach and the results of a usability study with real users indicate that the system is useful and effective.

**Keywords:** Remote collaboration, human factors, augmented reality, mobile computing, remote gestures, shared visual space.

## 1. INTRODUCTION

Globalization is an inevitable trend in the modern society and the need for collaboration between remotely located individuals has increased substantially. To meet this need, many systems have been proposed or developed in the literature (e.g., [1]). Technologies used to support remote collaboration include email exchanges, telephone calls, video conferencing and video-mediated gesturing. However, most of these systems in use are typically developed for supporting group activities or performing tasks without having to referring to or operating on external physical objects [2].

There are a range of real world situations in which remote expert guidance is required for a local novice to complete physical tasks. For example, in telemedicine a specialist doctor guiding remotely a non specialist doctor or nurse performing surgery for a patient [3]; in remote maintenance an expert guiding remotely a technician into repairing a piece of equipment [4]. Particularly in the field of the industrial and mineral extraction, complex technologies such as fully automated or semi automated equipments, teleoperated machines, are being introduced to improve productivity. Consequently, the maintenance and operation of these complex machines is becoming an issue. Operators/technicians rely on assistance from an expert (or more) in order to keep their machinery functioning. Personnel with such expertise, however, are not always physically located with the machine. Instead, they are often

in a major metropolitan city while the technicians maintaining equipment are in rural areas where industrial plants or mine sites may be located. Therefore, there is a growing interest in the use and development of technologies to support the collaboration between a maintenance worker and a remote expert.

It is often challenging to support interactions when collaborations take place over a distance, relying on computer mediated communication and interactions, and even more so to support collaborations between a remote helper and a local worker. For such remote collaborations, systems need to provide different interfaces and functions to support the specific actions taken by the worker and the helper and to facilitate communications and interactions between them. In general, one of the main issues when participants collaborate remotely is that there is no longer common ground for them to communicate in a way in which they do when they are co-located. Clark and Brennan [5] define common ground in communication as a state of mutual agreement among collaborative partners about what is referred to. In the scenario of an expert guiding a worker on physical tasks, the expert speaks to the worker by first bringing attention to the object that they are going to work on. To achieve this, the referential words such as "this", "that", along with gestures such as hand pointing, head nodding, eye contacts, and facial expressions, may be used [6]. Only when the mutual understanding is built can instructions on how to perform tasks be effectively communicated. As such, many attempts have been made to rebuild common ground. Among them, providing shared visual spaces is one that has been studied most. According to Tang *et al*. [7], "A shared visual workspace is one where participants can create, see, share and manipulate artefacts

*Address correspondence to this author at the Corner Vimiera and Pembroke Roads, Marsfield, NSW 2122, Australia; Tel: +61 2 9372 4366; Fax: +61 3 9545 8080;
E-mail: leila.alem@csiro.au

within a bounded space". Real world examples include whiteboards and tabletops.

When collaborating face-to-face on physical tasks, people use a range of visual cues available to them as a starting point and reference to communicate and interact with each other, therefore facilitating the negotiation of common ground during the process. The visual cues include facial expressions, body languages and actions of the partner, the view of the task objects and environments [2]. A series of studies have discussed and demonstrated that providing remote collaborators with access to these visual cues is beneficial to the completion of collaborative tasks. The visual cues are often provided in the form of video view of the workspace [8]. Further, prior research has indicated that the reason why face-to-face communication is more efficient than computer-mediated communication is mainly because in the face-to-face condition, participants are able to perform gesturing more easily and the gestures are visually available to all participants [6, 7, 9]. This suggests the importance of providing a shared visual space to the collaborators and supporting gesturing in the visual space.

A number of systems have been developed to support remote guidance providing a share visual space and using the space for gesturing. For example, Ou *et al.* [10] developed a DOVE (Drawing over Video Environment) system that integrates gestures of helper into the live video of the workspace to support collaboration on physical tasks. The main feature of the system is that the system allows a remote helper to draw on video streams of the workspace while providing task instructions. In their system, the gestures they support are mainly pointing and sketching. Kirk *et al.* [9, 11] presented a system MixedEcology. This system supports collaborative physical tasks through a mixed reality surface that aligns and integrates the ecologies of the local worker and the remote helper. In this system, the gestures of helper's hands are captured by a video camera and projected onto the desk of the worker. MixedEcology aims to promote mutual awareness between participants, by supporting the remote guiding through a mixed reality surface in which the remote helper's hands are overlaid on top of the local worker's hands. Kurata *et al.* [12, 13] developed a Wearable Active Camera with a Laser pointer system (WACL). In this system, the worker wears a steerable camera/laser head. The helper is allowed to control this steerable camera remotely, set his own viewpoint independently and point to real objects in the task space with the laser spot. Kuzuoka *et al.* [14] developed GestureMan systems in which remote gestures are conveyed by a mobile robot through the use of a laser pointer.

