Working Paper Series
ISSN 1170-487X

Use of Video Shadow for Small Group Interaction Awareness on a Large Interactive Display Surface

Mark Apperley, Laurie McLeod, Masood Masoodian, Lance Paine, Malcolm Phillips, Bill Rogers and Kirsten Thomson

Working Paper: 07/02
July 2002

Department of Computer Science
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

Use of Video Shadow for Small Group Interaction Awareness on a Large Interactive Display Surface

Mark Apperley, Laurie McLeod, Masood Masoodian, Lance Paine, Malcolm Phillips, Bill Rogers and Kirsten Thomson
The Computer Science Department
University of Waikato, Hamilton, New Zealand
Phone 64(7)838-4021
coms0108@cs.waikato.ac.nz

ABSTRACT

This paper reports work done as part of the Large Interactive Display Surface(LIDS) project at the University of Waikato. One application of the LIDS equipment is distributed meeting support. In this context large display surfaces are used as shared workspaces by people at collaborating sites. A meeting will start with a shared presentation document, typically an agenda document with summary and detail on agenda items as required. During the meeting, annotations will be made on the shared document, and new pages will be added with notes and drawings.

To prevent access collisions and generally mediate use of the shared space, mechanisms to provide awareness of actions of people at other sites are required. In our system a web camera is used to capture a low-resolution image of the person/people near the board on each side. Rather than transmit the image directly we compute a shadow/silhouette The shadow is displayed 'behind' other screen content. This provides awareness of position and impending write actions and allows intentional pointing to locations on the screen. It also has the advantages of being transmitted with low bandwidth, being relatively insensitive to low frame rates, and minimizing visual interference with the substantive data being displayed on the screen.

Keywords

Computer Supported Collaborative Work, Shadow, Video, Collaboration, Silhouette, Large Interactive Display Surface, Whiteboard Metaphor, Awareness.

INTRODUCTION

Much existing CSCW work is concerned with supporting individuals collaborating from personal workstations. However, collaborative work does not necessarily mean geographically separation. As computer design moves away from workstations towards 'ubiquitous' computing, it is appropriate to look at ways in which collaboration can be facilitated, not just at a distance, but also for groups of people in the same room. Without computer support, people use paper documents for agendas and pre-written reports. Shared document construction is done with blackboards, whiteboards, or large sheets of paper – in other words 'large display surfaces'. A computer managed large display surface – something that 'looks and feels' like a whiteboard – is a good example of ubiquitous computing; taking a familiar artifact and enhancing its capabilities. A computer managed display can offer: storage and replay of information; presentation of a wide variety of prepared materials from many sources; and it can offer traditional CSCW features – remote support, not just for individuals, but for cooperating groups at each of several locations. We use the term LIDS (Large Interactive Display Surface) for such a system. LIDS systems are not new. The Xerox Liveboard system [3] has been available for some time. What is new is that such systems can now be easily assembled from relatively inexpensive and readily available equipment.

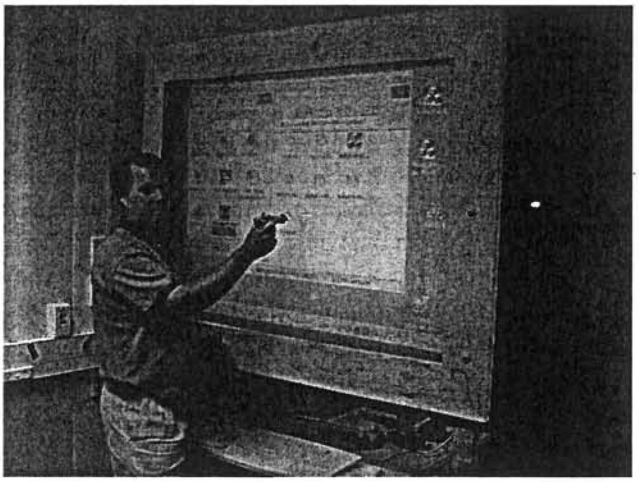

Figure 1. Experimental LIDS system.

The Waikato University LIDS project has as its objectives, taking advantage of readily available hardware components to build and make available Large Interactive Display Surfaces, and developing software to exploit this environment. Our LIDS systems are office whiteboard sized screens functioning as workstation displays, with digitization equipment at the front allowing a person to directly interact using a pen (with similar functionality to a mouse). Data projectors and a rear projection screen are used so as to allow the operator(s) to work directly in front of the screen without occluding the image. One of our experimental systems is shown in Figure 1.

The basic LIDS system – large screen and pen used in 'mouse' mode – is attractive to use in a number of settings: teaching, demonstration, software evaluation and testing. The large display makes it possible for a group of people to sit and watch. Alternatively, a small group can stand around the display and take turns at contributing. In contrast to a conventional computer and data projector setup the operation of a LIDS system is completely transparent. The person operating the system can work entirely in view of their audience. This is particularly powerful when demonstrating software or discussing usability issues. Even for the individual user the directness of object manipulation is useful. Rather than tinkering with a mouse out of view of an audience, or to the side of the display, the user actually touches object images with the pen. In a group situation the shared document presented on the screen is the focus of everyone's attention. *What is done is what is seen being done* by everyone present.

Our interest in software development has been to try to produce applications that retain as much as possible of the simple and intuitive interaction style people enjoy with real whiteboards. The benefits of free sketch interaction over formal and formatted presentation have been documented elsewhere [8]. We have developed experimental software in this style to use LIDS systems for lecture capture and remote meeting support. Of course, involving a computer allows us extra functionality. Prepared information can be displayed, sessions can be recorded for later playback, and content can be relayed to collaborating remote sites. The challenge is to minimize the cognitive overhead of using this functionality so that it does not spoil the simplicity of the basic interaction. We have reported on our software development in [1,2].

