IN Proceedings of the First International Symposium on Mixed Reality (ISMR '99). Mixed Reality – Merging Real and Virtual Worlds, pp. 261-284. Berlin: Springer Verlag.

Collaborative Mixed Reality

Mark Billinghurst, Hirokazu Kato
Human Interface Technology Laboratory
University of Washington
Box 352-142
Seattle, WA 98195, USA
{grof, kato}@hitl.washington.edu

Abstract

Virtual Reality (VR) appears a natural medium for computer supported collaborative work (CSCW). However immersive Virtual Reality separates the user from the real world and their traditional tools. An alternative approach is through Mixed Reality (MR), the overlaying of virtual objects on the real world. This allows users to see each other and the real world at the same time as the virtual images, facilitating a high bandwidth of communication between users and intuitive manipulation of the virtual information. We review MR techniques for developing CSCW interfaces and describe lessons learned from developing a variety of collaborative Mixed Reality interfaces. Our recent work involves the use of computer vision techniques for accurate MR registration. We describe this and identify areas for future research.

Keywords: Mixed Reality, Augmented Reality, Virtual Reality, Computer Supported Collaborative Work

1 Introduction

Computers are increasingly used to enhance collaboration between people. As collaborative tools become more common the Human-Computer Interface is giving way to a Human-Human Interface mediated by computers. This emphasis on collaboration adds new technical challenges to the design of Human Computer Interfaces. There are also many social factors that must be addressed before collaborative tools will become common in the workplace.

These problems are compounded for attempts to support three-dimensional Computer Supported Collaborative Work (CSCW). Although the use of spatial cues and three-dimensional object manipulation are common in face to face communication, tools for three-dimensional CSCW are still rare. However new interface metaphors may overcome this limtation. In this paper we describe how Mixed Reality techniques can be used to enhance remote and face to face collaboration, particularly 3D CSCW.

Mixed Reality (MR) environments are defined by Milgram as those in which real world and virtual world objects are presented together on a single display [27]. Single user Mixed Reality interfaces have been developed for computer aided instruction [9], manufacturing [7] and medical visualization [2]. These applications have shown that Mixed Reality interfaces can enable a person to interact with the real world in ways never before possible. For example, the work of Bajura et. al. overlays virtual ultrasound images onto a patients body, allowing doctors to have "X-Ray" vision while performing a needle biopsy task [2].

Although Mixed Reality techniques have proven valuable in single user applications, there has been less research on collaborative applications. We believe that Mixed Reality is ideal for collaborative interfaces because it addresses two major issues in CSCW: *seamlessness* and *enhancing reality*. In the next section we describe these issues in depth. We then review approaches for 3D CSCW and describe examples from our work and others of how Mixed Reality can be used to support local and remote collaboration. Finally we conclude with a description of new computer vision techniques for collaborative Mixed Reality interfaces.

2 Motivation: Why Collaborative Mixed Reality

2.1 Seamless Computer Supported Collaborative Work

When people talk to one another in a face-to-face conversation while collaborating on a real world task there is a dynamic and easy interchange of focus between the shared workspace and the speakers' interpersonal space. The shared workspace is the common task area between collaborators, while the interpersonal space is the common communications space. In face-to-face conversation the shared workspace is often a subset of the interpersonal space, so there is a dynamic and easy change of focus between spaces using a variety of non-verbal cues. For example, if architects are seated around a table with house plans on it, it is easy for them to look at the plans while simultaneously be aware of the conversational cues of the other people.

In most existing CSCW tools this is not the case. Current CSCW interfaces often introduce seams and discontinuities into the collaborative workspace. Ishii defines a seam as a spatial, temporal or functional constraint that forces the user to shift among a variety of spaces or modes of operation [18]. For example, the seam between computer word processing and traditional pen and paper makes it difficult to produce digital copies of handwritten documents without a translation step. Seams can be of two types:

- *Functional Seams:* Discontinuities between different functional workspaces, forcing the user to change modes of operation.
- *Cognitive Seams:* Discontinuities between existing and new work practices, forcing the user to learn new ways of working.