Despite the progress made for supporting remote collaboration, current systems either assume that the workspace of the worker is limited on a fixed desktop, or support only limited gestures such as pointing and sketching. In the mining environments, the traditional desktop workspaces are rare; the workspace conditions are usually dusty and unpredictable. Workers are often required to walk around to inspect the machine and fetch tools during the maintenance. In addition, complex hand gestures are needed to facilitate the communication process. 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 understood.

In an attempt to fill this gap, we have developed a system called HandsOnVideo following a participatory design approach. The key drivers for designing this system are:

1.    Support mobility of the worker

2.    Support the richness of hand gestures beyond pointing and digital sketching

3.    Easy to use with natural interaction

4.    Can be used in a mining environment

To be more specific, HandsOnVideo uses a near-eye display to support mobility and uses unmediated representations of hands to support remote gestures. A usability evaluation has been conducted and the results confirm the usability and usefulness of this system and also indicate possible improvements for future work.

This paper is an extension of its conference versions [15, 16]. The remainder is organized as follows. We first provide a review of the literature on systems for maintenance guidance. Then technical specifications and features of HandsOnVideo are described, followed by a usability study. Design tradeoffs and system limitations we have encountered are discussed. Finally we conclude the paper with a short summary and future work.

## 2. RELATED WORK

In this section, we selectively review related work. Systems for remote guidance are reviewed first. Then a review of previous research on supporting remote gestures is presented.

### 2.1. Systems for Remote Guidance

There are many real world collaborative scenarios in which the worker is engaged in a mobile task or performing tasks on objects that are consistently moving. The mobility of the worker presents unique challenges for system design and a few attempts have been made by researchers to address the challenges.

Kuzuoka *et al.* [14] developed a system GestureMan for supporting remote collaboration. GestureMan uses mobile robots as communication media (see Fig. **1**). The instructor controls the robot remotely and the operator receives instructions *via* the robot. In their system, the robot is mounted by a three-camera unit for the environment of the operator. It also has a laser pointer for hitting the intended position and a pointing stick for indicating the direction of the laser pointer. The movement of the robot is controlled by the instructor using a joystick.

Kurata *et al.* [12, 13] developed the WACL system that involves the worker wearing a steerable camera/laser head. WACL allows the remote instructor not only to independently look into the worker's task space, but also to point to real objects in the task space with the laser spot. As shown in Fig. (**2**, left), the laser pointer is attached to the active camera-head and it can point a laser spot. Therefore, the instructor can observe around the worker, independently of the worker's motion, and can clearly and naturally instruct the worker in tasks.
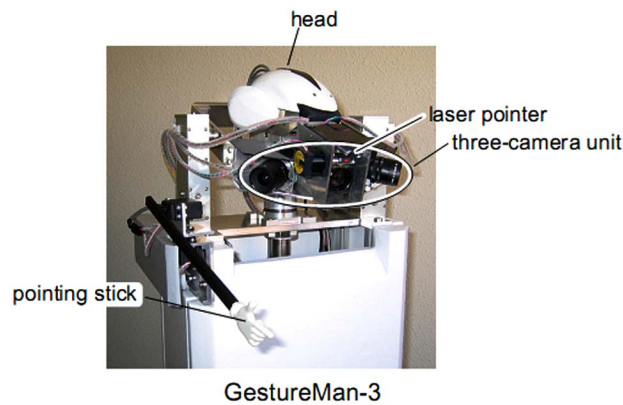
**Fig. (1).** Overview of GestureMan-3 system [14].

Previous work in the area of remote guiding of mobile workers has mostly focused on supporting pointing to remote objects, using a projection based approach such as the laser pointing system in WACL [12, 13], or using a see through based approach such as in REAL [17] (see Fig. **2**). While pointing (with a laser or a mouse) is an important aspect of guiding, research has indicated that projecting the hands of the helper supports a much richer set of non verbal communications and hence is more effective for remote guiding (e.g., [9, 11]). The next section reviews the work in this space.

## 2.2. Supporting Remote Gestures

Importance of gestures can be intuitively illustrated by hand movements that we use together with verbal and nonverbal communications in our everyday life. In fact, the use of hand gestures in support of verbal communications is so natural that they are even used in communications when people speak on the phone. Recent empirical studies have also shown that gestures play an important role in building common ground between participants in remote guiding [6].

Given that gesturing is of such importance to collaborative physical tasks a variety of systems are being developed to facilitate remote gesturing (e.g., [2, 5, 6, 9]). Most of these systems are explicitly built with the intention of enabling a remote helper (expert) to guide the actions of a local worker, allowing them to collaborate over the completion of physical tasks. Results have so far suggested that such tools can increase performance speed and also improve the worker's learning of how to use the system and perform tasks (when compared to standard video-mediated communication methods).