In our meeting support software the LIDS screen provides a common workspace. Initial display content at the start of a meeting is provided as a presentation file using Microsoft PowerPoint™. This might include an agenda and its supporting documents. Participants at each site can then draw on the surface. This new data appears as overlaid annotation on the original information, and is visible at each site. The software also provides an audio connection between the sites and this is the primary channel of communication [9]. Of particular interest is the problem of mutual awareness for participants at each site. There is a big difference between watching a person draw on a screen and simply seeing lines appear without warning. The advantage of transparency of operation as described earlier is lost for remote participants.

To resolve this problem we investigated options for using a video channel. Other workers have used video in a number of different ways; some of this work is summarized in the next section of this paper. As our whole project is based around using off-the-shelf components to provide relatively inexpensive solutions, we did not want to introduce complex or expensive video equipment. We are also interested in communicating over the internet, possibly through modems, but certainly without special high capacity data channels, so again we were limited in the style of video we could use. We saw the major purpose of the video channel as helping people cooperate in manipulating their shared document on the screen. This seemed to require some awareness of people's position, especially when they were about to write on the screen; and ability to gesture at features on the screen. Less important were detailed facial images or awareness of people's positions in the room beyond each screen.

To this end our system takes on the idea of the screen as a 'window' between meeting rooms, as described and developed by Fish et al [4], but instead of providing a full video view through the window, implements a monochrome outline image 'behind' the other information displayed on the screen. The analogy is to the screen being a translucent window into the other site, with a bright light at the back of the room, so that a person near the window casts a shadow on the screen. In fact the implementation is not exactly that of a shadow because it is reversed left to right and can have highlighted areas, but it tends to be interpreted as such. For the purposes of this discussion we will refer to the image as a shadow. We feel that our system provides a good balance between awareness and distraction from the document being shared.

In the following sections we will outline other work with video in CSCW; describe our system in more detail, comment briefly on our implementation; and present some preliminary findings from usability studies.

OTHER WORK

There has been a considerable amount of work done on the use of video in CSCW situations, ranging from video conferencing through a variety of systems in which video is used to provide position and even gaze awareness for participants sharing the task of constructing a document.

In the 1970's Krueger's VideoPlace and Video Desk Systems [7] made use of silhouettes for communication between remote participants. In VideoPlace users stand in front of large display screens. A camera mounted below or to the side of the screen takes a nearly front on view of the participants. Careful control of lighting and background allows the camera to directly capture a silhouette. The resulting images were combined between sites and augmented with computer created graphics, to create a virtual world in which people could interact with each other and with graphic objects. Digitised versions of the silhouettes were subject to image recognition for the purposes of software control. Krueger provides strong evidence for the usefulness of the silhouette in communication, reporting that the silhouette enabled quite elaborate interaction between participants, even allowing people to dance together. We note however, that Krueger's participants were deliberately creating images on the screen. The focus was not on awareness of participants incidental activity. The silhouettes were a fundamental part of the shared 'document' being created. As a result, exaggerated gestures and movements were appropriate. A problem with the system was that hand and arm movements could only be seen when the arm was out to the side of the body. In VideoPlace this was a problem for control software, but not an issue in the display as participants would move to achieve whatever effect they required on the screen.

In VideoDesk Krueger uses a lighted desk surface and a front mounted camera, above and behind the user, to capture silhouettes of participants' hands. Software allowed graphic operations under control of hand gestures. This system did not involve collaborative work, but again demonstrates that simple silhouettes can provide reasonably precise position information.

Morikawa and Moesako's HyperMirror system [10] is somewhat similar in application to VideoPlace. Instead of using silhouettes HyperMirror uses full video images which it superimposes to generate a shared meeting place. Images of people are isolated from a blue painted background by using chromakey technology.

A system using a shadow system to support collaborative drawing between two sites is the VideoWhiteBoard system by Tang and Minneman [11]. Participants used whiteboard markers to draw on translucent screens. At each site a camera mounted behind the screen captured the screen content and any shadow that the user cast on the screen surface. The image was relayed to a video projector mounted behind the screen, close to the camera, at the other site. In this system there was no computer mediation. A shared picture could be produced, but participants were restricted to editing their own contributions – they could not erase each other's work for example.

Tang et al report use of hand gestures referencing the images on the screen and of upper body gestures, for example shrugs. In one case they observed a participant cock her ear towards the projection screen to indicate that she was having difficulty hearing what was being said. The person on the other side responded by speaking more loudly. The authors also comment on the importance of seeing an image being drawn by someone, albeit just that person's shadow, as opposed to having content simply appear on the screen. They conclude that this contributes positively to the sense of co-presence and is superior in that effect to the use of tele-cursors. They also comment on the advantages of superimposing the shadow on the entire shared workspace, allowing awareness without division of attention such as might occur with separate workspace and video

A limitation in providing position awareness was that the VideoWhiteBoard system only generated a detailed shadow when a person was close to the screen. As they move back the shadow becomes more diffuse and can disappear leading to a sense of lost contact. Because a full video image is being projected the effect of moving back is presumably that the shadow blurs, becoming fainter around the edges. We note though that loss of contact is always a problem with video systems, unless some kind of camera tracking is used, as stepping out of the field of view can happen easily. An interesting question, not addressed in Tang et al's paper is the question of awareness of a group of people. The fact that the shadow becomes more diffuse with distance may be an advantage here – providing a loose awareness of a 'shadowy' group of distant participants as well as the detailed view of anyone close to the board.