One of the most important functional seams is that between shared and interpersonal workspaces. However, most CSCW systems have an arbitrary seam between the shared workspace and interpersonal space; for example, that between a shared white board and a video window showing a collaborator (Figure 1). This prevents users who are looking at the shared white board from maintaining eye contact with their collaborators, an important non-verbal cue for conversation flow [20].

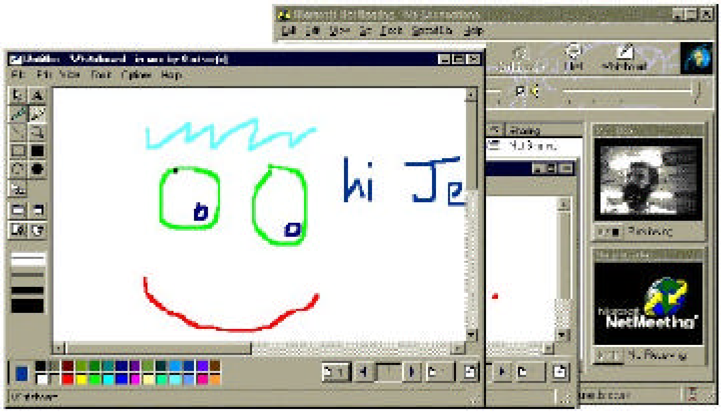

Figure 1. The functional seam between a shared whiteboard and video window.

A common cognitive seam is that between computer-based and traditional desktop tools. This seam causes the learning curve experienced by users who move from physical tools to their digital equivalents, such as the painter moving from oils to digital tools. Grudin [12] points out that CSCW tools are generally rejected when they force users to change the way they work; yet this is exactly what happens when collaborative interfaces make it difficult to use traditional tools in conjunction with the computer-based tools.

Functional and cognitive seams in collaborative interfaces changes the nature of collaboration and produces communication behaviors that are different from face-to-face conversation. For example, even with no video delay, video-mediated conversation doesn't produce the same conversation style as face-to-face interaction [13]. This occurs because video cannot adequately convey the non-verbal signals so vital in face-to-face communication introducing a functional seam between the participants [18]. Thus, Sharing the same physical space positively affects conversation in ways that is difficult to duplicate by remote means.

Ishii et. al. showed that it is possible to build seamless collaborative interfaces through their work on the *TeamWorkStation* and *ClearBoard* projects [16][17]. *TeamWorkStation* reduces the seam between the real world and collaborative workspace by combining video and computer based tools. Video overlay on the collaborative workspace allows collaborators to use pencil and paper together with their computer-based tools in remote collaboration. However there is still a seam between the collaborative workspace and video images of the collaborators. The *ClearBoard* series of interfaces address the seams between the individual and shared workspace. The interfaces are based on the metaphor of users drawing on a glass between them. By using surfaces made of large mirrors, and video projection techniques, users can look directly at their workspace and see their collaborator directly behind it (fig. 2). This removes the seam between the interpersonal and shared workspace. Users can effectively and easily change focus, maintain eye contact, and use gaze awareness in collaboration; the result is an increased feeling of intimacy and copresence.

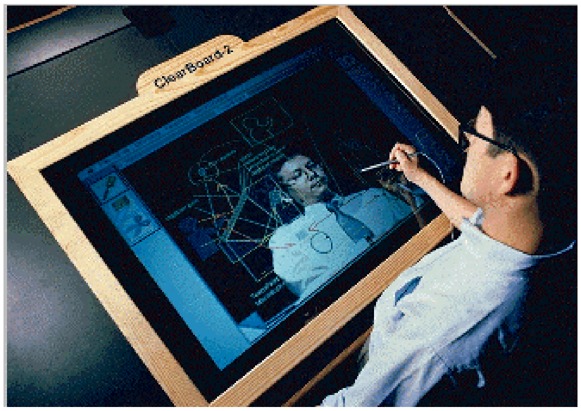

Figure 2. The ClearBoard seamless interface
(Image courtesy of H. Ishii, MIT Media Laboratory).

These projects show that in order for an interface to minimize functional and cognitive seams it should have the following characteristics:

- It must support existing tools and work techniques;
- Users must be able to bring real-world objects into the interface;
- The shared workspace must be a subset of the interpersonal space;
- There must be audio and visual communication between participants;
- Collaborators must be able to maintain eye contact and gaze awareness.