More specifically, Fussell *et al*. [6] introduced a system in which the helper can perform gestures over the video streams. In their system, gestures were instantiated as a digital form (see Fig. **3**). A user study conducted by Fussell *et al*. demonstrated the superiority of the digital sketches over cursor pointers. More recently, Kirk *et al*. [9] explored the use of gestures in collaborative physical tasks using augmented reality. In particular, the guiding is supported through a mixed reality surface that aligns and integrates the ecologies of the local worker and the remote helper (Fig. **4a**). The system allows the helper to see the objects in the worker's local ecology, the worker's actions on the objects in the task space, and his/her own gestures towards objects in the task space (Fig. **4b**). The work of both Fussell *et al*. and Kirk *et al*. demonstrated the importance of supporting remote gestures. However, how gestures can be better supported with a mobile worker has not been fully understood.

## 3. HANDSONVIDEO

Our literature review suggests the following requirements for remote guiding systems in industry.

- The need to support the mobility aspect of the task performed by the worker using wearable computers and wearable cameras.

- The need to allow helpers to guide remotely using their hands in order to provide reference to remote objects and places but also support procedural instructions.
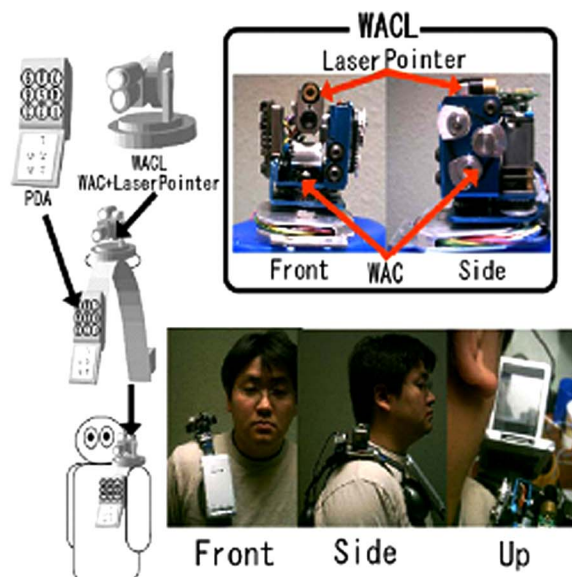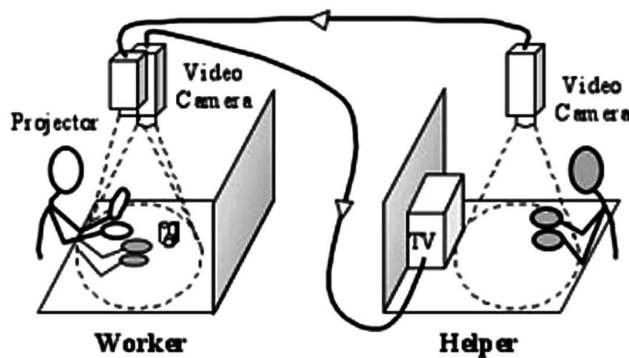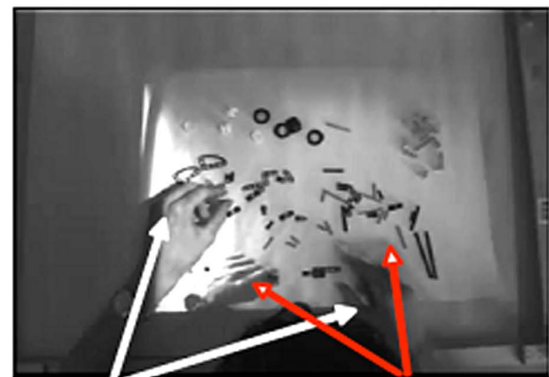


**Fig. (2).** The WACL system [12] (left) and the REAL system [17] (right).

**Fig. (3).** Sketches of hand gestures [6].



a: Projecting gesture into the Worker's local ecology

b: The mixed reality surface: a reciprocity of perspectives

**Fig. (4).** Projected hands [9].



**Fig. (5).** Worker interface.

In this section, HandsOnVideo is introduced to address the above needs. In particular, HandsOnVideo captures the hand gestures of the helper and projects them onto a near eye display worn by the worker. It is composed of 1) a helper user interface that is used to guide the worker remotely using a touch screen device and an audio link, and 2) a mobile worker interface that includes a wearable computer, a camera mounted on a helmet and a near eye display (a small device with two screens); see Fig. (**5**). In the following subsections, the design of our remote guiding system and the technical platform are described in more detail.