The TeamWorkStation system of Ishii and Miyake [5] provided an elaborate workstation environment in which video could be used in two different ways. A camera focused on the workstation operator could provide a 'talking head' window on the remote workstation. A second camera pointing down at a desk surface transmitted an image of paper documents, and the hands and arms of the operator. In addition the system provided images of remote computer screens, remote computer mouse/keyboard operation and an audio channel. A most interesting novel aspect of this system was the use of overlay in the video. Participants have two display screens; one private and one displaying shared content. Display content is assembled by mixing video layers. For example, a user could view a remote computer screen overlaid by a view of the remote desk surface. This provided a variety of operation modes allowing gestures referring to paper documents or to artifacts on the screen. The authors report that users were able to *differentiate up to three overlaid video images without much difficulty*, although they note that quality of the overlaid images was a problem, limiting options for sharing detailed documents.

The video image was of sufficient detail to allow interactive sessions teaching calligraphy. A special advantage here (also reported by the VideoWhiteBoard team) was that collaborating participants could simultaneously work on the 'same' part of a shared document. In calligraphy lessons, for example, a teacher could write directly over a student's strokes to show how errors should be corrected. In usability experiments the TeamWorkStation group also quote positive comments about the advantages of direct pointing at documents, reducing the need for verbal descriptions of position.

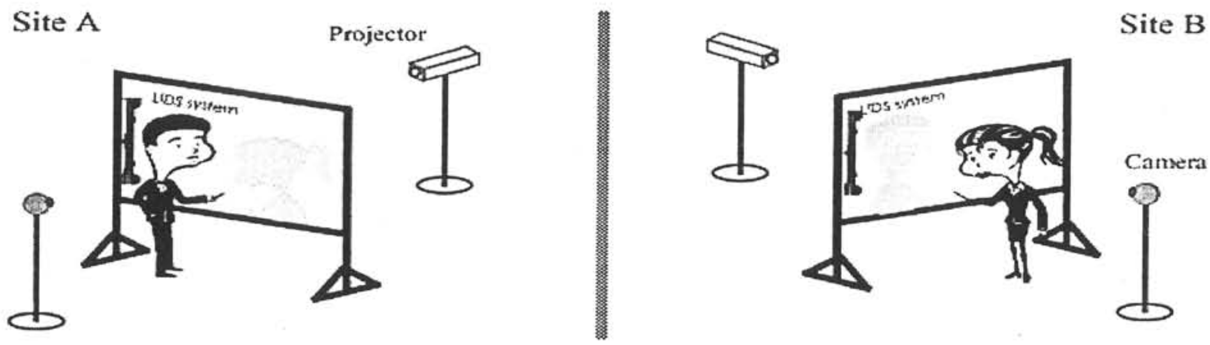

Figure 2. The LIDS video shadow system.

silvered mirror allows the ClearBoard system to capture video images of each participant as though the camera were looking at the user directly from on or behind the display. The result is that the image each person sees of the other is correctly placed with respect to the shared workspace and each person can see at just what part of the surface the other is looking. Each participant can be aware of the other's gaze. In experiments it was found that this considerably facilitated cooperative activities.

Each of the systems described here has attempted to provide awareness of other participants in a shared document creation environment by use of video. In each case the video is displayed as part of a composite image, rather than being displayed in a window to the side. VideoPlace and VideoWhiteboard demonstrate that shadow or silhouette forms of video can provide useful information and contribute strongly to a sense of co-presence. The TeamWorkstation results show that participants can interpret layered image displays, and switch attention between layers with little effort. It also demonstrates the value of allowing participants to gesture to points on the display surface. ClearBoard shows what can be achieved with more elaborate technology and establishes the importance of gaze awareness. Each of these systems made use of quite complex video equipment and required reasonably high communications bandwidth.

The system we have designed doesn't attempt to achieve as sophisticated a result as ClearBoard, but does work towards results comparible to the other systems. Our goal was to develop good functionality with simple and inexpensive equipment.

THE LIDS SHADOW SYSTEM

Figure 2 shows the physical setup we use. The large display surface is rear projected. A Mimio™ digitiser is used to track pen position. In the room space a small web camera is mounted on a stand, so as to be pointing directly at the center of the screen, distanced so that the screen display fills its field of view. In our system this means that the camera is approximately two metres from the screen; that distance varies with the focal length of the camera. A flatter image could be obtained by using a longer focal

Figure 3. Participant with back to camera, pointing.

The image picked up by the camera is converted to shadow form and transmitted to other sites. Figure 3 is a screen shot from system B showing the shadow of a person standing in front of system A with a nearly blank screen. The person is pointing at a point in the upper right quadrant of the display surface. (The large blob at lower right is a clipboard they are holding.) Provided that the camera is well aligned with the display, pointing is quite accurate – it is possible to indicate a point to the nearest centimeter or two. In particular, it is easy to select a word or a line in a typical presentation size font.

Shadow transmission occurs at about ten frames per second. Lag from image capture to display is not much more than one tenth of a second. The result is fast enough to watch comfortably. The system benefits from the shadow format in this respect. If the image were showing a human face in detail we would expect it to update fast enough to follow facial expressions and lip movements during speech. Because the shadow is only being used for rough position awareness, the low frame rate is not significant to the viewer.