2.2 Enhancing Reality

Removing the seams in a collaborative interface is not enough. As Hollan and Stornetta point out [15], CSCW interfaces may not be used if they provide the same experience as face-to-face communication; they must enable users to go "beyond being there" and enhance the collaborative experience. When this is not the case, users will often stop using the interface or use it differently that what it was intended for. For example, studies of use of the Cruiser video conferencing system between users in the same building found that most people used the system for brief conversations and setting up face-to-face collaboration rather than for replacing face-to-face meetings [10].

The motivation for going "beyond being there" can be found by considering past approaches to CSCW. Traditional CSCW research attempts to use computer and audio-visual equipment to provide a sense of remote presence. Measures of social presence [38] and information richness [8] have been developed to characterize how closely CSCW tools capture the essence of face-to-face communication. The hope is that collaborative interfaces will eventually be indistinguishable from actually being there.

Hollan and Stornetta suggest this is the wrong approach. Considering face-to-face interaction as a specific type of communications medium, it becomes apparent that this approach requires one medium to adapt to another, pitting the strengths of face-to-face collaboration against other interfaces. Mechanisms that are effective in face-to-face interactions may be awkward if they are replicated in an electronic medium, often making users reluctant to use the new medium. In fact, because of the nature of the medium, it may be impossible for mediated collaborations to provide the same experience as face-to-face collaboration [11].

Hollan and Stornetta argue that a better way to develop interfaces for telecommunication is to focus on the *communication* aspect, not the *tele-* part. Rather than using new media to imitate face-to-face collaboration, researchers should be considering what new attributes the media can offer that satisfy the needs of communication so well that people will use it regardless of physical proximity. So one way to develop effective collaborative interfaces is to identify unmet needs in face-to-face conversation and create interface attributes that address these needs.

In this section we have described the need for collaborative interfaces to support seamless interaction and enable collaborators to go beyond being there. Mixed Reality interfaces are ideal for CSCW because they meet both these needs. We demonstrate this in the next section by reviewing other types of interfaces for three-dimensional CSCW, and comparing them to a Mixed Reality approach.

3. Collaborative Interfaces for Three Dimensional CSCW.

There are several different approaches for facilitating three-dimensional collaborative work. The most obvious is adding collaborative capability to existing screen-based three-dimensional packages. However a two-dimensional interface for three-dimensional collaboration can have severe limitations. For example, Li-Shu [24] developed a workstation based collaborative CAD package but users found it difficult to visualize the different viewpoints of the collaborators making communication difficult. Communication was also restricted to voice and pointing with a graphical icon, further compounding the problem.

Alternative techniques include using large parabolic stereo projection screens or holographic optical systems to project a three-dimensional virtual image into space. CAVE-like systems [7] and the responsive workbench [22] allow a number of users to view stereoscopic 3D images by wearing LCD-shutter glasses. These images are projected on multiple large screen projection walls in the case of the CAVE, or a large opaque tabletop display for the responsive workbench. Unfortunately in both cases the images can be rendered from only a single user's viewpoint, so only one person will see true stereo. This makes it impossible for users to surround the Responsive Workbench table, or to spread themselves throughout the CAVE and see the correct stereoscopic image. The devices are also need bulky hardware such as a projection screen or large beam splitter, are not portable and require expensive optics.

Mechanical devices can also be used to create volumetric displays. These include scanning lasers onto a rotating helix to created a three-dimensional volumetric display [39] or using a rotating phosphor coated plate activated with electron guns [4]. These devices are also not portable, do not permit remote collaboration, and do not allow direct interaction with the images because of the rotating display surface.

Multi-user immersive virtual environments provide an extremely natural medium for three dimensional CSCW; in this setting computers can provide the same type of collaborative information that people have in face-to-face interactions, such as communication by object manipulation and gesture [41]. Work on the DIVE project [5], GreenSpace [25] and other fully immersive multi-participant virtual environments has shown that collaborative work is indeed intuitive in such surroundings. Gesture, voice and graphical information can all be communicated seamlessly between the participants. However most current multi-user VR systems are fully immersive, separating the user from the real world.