**3.1. Worker Interface Design**

When it comes to display information to a mobile worker, there are a range of types of displays that can be employed for this purpose [18]. For example, hand-held displays, wrist-worn displays, shoulder-worn displays, head-worn displays, displays that are embedded in the environment, displays that are specifically set up in the workspace, or devices that project images on the surfaces of arbitrary objects. According to Holler and Feiner [18], displays used for mobile AR can be generally classified into

two categories: displays that make use of resources in the environments and displays that the worker carries on the body or by hand. The display used in the system of Fussell *et al*. [6] mentioned in section 2 falls in the first category. Displays in this category require space, supporting facilities, and sophisticated infrastructure to be readily available. For remote guiding systems, one would need camera, projectors, monitors, desks, cables, power source and proper lighting conditions. The list could go on. All these devices need spaces and require strict environmental conditions to work properly. Even setting them up and making them all working together require extra expert instructions. Considering that our system is designed to be used in mining sites, in which the environment can be unpredictable. We therefore limit our options to the second category: displays that the worker carries on the body or by hand.

Since local workers need to move around and use their hands to operate on physical objects, it is not practical to use hand-held displays. In regard to head-worn displays, optical see-through and video see-through displays have been used for mobile AR guiding systems [18, 19]. Optical see-through displays are semi-transparent. They overlay computer-generated images on top of the worker's view of the real world. On the other hand, video see-through systems present an indirect and mediated view of the environment. They combine video feeds from cameras with computer-generated images and display the video representation of the real world in front of the user's eyes. Examples of contents from the two displays are shown in Fig. (**6**). Although both displays can be useful for general mobile AR systems, applications for mining sites have strict requirements in terms of human and environmental factors [20]. These factors include safety, ease of use, changing light and dust conditions of environments, and so on. The end users had indicated to us that mining sites are often very dusty. Dusts can easily spread on the surface of optical displays, blocking the view of the worker. This hardly makes optical see-through a practical option for consideration. In addition, both optical and video see-throughs offer a limited view of worker's workspace. And not being able to see the surrounding environment fully and directly is risky for the worker in mining sites.

Workers usually wear helmets while working in mining sites for safety reasons. We therefore make use of the helmet and attach a near-eye display under the helmet. As shown in Fig. (**5**), the near-eye display is light, easy to put on and comfortable to wear, compared to other head-worn displays. The worker can easily look up and see video instructions shown on the two small screens, and at the same time he/she can see the workspace in front of him/her with little constraint. We also tested the display with real users. The feedback from them during the design process was very positive with the near-eye display.

## 3.2. Helper Interface Design

We adopted a participatory approach for the design of the helper interface. Our aim was to come up with a design that fulfils the users' needs and that is as intuitive to use as possible. Our initial step consisted of observing maintenance workers and developing a set of requirements for the helper user interface (UI) based on our understanding of their needs.

- The need for supporting complex hand movements such as: "take this and put it here", "grab this object with this hand", and "do this specific rocking movement with a spanner in the other hand".

- Mobility of the worker during the task, as they move from being in front of the machine to a tool area where they access tools, to the back of the machine to check valves etc.

- The helper may need to point/gesture in an area outside the field of view of the worker. Therefore there is a need to provide the helper with a panoramic view of the remote workspace.

We then designed a first sketch of the interface consisting of a panoramic view of the workspace and a video of the worker's view. The video provides a shared visual space between the helper and the worker that is used by the helper for pointing and gesturing with their hands (using unmediated gesture). This shared visual space augmented by the helper's gestures is displayed real time on the near eye display of the worker (image + gestures).



**Fig. (6).** Optical see-through (left) and Video see through indoor AR (right) [18].

The helper UI consists of:

- A shared visual space which displays, by default, the video stream captured by the remote worker's camera. This space occupies the central area of the touch table.

- A panoramic view of the worker's workspace which the helper can use for maintaining an overall awareness of the workspace. This view can also be used by the helper for bringing the worker to an area that is outside their current field of view. The panoramic view occupied the lower end of the touch table.

- Four storage areas, two on each side of the shared visual space, to allow the helper to save a copy of the shared visual space in case there is a need to reuse it.

We performed four design iterations of our UI, testing and validating each design with a set of representative end users on the following three maintenance/repair tasks (Fig. 7):

- Repairing a photocopy machine

- Removing a card from a computer mother board

- Assembling Lego toys

Over 12 people have used and trialled our system, providing valuable feedback on how to improve the helper UI and more specifically the interactive aspect of the UI: the selection of a view, the changing of the view in the shared visual space and the storage of a view. The aim was to perform these operations in a consistent and intuitive manner, for ease of use. The overall response from our representative end users pool is that our system is quite intuitive and easy to use. No discomfort has been reported to date with the near eye display of the worker system.

### 3.3. The Platform and technical specifications

Our platform draws on previous experience in the making of the REAL system, a commercial, wearable, low-power augmented reality system. REAL employs an optical see trough visor (LiteEye 750) for remote maintenance in industrial scenarios. In particular, HandsOnVideo makes use of the XVR platform [17], a flexible, general-purpose framework for VR and AR development. The architecture of our system is organized around two main computing



**Fig. (7).** Maintenance and assembly task.



**Fig. (8).** The helper control console (left) and the worker wearable unit (right).

components: the worker wearable device and the helper station, as seen in Fig. (**8**).