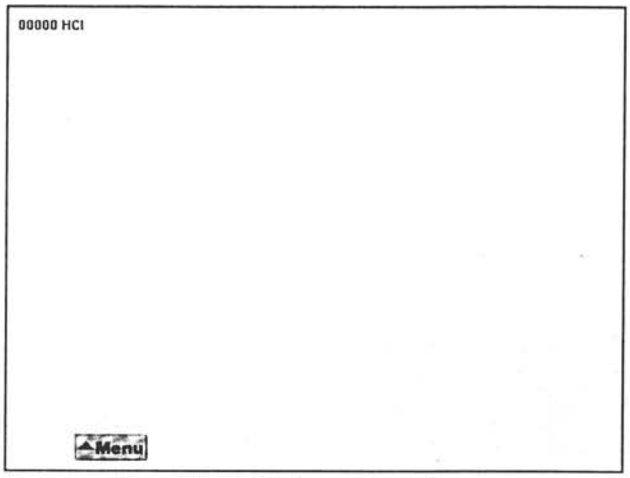

Figure 4: Subject facing partly towards camera.

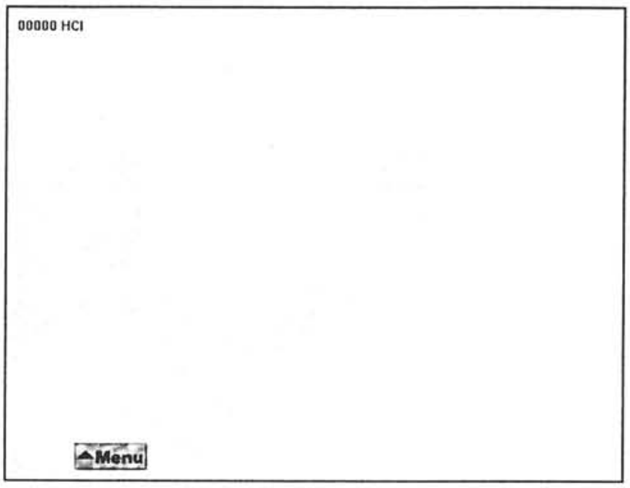

Figure 5: Subject with pale coloured shirt.

The example in Figure 3 looks very much like a shadow. In fact, it is obtained by colouring parts of the image that are darker than the white of the screen background. In Figure 4 the same person has turned to partially face the camera. There was a bright ceiling light above and just in front of the camera that highlighted the forehead and the tops of nose and cheeks. The shadow algorithm therefore leaves those areas as gaps in the shadow. This is strictly not a 'shadow', but a posterised form of the image. The same effect can be seen in Figure 5. The person in this image was wearing a light coloured t-shirt. Folds in the fabric show up around the shoulder and much of the body is missing.

We decided that this style of shadow generation offered useful features. Despite the low resolution, the moving form of the image appears to carry a surprising amount of detail. Recognition of facial expressions is not possible, but the detail is such that it is usually easy to identify people. Because the brain integrates information from several images, quite a strong impression of facial appearance can be experienced.

There is another important sense in which the image is not a shadow: because the image is captured from the front, it is a mirror image of a true shadow. When a person at another site writes on the screen it is essential that the writing be displayed the right way round - if the person was really standing behind a window, their writing would appear reversed. The effect is rather curious to watch. When text is being written there is a strong impression that it is being deliberately written in reversed form from behind the screen. When lines are being sketched there is no such impression. The person simply appears to be drawing lines.

The displayed screen image is assembled from three layers: the shadow; the presentation; and any annotations that have been drawn. The shadow is at the back and therefore covered by text and images. Figure 6 shows the shadow with presentation data.

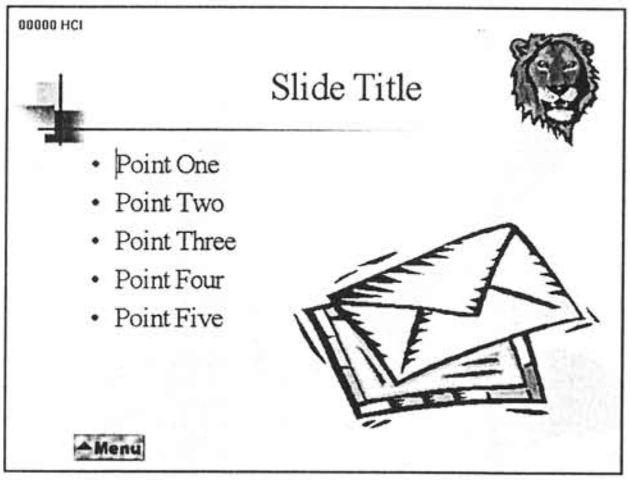

Figure 6: Shadow behind text and pictures.

Notice that the text is clearly on top, the picture of envelopes is partly transparent so the shadow arm can be seen in parts through it and the lion face is opaque so that the shadow hand is truncated at its edge. The illusion of 'being behind' is much stronger on the real system than it is with the screen shot printed here, because the shadow is moving while the foreground document is not. This strongly promotes the impression that the shadow is behind. Also, in the real system the shadow is paler than it appears in printed images from screen shots. This makes it less intrusive.

The fact that the users are annotating slides, and that slide content changes as we move from slide to slide poses a considerable problem in generating the shadow image. (Also note the cursor on the screen before the 'P' of 'Point One'. Editing of slide text during a presentation is supported.) The camera picks up the content and annotation and without additional processing would interpret this material as contributing to the shadow. An actual camera view is shown in Figure 7, overlaid on the same foreground image as Figure 6. (The camera image is of much lower resolution than the display image, and so appears smaller.) Notice that the camera picks up the screen content quite clearly. If the camera image is scaled to the size of the display the result is as shown in Figure 8.