3.1 Collaborative Mixed Reality

Unlike the other methods for three-dimensional CSCW, Mixed Reality interfaces can overlay graphics and audio onto the real world. This allows the creation of MR interfaces that combine the advantages of both virtual environments and seamless collaboration. Information overlay may be used by remote collaborators to annotate the user's view, or may enhance face-to-face conversation by producing shared interactive virtual models. In this way Mixed Reality techniques can be used to enhance communication regardless of proximity. Thus the use of Mixed Reality facilitates the development of collaborative interfaces that go "beyond being there". Mixed Reality also supports seamless collaboration with the real world, reducing the functional and cognitive seams between participants. These attributes imply that Mixed Reality approaches would be ideal for many CSCW applications.

Despite this, there are few examples of multi-user Mixed Reality systems. Amselen [1] and Rekimoto [33] have explored the use of tracked hand held LCD displays in a multi-user environment. Amselen uses LCD panels as portable windows into a shared multi-user immersive environment, while Rekimoto attaches small cameras to LCD panels to allow virtual objects to be composited on video images of the real world. These displays have the advantage that they are small, light weight, portable and higher resolution than head mounted displays. Unfortunately they do not support a true stereoscopic view, and are not hands free. Users must also hold the LCD panel in front of their face - obscuring their facial expressions.

Klaus et. al. [19] also use video compositing techniques to superimpose virtual image over a real world view. Their system is also multi-user, but is monitor and workstation based so users get the impression that the virtual objects are superimposed on a remote real environment rather than their local environment.

Their architecture is designed to support distributed users viewing the same real environment remotely rather than local users interacting in the same real environment.

Unlike these systems, our approach is to use see-through head mounted displays with head and body tracking in a collaborative interface. These types of Mixed Reality interfaces allow multiple users in the same location or remote to work in both the real and virtual world simultaneously, facilitating CSCW in a seamless manner. This approach is most closely related to that of Schmalsteig et. al. [34]. They use see-through head mounted displays to allow users to collaboratively view 3D models of scientific data superimposed on the real world. They report users finding the interface very intuitive and conducive to real world collaboration because the groupware support can be kept simple and mostly left to social protocols. The AR^2 Hockey work of Ohshima et. al. [30] is also very similar. In this case two users wear see-through head mounted displays to play a Mixed Reality version of the classic game of air hockey. Like Schmalsteig, they report that users are able to naturally interact with each other and collaborate on a real world task.

From their work Schmalsteig et al. [34] identify five key advantages of collaborative MR environments:

- *Virtuality:* Objects that don't exist in the real world can be viewed and examined.
- *Augmentation:* Real objects can be augmented by virtual annotations.
- *Cooperation:* Multiple users can see each other and cooperate in a natural way.
- *Independence:* Each user controls his own independent viewpoint.
- *Individuality:* Displayed data can be different for each viewer.

Compared to immersive virtual environments, MR interfaces allow users to refer to notes, diagrams, books and other real objects while viewing virtual images, and they can use familiar real world tools to interact with the images, increasing the intuitiveness of the interface. More importantly, users can see each other's facial expressions, gestures and body language, increasing the communication bandwidth. Finally the entire environment doesn't need to be modeled, considerably reducing the graphics rendering requirements.

4 OUR WORK

In this section we present two of our prototype MR interfaces:

WearCom: An interface for multi-party conferencing that enables a user to see remote collaborators as virtual avatars surrounding them in real space. Spatial cues help overcome some of the limitations of current multiparty conferencing systems.

Collaborative Web Space: An interface which allows people in the same location to view and interact with virtual world wide web pages floating about them in space. Users can collaboratively browse the web while seeing the real world and use natural communication to talk about the pages they're viewing.

MR interfaces can be distinguished by how they present information in the information space. In a MR interface with a head mounted display, information can be presented in a combination of three ways:

- *Head-stabilized* **-** information is fixed to the user's viewpoint and doesn't change as the user changes viewpoint orientation or position.
- *Body-stabilized* **-** information is fixed relative to the user's body position and varies as the user changes viewpoint orientation, but not position.
- *World-stabilized* **-** information is fixed to real world locations and varies as the user changes viewpoint orientation and position.