Wearable computers usually have lower computing capability with respect to desktop computers. To take into account the usual shortcomings of these platforms, all our software has been developed using an Intel Atom N450 as a target CPU (running Microsoft Windows XP). It presents reasonable heat dissipation requirement and a peak power consumptions below 12 watts, easily allowing for battery operation. A Vuzix Wrap 920 HMD mounted on a safety helmet was used as the main display of the system. The arrangement of the display is such that the upper part of the worker field of view is occupied by the HMD screen. As a result, the content of the screen can be seen by the worker just looking up, while the lower part remains non-occluded. With such an arrangement, what is displayed on the HMD gets used as a reference, but then the worker performs all his/her actions by directly looking at the objects in front of him/her. CMOS USB camera (Microsoft Lifecam HD) is mounted on top of the worker's helmet (as seen in Fig. **8**). This allows the helper to see what the worker is doing in his/her workspace. A headset is used for the worker-helper audio communication.

The main function of the wearable computer is to capture the live audio and video streams, compress them in order to allow network streaming at a reasonable low bit rate, and finally deal with typical network related issues like packet loss and jitter compensation. To minimize latency we use a low level communication protocol based on UDP packets, data redundancy and forward error correction, giving us the ability to simulate an arbitrary values of compression/ decompression/network latency, with a minimum measured value around 100 ms. Google's VP8 video compressor [21] is used for video encoding/decoding, and the Open Source SPEEX library is used for audio, with a sampling rate of 8Khz. Please note that at the same time the wearable computer also acts as a video/audio decoder, as it receives live streams from the helper station and renders them to the local worker.

The main component of the helper station is a large (44 inches) touch-enabled display. The display is driven by a NVidia GeForce graphic card mounted on a Dual Core 2.0 Ghz Intel workstation (Windows XP). The full surface of the screen is used as a touch-enabled interface, as depicted in Fig. (**9**).

Occupying the central portion of the screen is an area that shows the video stream captured by the remote worker camera: it is on this area that the helper is using his/her hands to guide the worker. On the side of the live stream, there are 4 slots, initially empty, where at any moment it is possible to copy the current image of the stream. This can be useful to store images of particular importance for the collaborative task, or snapshots of locations/objects that are recurrent in the workspace. Another high-resolution webcam (Microsoft Lifecam HD) is mounted on a fixed support attached to the frame of the screen, and positioned to capture the area on the screen where the video stream is displayed (see Fig. 10): the camera capture what is shown on the touch screen (see arrow 1) and the hand performed by the helper over that area too (see arrow 2). The resulting composition

(original image plus the hand gesture on top) is once again compressed and streamed to the remote worker, to be displayed on the HMD (see arrow 3). The overall flow of information is represented in the diagram of Fig. (**10**).
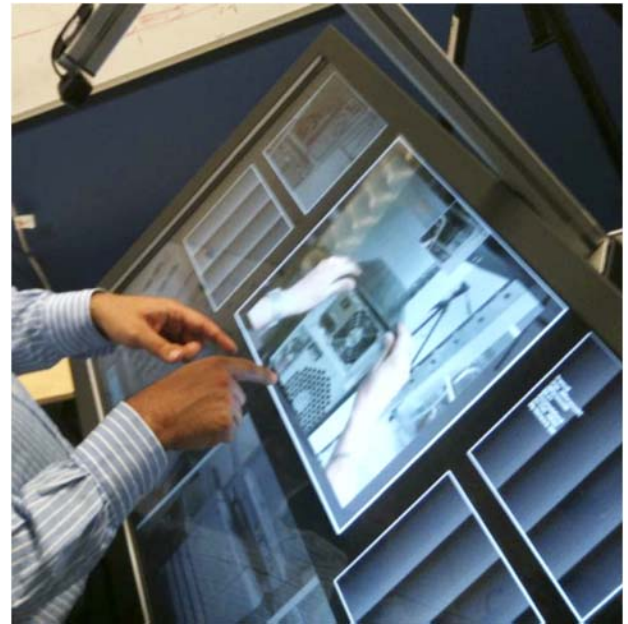


**Fig. (9).** Layout of the helper screen.

## 4. USER TESTING

In this section, we present a user study we conducted with representative end users. The main objectives were to collect feedback on the usability of HandsOnVideo and to identify possible directions of future work.

### 4.1. 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 a mining site. Both rooms were about 20 meters away from each other. The helper and worker could talk to each other through a headphone. Since the system was developed specifically for mining workers, end users who had experience with remote collaboration were asked to perform two different physical tasks. The users were randomly grouped in pairs with one playing the role of helper and the other playing the role of the worker. If one participant played as a helper in the first task, then in the second task, he or she changed to play as a worker. For each session, the whole process was video recorded on both helper and worker sites 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.