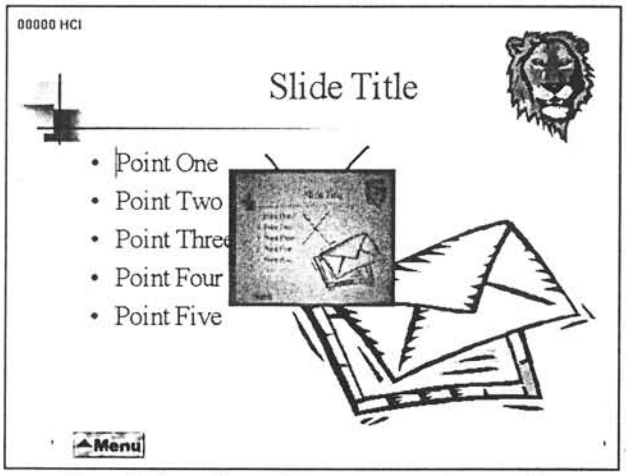

Figure 7. Screen image with actual camera view overlaid.

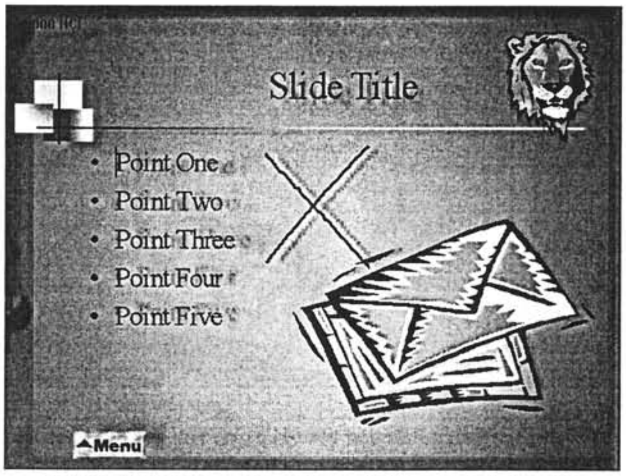

Figure 8. Camera image magnified underlying presentation data.

The alignment between camera view and objects on the screen shown in Figure 8 is quite close. It is obtained by careful physical placement of the camera. In fact this proximity suffices to provide accurate enough shadow positioning to allow a person to clearly point to something on the screen. The display with shadow therefore, is essentially the view shown in Figure 8, after a posterisation algorithm has been applied to the video data.

The difficulty that arises is in doing the posterisation. Dark areas of the camera image may be a result of a person between the camera and screen, or they may just be images of data displayed on the screen. The misalignment of image and camera view is sufficient to produce 'shadows' of screen artifacts that are close to, but obviously displaced from those artifacts. In Figure 8 the displacement is most obvious for the large X drawn in the middle of the screen. The camera image appears clearly slightly to the right of the X itself. When a shadow is generated the misaligned versions of artifacts appear as part of the shadow image.

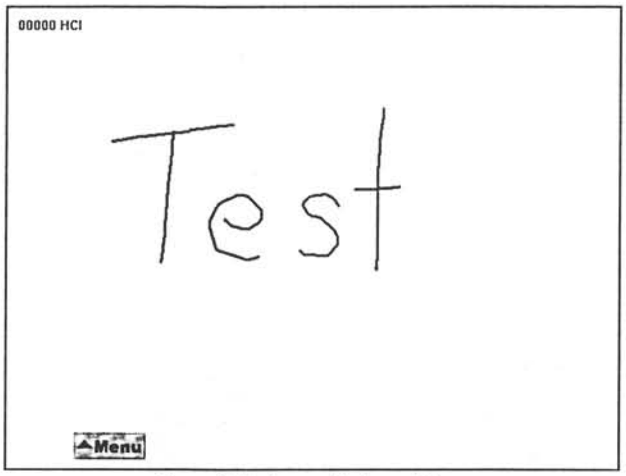

Figure 9: Background echo problem.

The effect is illustrated in Figure 9 where the word 'TEST' drawn on the screen has a clear shadow. The effect is particularly annoying when there are many thin lines on the screen, as occur with text displays and most annotations. Opaque images do not matter quite so much as the bulk of the additional shadow is hidden behind the images.

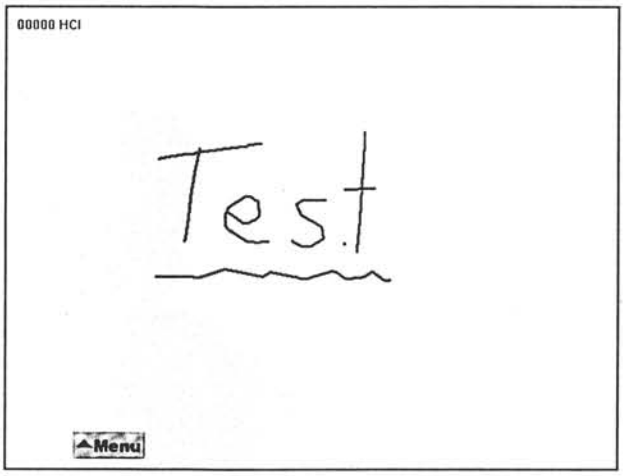

Figure 10: Dynamic background elimination.

It might seem that more careful alignment of the camera would improve the situation. If the camera and screen image were perfectly aligned such shadow additions would disappear behind the content or annotation responsible. Unfortunately, sufficiently accurate alignment is not possible. Instead we have implemented a dynamic background elimination algorithm. Based on an initial calibration process, the algorithm estimates the position of additional shadow and removes most of it. Figure 10 shows a similar experiment to that shown in Figure 9 with the background elimination algorithm enabled.