**Fig. (10).** Data capture and display.

## 4.2. Participants

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

## 4.3. Tasks

Two types of task were used. One is the assembly task using Lego toy blocks. This task has been used in previous research for similar purposes [6, 9]. This task is considered representative because it has a number of components that can be found in a range of physical tasks such as assemble, disassemble, select, move, and rotate. During the task, the worker was asked to assemble the Lego toys into a complex model under the instruction of the helper.

The other task is repair task. This is a real task that may occur in mining sites. Since we did not have access to mining equipment, we use the repair of a PC as our second task. During this task, the worker was asked to take the cover of the PC off, and 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.

## 4.4. 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.

## 4.5. 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 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. The worker was able to
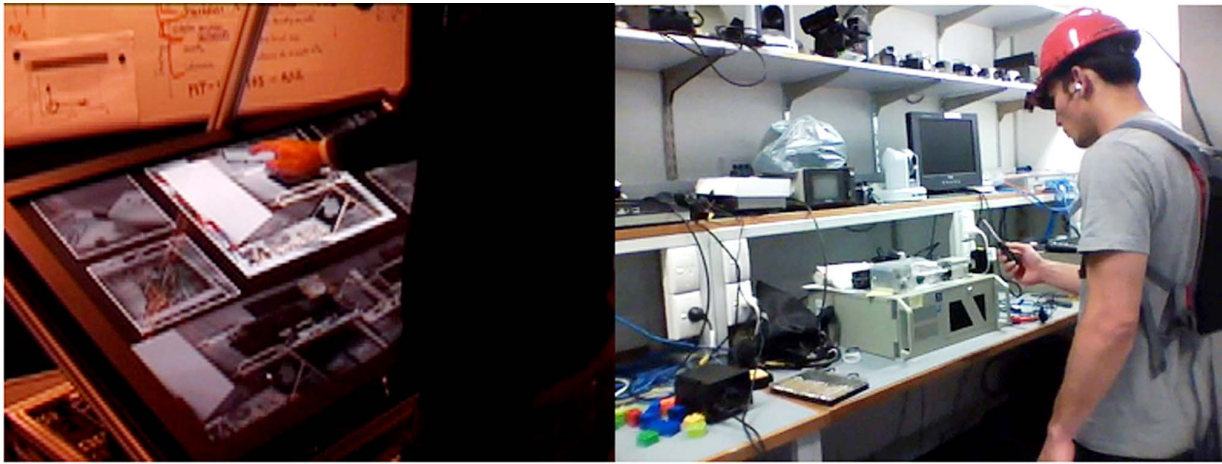
**Fig. (11).** Confusion with the view being displayed.

walk around the workspace without apparent difficulties. This demonstrates that the worker was able to be aware of the environment with the near-eye display. The communications between the pair seems smooth and effective. During the observation, we also identified some usability issues, which were detailed as follow.

Confusion with the view being displayed: The worker view on the near eye display does not correspond to what they see in their physical workspace (See Fig. **11**). This might be due to the network delay as mentioned by participants: "video lag is annoying", "video delay makes it harder to use", "video lag is not good".

During the study we observed from three of the participants that there could be an issue of spatial awareness with the use of the near-eye display (see Fig. **12**):

- 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 help 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 place. 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 checking the help being displayed. A further exploration on how the near-eye display should be configured is needed in order to prevent such issues from happening again.

### 4.6. Questionnaire Results

Six participants filled two questionnaires: the helper questionnaire and worker questionnaire. We got 12 copies of questionnaires 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: strongly disagree to 7: strongly agree. The results are shown in Table **1**,

As can be seen from Table **1**, in general, the participants thought that the HandsOnVideo system is easy to learn and use. On average, co-presence and environmental awareness were rated just above being neutral, while perceived task performance and interaction were rated relatively high. Specifically, 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



**Fig. (12).** Limitations of the near-eye display.

comfortable with the helper interface and that on the other hand, they might need more time to get used to the worker interface.

**Table 1.    Results of Likert Score Questions**

| Question | Worker | Helper | Average |
|---|---|---|---|
| ease of learning | 5.00 | 6.00 | 5.50 |
| ease of use | 4.83 | 5.66 | 5.25 |
| task satisfaction | 5.50 | 5.66 | 5.58 |
| co-presence | 4.50 | 4.16 | 4.33 |
| awareness of environment | 4.66 | 5.00 | 4.83 |
| perception of interaction | 4.83 | 5.66 | 5.25 |

In regard to co-presence, generally speaking, co-presence was rated relatively low compared to other measurements, it is reasonable since both the helper and worker knew they were located in different rooms and we did not expect that the system would present a sense of "being together" as strong as virtual environments would do. On the other hand, we expected that helper and worker might have different levels of senses of "being together", as there were different interfaces on both sides. The helper used relatively large touch display that showed the view of worker's workspace, while the worker used the small near-eye display that showed the workspace with hands of the helper. As can be seen from Table **1**, co-presence on the worker side was rated as 4.50, which is relatively higher than what was rated on the helper side (4.16). This indicated that the worker had a greater sense of co-presence. According to Li *et al.* [8], this difference might be the result of one of the key features that our HandsOnVideo system offered: the worker being able to see hands of the helper.