The fact that each camera is looking at the shadow generated from the other's image might be expected to cause a further video 'echo' phenomenon. We avoid such an effect by making sure that the shadow is displayed quite faintly. This keeps the shadow image below the threshold of the image posterisation and ensures that shadows of the shadows don't propagate around the system. It is important to keep the contrast of the shadow low in comparison to other data on the screen anyway, to minimise distraction.

IMPLEMENTATION ISSUES

The most significant implementation difficulty was managing the camera image. The camera we used had automatic brightness and contrast adjustment which could not be disabled. We persevered with the camera because we felt that building software that was reasonably robust to lighting levels was important, and the camera provided abundant challenge in that respect. The current version of the software works reasonably well in a range of lighting conditions, but does depend heavily on a calibration phase. Lighting change during a session will degrade the shadow generation.

As indicated in the previous section we relied on physical positioning of the camera to achieve alignment between the shadow and the computer-displayed data on the screen. This alignment is not very critical and is quite easy to establish. Precise alignment information is needed for background elimination when generating the shadow.

The shadow background elimination algorithm requires an accurate calibration of the mapping between camera image and screen. This is done during camera calibration when the program starts. A series of squares is displayed on the screen. An image of the screen is analysed for each square to determine its actual location in the camera image. The resulting pattern (10 by 8 square locations) is used to generate a piecewise linear mapping between display screen and camera image. When data is being displayed, for example when drawing a line on the screen, the mapping is used to generate an expected 'background echo' in camera image space.

When generating a shadow, the posterisation algorithm ignores brightness data in the echo regions of camera images. By itself, this removes the bulk of background contribution to the shadow. However the algorithm is only accurate to the nearest pixel, and some background effect remains. A further reduction in background is achieved by filtering out isolated pixels from the generated shadow. The algorithm does not need to be perfect because the shadow is a background feature of low intensity. Imperfections in the display are not a severe problem.

Background elimination does complicate the shadow generation in one respect however. It is necessary to decide, for each pixel in the camera image, whether the point is part of the shadow or not. For pixels in the background echo areas there is no basis for making the decision. If all such pixels are treated as not being part of the shadow, the result is a shadow with white gaps in it. Fortunately the problem is only an issue for small areas of echo. Camera/screen alignment is good enough to ensure that large areas of uncertain shadow are out of sight behind the front video layer. We resolve shadow content for pixels in 'echo' areas on a last known horizontal neighbour basis. This is not quite sophisticated enough and does lead to some anomalies in the shadow. Notice in Figure 10, just at the junction of top and down strokes of the capital 'T' that there is a spurious line of shadow. A nearest known neighbour algorithm should further reduce this error, but we have yet to implement it.

In contrast to earlier systems we have been able to take advantage of the speed of modern computers. A single desktop machine, with a 1GHz processor, at each site does everything. It takes a raw video image, applies the background cancellation algorithm and generates the outgoing shadow. It builds an image from the incoming shadow data, mixes the three video layers and presents the result on the screen. Simultaneously it manages the audio stream and saves audio to disk if required. Most importantly we avoid the complexity and cost of video equipment. The system is therefore relatively simple to set up and operate.

The penalty we have paid for single computer operation is that we cannot support high screen resolution. Our experimental system has a display resolution of 640 by 480 pixels and a camera (shadow) resolution of 176 by 144 pixels. With everything operating the computer runs at approximately 50% processor utilization. In presentation mode the resolution is not limiting because the audience could not see fine detail anyway. In close interactive use, where people stand around the display, higher resolution would be a benefit.

USABILITY ASSESSMENT

We have conducted a preliminary usability trial of our system, mostly to obtain qualitative results on user reactions to our shadow technology and to identify operational problems. The experiment was focused on the shadow, rather than other aspects of the LIDS equipment and software. We used pairs of people and looked at the way the shadow facilitated their interaction. Although the intended application of the LIDS system is meeting support, which would involve larger numbers of people, we wanted to look at the way in which the system facilitated (or failed to facilitate) the interaction of the people actually standing at the screen – whether it helped them in the social process of sharing the drawing space. Quite separately we have tried the shadow in a meeting setting and had favourable reactions from observers.

Trials were conducted with pairs of participants under three conditions: working side by side at an ordinary whiteboard; collaborating via LIDS screens with audio connection but no shadow, and collaborating via LIDS screens with the shadow. A limitation of the trial was that we were only able to provide the shadow in one direction at the time, ie only one participant could see the other's shadow. The participants were swapped over during the experiment in order that both experienced seeing the shadow. In addition, the background cancellation algorithm had not been implemented at the time of the trial, so that a significant amount of video noise was present in the shadow display.

Twelve pairs of participants were run through the experiment, 9 females and 14 males, recruited by advertising posters around our university campus. Each pair of participants participated in three tasks: playing noughts and crosses (tic-tac-toe); solving a crossword style puzzle in which a provided list of vowels had to be placed into crossword cells; and a word search puzzle. The order in which the different tasks were presented to the participants, and the technology conditions under which they tried each task were counter-balanced. Each of the tasks seemed successful in engaging the participants' interest.

A number of faults in the software were uncovered during the trial, particularly during the first sessions. In particular a need for performance improvement and the background echo cancellation were noted and have been implemented. Faults in operation of the system limit the value of statistical results, and many qualitative results simply concern the faults. However, the overall results were still encouraging.