Second, the questionnaires also had open questions asking about their experience with the study and asking for their feelings about specific interface features. The comments indicated that 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"; "The near-eye display helped me to see what my partner could see". The pros and cons mentioned by the participants were summarized as follow.

Pros:

- Helper UI is intuitive and enjoyable to use, easy to learn.

- Guiding with hands is natural; being guided seeing hands is easy to understand.

Cons:

- Latency is annoying: users get the audio instruction first and have to wait for the visual instructions

- Image quality is not very good

- Jerkiness of the images affects users

Specifically, in response to the question: *What is your view of seeing the hands of person helping you?* Participants playing the role of worker commented that it was useful. "It was useful. I could see what he was pointing to, even though he couldn't see the exact colors or shapes on his screen, I could tell which object he meant."

On the other hand, in response to the question: *Please explain your experience of guiding/instructing using your hands.* Participants playing the role of helper commented that it is helpful. "I had to make sure I only had one hand on the screen. It was pretty easy to use. Might be useful to have some feedback of hand gestures thought, perhaps if we could see what the partner sees."

## 5. DISCUSSION

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 [9] 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 [9]. 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.

Designing a useable remote guiding system requires a close involvement of end users, and the ability to capture and address the interaction issues they raise while using the system. This is not an easy task and during the system design and development process, we have encountered a number of challenges. These challenges include:

- The trade off between the richness of the gesture supported by the system and the resulting latency it introduces. In our system, the two-way sequential process of capturing/encoding/streaming/decoding visual information introduces some intrinsic latency between the time the image is captured and it is displayed on the screen. We are currently exploring means by which we can extract the hands of the helper from the shared visual space and display the hands on the local video view of the worker.

- The trade off between the quality of the image/video projected and network latency. We are currently exploring ways in which we can provide a high resolution video of a subset of the shared visual space.

- The trade off between supporting the mobility of the worker while maintaining spatial coherence. 1) Because of the worker's mobility there are sometimes discrepancies between the view projected in the worker's display and the view the worker has of his physical workspace. This may disorient workers. There is a need for workers to maintain a spatial coherence. We are currently exploring gesture based interactions to allow the worker to change the view displayed on their view. 2) As the worker moves around, the shared visual view changes, gesturing on a changing target could become challenging. To address this issue, gesturing (pointing to a location or an object and showing orientation and shape, etc.) was initially performed on a still image. Helpers were required to freeze the video view in order to gesture. Gesturing on a still image not only adds an extra workload for the helper, but also results in the shared visual space not being synchronised with the view of the physical workspace. We are currently investigating how to allow the helper to gesture on the video view.

We believe that exploring these tradeoffs in series of laboratory experiments will provide a solid basis on how to design useful remote guiding systems.

## CONCLUSION AND FUTURE WORK

In summary, in this paper we have reviewed the literature on remote guiding, and put forward the case for supporting richness of gesture and mobility of the worker. We described HandsOnVideo, a system for a remote helper guiding a mobile worker working in non-traditional-desktop environments. The system was designed and developed using a participatory design approach. Our key research drive was to develop a remote guiding system that is truly useful, enjoyable, easy to use, reliable effective and comfortable for end users. The design approach we have taken has allowed us to test and trial a number of design ideas. It also enabled us to understand from a user's perspective some of the design tradeoffs. The usability study with end users indicated that the system is useful and effective. The users were also positive about using the near-eye display for mobility and instructions and using unmediated representations of hands for remote gestures.

Our future work is to investigate the expansion of the current system to a mobile helper station [22, 23]. In the remote guiding system currently developed the gesture guidance is supported by a large touch table. A fully mobile remote guiding system using similar technologies for the two parts of the system, the expert station and the operator station, will be easily deployable and adaptable in the mining industry.

We are currently engineering a rugged version of the system for initial field deployment and field studies. Industry deployment and the study of system use in its real context are crucial in understanding the human factors issues prior to prototype development and commercialization of the system.

The deployment of a rugged HandsOnVideo system to a mine site would allow us to investigate the following questions:

- What is required for mining operators to use the system effectively?

- What measurable benefits can be achieved from the system use in a mine, such as, productivity and safety?

- What ROI on maintenance cost could be obtained by means of a large deployment of several similar units?

## ACKNOWLEDGEMENT

## CONFLICT OF INTEREST

None declared.

## REFERENCES

[1]   Karsenty L. Cooperative work and shared visual context: an empirical study of comprehension problems in side-by-side and remote help dialogues. Hum-Comput Interact 1999; 14(3): 283-315.