Participant responses to perception questions, are summaried in figure 11. Generally participants felt that the shadow software was simple, low tech, easy to use, friendly and mostly they liked it. There were significant reservations about its attractiveness and quality. These issues seem to be mostly related to the need for background elimination in the version of the software used in the trial. Respondents disliked the video noise that accompanied the shadow.

The questions about group interaction were as follows. Participants were asked to respond on a scale of 1 to 5, representing 'strongly disagree' to 'strongly agree'. Mean response values are shown in Table 1.

(1) I found my group worked well together.
(2) I found that I was aware of the other user at all times.
(3) I found that I was aware of where the other user was standing at all times.
(4) I found that I was confident that the other user was aware of what I was doing at all times.

The major result is the answer to question 3. The shadow lifts participants' confidence about the other user's position to nearly the same value as in the non-distributed situation where participants were together in the same room (the difference is not statistically significant). Without the shadow participants did not have that awareness.

Q	Non-Distributed ΦN	Distributed No Shadow ΦD	Distributed Shadow ΦS
1	4.2	4.1	3.9
2	4.3	3.8	4.3*
3	4.3	1.9	4.0**
4	4.6	3.4	3.6#

* ΦS was significantly different from ΦD at p=0.05

** ΦS was significantly different from ΦD at p=0.01

ΦS was significantly different from ΦN at p=0.01

Table 1: Group interaction responses for technology scenarios.

The answers to question 4 do not show any benefit from the shadow. Only limited conclusions can be drawn here, as the shadow image was not presented to both participants. Nevertheless it is a disappointing result as it indicates that participants did not confidently expect to be able to have pointing and other gestures interpreted by their partner. To some extent this result is contradicted by some of the free response question answers. It may also be that participants were not very concerned as to whether their partner knew where they were in the distributed situations, perhaps because the problem set did not make enough demands on mutual position awareness. Indeed it was observed that the word search problem was often handled by an agreement to partition the search set between participants.

A representative sample of free response question answers follows. In each category the full number of answers is shown.

Awareness of position (12)

shadow gave relative position of other person
aware of other's position but not details
useful to know someone is at other location
my focus was drawn to her because of the shadow
without shadow could not see partner's location

Awareness of actions (11)

could predict other person's moves from their shadow
without shadow did not know where other person was going to write
distributed but without shadow, sometimes both started on same task at same time
they could see what I was doing

Assisted communication (5)

shadow meant less verbal communication needed

Audio more important (7)

relied on voice communication
used audio as well as shadow

More personal (7)

shadow (knowing where they were) made me feel more comfortable

Fun (5)

exciting/interesting to see other user

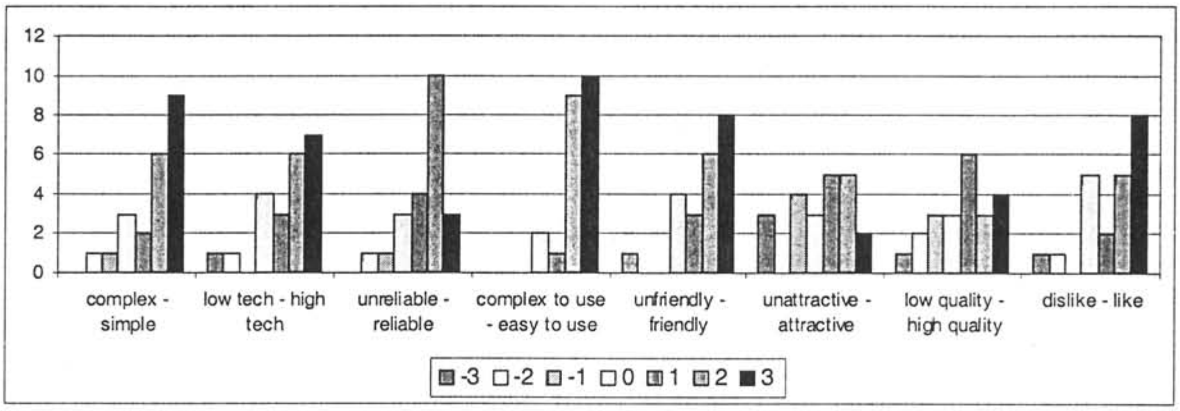

Figure 11: Participants' perceptions of the shadow technology

Easy to use (6)

Distracting (7)
sometimes shadow might distract
shadow can be annoying
shadow from letters was distracting

Not distracting (5)
easy to adapt to as does not impose on the screen
after a while shadow was irrelevant (ignored it)

Focus (3)
concentrated on other person as well as task when had shadow

Obscuring (2)
cannot see partner's actions if writing in front of themselves
shadow obscured a little

Helpful (5)
shadow technology added that extra dimension
partner displayed in front of me rather than peripherally and could see what partner was doing

Not helpful (4)
added no value in helping perform task
just as easy to use without shadow(18S10)

Shadow quality (7)
blocky but recogniable
large => positioning not exact
shadow actions not always clear
parts of other person did not cast a shadow
shadow slow moving and should have smoother edge
little room for gestures
shadow is not like having a proper image

Technical problems (2)
computer too slow with shadow
somewhat disconcerting talking to a shadow with a rough voice

The general tenor of comments supports our design intentions and are mostly positive. In particular participants say that the shadow provides useful position awareness and helps provide a sense of the other person's presence and impending actions. The results are weak on providing direct evidence of use of gestures. The crossword puzzle was the problem that required the greatest special cooperation. The issue facing the participants was to choose the right place to put each of the letters they were given. We have observed some use of gestures in this problem. It would be useful to run another experiment that required the participants to work more cooperatively.