[2]   Kraut, RE, Fussell SR, Siegel J. Visual information as a conversational resource in collaborative physical tasks. Hum.-Comput Interact 2003; 18(1): 13-49.

[3]   Palmer D, Adcock M, Smith J, *et al*. Annotating with light for remote guidance. In: Proceedings of the 19th Australasian Conference on Computer-Human interaction: Entertaining User interfaces (Adelaide, Australia, November 28 - 30, 2007). OZCHI '07, vol. 251. ACM, New York, NY, pp. 103-10.

[4]   Kraut RE, Miller, MD, Siegel J. Collaboration in performance of physical tasks: effects on outcomes and communication. In: Proceedings of the 1996 ACM conference on Computer supported cooperative work, ACM 1996; pp. 57-66.

[5]   Clark HH, Brennan SALB, Resnick JM. Grounding in communication. Perspectives on socially shared cognition, American Psychological Association 1991.

[6]   Fussell SR, Setlock LD, Yang J, Ou J, Mauer E, Kramer ADI. Gestures over video streams to support remote collaboration on physical tasks. Hum Comput Interact 2004; 19: 273-309.

[7]   Tang A, Boyle M, Greenberg S. Display and presence disparity in Mixed Presence Groupware. Proceedings of the 5th conference on Australasian user interface, Australian Computer Society, Inc., 2004; pp. 73-82.

[8]   Li J, Wessels A, Alem L, Stitzlein C. Exploring interface with representation of gesture for remote collaboration. In: Proceedings

of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces (Adelaide, Australia, November 28 - 30, 2007). OZCHI '07, Vol. 251. ACM, New York, NY, pp. 179-82.

[9] Kirk D, Stanton FD. Comparing remote gesture technologies for supporting collaborative physical tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM 2006; pp. 1191-200.

[10] Ou J, Fussell SR, Chen X, Setlock LD, Yang J. Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In: Proceedings of the 5th International Conference on Multimodal Interfaces, ACM 2003; pp. 242-9.

[11] Kirk D, Crabtree A, Rodden T. Ways of the hands. In: Proceedings of the 9th Conference on European Conference on Computer Supported Cooperative Work. Springer-Verlag New York Inc. 2005; pp. 1-21.

[12] Kurata T, Sakata N, Kourogi M, Kuzuoka H, Billinghurst M. Remote collaboration using a shoulder-worn active camera/laser. 8th International Symposium on Wearable Computers 2004; vol. 1: pp. 62-9.

[13] Sakata N, Kurata T, Kato T, Kourogi M, Kuzuoka H. WACL: supporting telecommunications using - wearable active camera with laser pointer. In: Proceedings 7th IEEE International Symposium on Wearable Computers 2003; pp. 53-6.

[14] Kuzuoka H, Kosaka J, Yamazaki K, *et al.* Mediating dual ecologies. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, ACM, 2004; pp. 477-86.

[15] Alem L, Tecchia F., Huang W. HandsOnVideo: towards a gesture based mobile AR system for remote collaboration. In: Alem L,

[16] Huang W, Eds. Recent trends of Mobile collaborative augmented reality, 2011. Springer, NY, USA, pp. 127-138.

[16] Huang W, Alem L. Supporting hand gestures in mobile remote collaboration: a usability evaluation. In: proceedings of the 25th BCS Conference on Human Computer Interaction, 4 July - 8th July, 2011; UK: Newcastle-upon-Tyne.

[17] R.E.A.L. (REmote Assistance for Lines), (c)(TM) SIDEL S.p.a. & VRMedia S.r.l, http://www.vrmedia.it/Real.htm. Accessed October 2010.

[18] Hollerer T, Feiner S. Mobile augmented reality. In: Karimi H, Hammad A, Eds. Telegeoinformatics: Location-based Computing and Services; Taylor & Francis Books Ltd. 2004.

[19] Rolland JP, Fuchs H. Optical versus video see-through head-mounted displays in medical visualization. Presence: Teleoper. Virtual Environ 2000; 9(3): 287-309.

[20] Horberry TJ, Burgess-Limerick R, Steiner LJ. Human factors for the design, operation, and maintenance of mining equipment. CRC Press 2011.

[21] The Google WebM project, http://www.webmproject.org/, Google Inc. Accessed October 2010.

[22] Alem L, Huang W. Developing Mobile Remote Collaboration Systems for Industrial Use: Some Design Challenges. In proceedings of the 13th IFIP TC13 Conference on Human-Computer Interaction (Interact11), Lisbon, Portugal, September 5-9, 2011.

[23] Huang W, Alem L. HandsInAir: A Wearable System for Remote Collaboration. In: Proceedings of the 2011 2nd Asia-Pacific Conference on Wearable Computing Systems (19 Mar - 20 March 2011, Changsha, China), APWCS11, IEEE Press, pp. 171-174.