It is clear from the results that some aspects of the implementation need attention. As already noted participants saw the background cancellation and tidiness of the shadow as important. Performance is also important in maintaining a sense of interaction. It is clear that improving the resolution of the shadow is an issue. We note here that the blocky, pixely edges of the shadow are particularly noticeable to participants standing close to a large screen display. They are not nearly as apparent in the screen shots included with this paper, simply because the printed images are so small. To some extent this problem can be managed by controlling the contrast between the shadow and the background. If the shadow is pale, on a white background, the edges are less noticeable than for a dark shadow.

CONCLUSIONS

Our prototype shadow video system appears to satisfy our design goals for contributing to a meeting support system. It appears to be successful at providing awareness of the presence of, and impending contribution to a shared document, by people at a remote site. It supports our whiteboard ideal of *what is done is what is seen being done* in a presentation context.

The system permits reasonably accurate, to the nearest centimeter or two, pointing to items on a large (1.2m by 0.8m) display surface. This does depend on the user being careful not to obscure the camera view with their body. Given that the application is meeting support it would be expected that a presenter would be careful to ensure that their gestures were clear to other people in the room anyway.

In contrast to VideoWhiteBoard the shadow is well defined and its clarity does not depend on very close proximity to the board. In this sense it is better suited to large-scale gestures involving the upper body.

Setting up is relatively simple. The entire system operates with through a single computer at each site, with a standard network connection. The only video equipment required is an inexpensive USB web camera. The bandwidth required for the video signal is limited to 20 kilobytes per second or less, depending on the complexity of the shadow image. Display equipment is a rear projection screen and data projector. Alignment of the camera is necessary, but need not be very high. Calibration in the software takes care of the fine detail.

The use of a layered video display allows the user position and impending action awareness without the need to shift their gaze from the document being developed. Our shadow system offers good control over the degree of distraction caused by the background. Because it is entirely computer generated and monochrome we have a choice of colour and intensity. The image can be pale enough to be barely visible if there is a need to minimize distraction. There is also the possibility of varying the display style. We have chosen a solid style for the shadow, because most of the information displayed on top is composed of lines and text. The pale shadow contrasts maximally with this and is therefore easy to distinguish. With minimal effort we could reprogram the shadow to appear in a 'sketch' mode, composed of lines, if that were helpful in presenting it against a background that had significant areas of pale solid colour.

Our meeting support software is intended to work with more than two sites. In particular the audio and annotation channels will work comfortably over three or four locations. We have not yet extended the shadow video to operate with more than two sites. In principal it is not difficult to do this. The processing cost would not be excessive, as several incoming shadows would be assembled into a single image at camera resolution, and only expanded to full screen size (the most time consuming step) once. Shadows from different locations could be distinguished by colour, however it is not clear how best to deal with overlap.

Whilst lacking the detailed video of systems like ClearBoard we think that the LIDS shadow video system provides useful functionality in a simple and easily deployed package.

ACKNOWLEDGMENTS

The research described in this paper has been carried out with the support of the New Zealand Foundation for Research, Science and Technology under the New Economy Research Fund contract number UOWX9910.

REFERENCES

1. Apperley, M.D., Dahlberg, B., Jeffries, A.Y.J., Paine, L., Phillips, M., Rogers, W.J., Applications of Large Interactive Display Surfaces, *Proceedings of SIGCHI NZ Symposium on Computer Human Interaction*, Palmerston North, New Zealand, July 2001
2. Apperley, M.D., Dahlberg, B., Jeffries, A.Y.J., Paine, L., Phillips, M., Rogers, W.J., Lightweight Capture of Presentations for Review, *Proceedings of IHM-HCI 2001*, Lille, September 2001
3. Elrod, S., Bruce, R., Gold, R, Goldberg, D., Halasz, F., Janssen, W.; Lee, D., McCall, K., Pedersen, E., Pier, K., Tang, J., and Welch, B. Liveboard: a large interactive display supporting group meetings, presentations and remote collaboration. *CHI '92 Conference Proceedings*, p 599-607, 1992.
4. Fish, R.S., Kraut, R.E., Chalfonte, B.L., The VideoWindow System in Informal Communications, *Proceedings of CSCW 90*, Los Angeles, 1990.
5. Ishii, H., Miyake, N., Towards an Open Shared Workspace: Computer and Video Fusion Approach of TeamWorkstation, *Communications of the ACM*, V34/12, December 1991
6. Ishii, H., Kobayashi, M., Grudin, J., Integration of Interpersonal Space and Shared Workspace: ClearBoard design and experiments, *ACM Transactions on Information Systems*, 11(4), October 1993, pp 349-375.
7. Krueger, M.W., *Artificial Reality II*, Reading MA, Addison Wesley, 1991
8. Landy, J, and Myers, B., Sketching Interfaces: Towards more human interface design, *Computer*, 34(3), pp 56-64, 2001
9. Masoodian, M., Apperley, M. & Frederikson, L., Video support for shared work-space interaction: an empirical study. *Interacting with Computers*, 7(3), 237-253, 1995
10. Morikawa, O., Maesako, T., HyperMirror: Toward Pleasant-to-use Video Mediated Communication System, *Proceedings of CSCW 98*, Seattle, Washingon, November 1998.
11. Tang, John C., Minneman, Scott L., VideoWhiteBoard: Video Shadows to Support Remote Collaboration, TANG, J., AND MINEMAN, S. *Proceedings of ACM* SIGCHI, 1991, pp 315-322