Each of these methods require increasingly complex head tracking technologies; no head tracking is required for head-stabilized information, viewpoint orientation tracking is needed for body-stabilized information, while position and orientation tracking is required for world-stabilized. The registration requirements also become more difficult; none are required for head-stabilized images, while complex calibration techniques are required for world stabilization.

Body- and world-stabilized information display is attractive for a number of reasons. A body-stabilized information space can overcome the resolution limitations of head mounted displays. For example, Reichlen [31] tracks only head orientation to give a user a "hundred million pixel" hemispherical information surround. World-stabilized information allows annotating the real world with context dependent data and creating information enriched environments [32]. Spatial information displays enable humans to use their innate spatial abilities to retrieve and localise information and to aid performance.

In a Mixed Reality setting, spatial information display can be used to overcome the resolution and field of view limitations of the HMD and provide information overlay on the surrounding environment. This is important because the information presented in a MR interface is often intimately linked to the user's real world location and task. The prototype *WearCom* and the Collaborative Web Space interfaces both use body-stabilized information, however at the end of this chapter we present some more recent applications which use a world-stabilized information display.

4.1 Mixed Reality Interfaces for Remote Collaboration

We can use previous research in teleconferencing and CSCW interfaces to suggests attributes of the ideal collaborative MR interface. Research on the roles of audio and visual cues in teleconferencing has produced mixed results. There have been many experiments conducted comparing face-to-face, audio-and-video, and audio-only communication conditions, as summarized by Sellen [36]. While people generally do not prefer the audio-only condition, they are often able to perform tasks as effectively or almost as effectively as in the face-to-face or video conditions, suggesting that speech is the critical medium [42].

Based on these results, it may be thought that audio alone should be suitable for a creating a shared communication space. However attempts to build audio-only communication spaces, such as Thunderwire [14], have found that while audio can be sufficient for a usable communication space, there are several shortcomings. Users are not able to easily tell who else is present and they can't use visual cues to determine other's willingness to interact and discriminate between speakers. These problems suggest that while audio-only may be useful for small group interactions, it is less usable the more people present.

These shortcomings can be overcome through the use of visual and spatial cues. In face-to-face conversation, speech, gesture, body language and other non-verbal cues combine to show attention and interest. Visual cues are present in videoconferencing applications, however the absence of spatial cues in most video conferencing systems means that users often find it difficult to know when people are paying attention to them, to hold side conversations, and to establish eye contact [37].

Virtual reality can provide an alternative medium that allows groups of people to share the same communications space. In collaborative virtual environments (CVEs) spatial visual and audio cues can combine in natural ways to aid communication. Users can freely move through the space setting their own viewpoints and spatial relationships; enabling crowds of people to inhabit the same virtual environment and interact in a way impossible in traditional video or audio conferencing [3]. The well known "cocktail-party" effect shows that people can easily monitor several spatialized audio streams at once, selectively focusing on those of interest [35]. Even a simple virtual avatar representation and spatial audio model enables users to discriminate between multiple speakers [28].

These results suggest that an ideal MR interface for remote collaboration should have high quality audio communication, visual representations of the collaborators and an underlying spatial metaphor.

4.1.1 WearCom

The WearCom project explores how wearable computers can be used to support remote collaboration. Wearable computers are the most recent generation of portable machines. Worn on the body, they provide constant access to computing and communications resources. In general, a wearable computer may be defined as a computer that is subsumed into the personal space of the user, controlled by the wearer and has both operational and interactional constancy, i.e. is always on and always accessible [26]. Wearables are typically composed of a belt or back pack PC, see-though or see-around head mounted display (HMD), wireless communications hardware and an input device such as touchpad or chording keyboard. The use of a see-through display means that wearable computers are an ideal platform for portable MR interfaces.

Many of the target application areas for wearable computers are those where the user could benefit from expert assistance, such as vehicle maintenance or emergency response. Network enabled wearable computers can be used as a communications device to enable remote experts to collaborate with the wearable user. In such situations the presence of remote experts have been found to significantly improve task performance [21]. The question WearCom addresses is how a portable MR interface can be used to support collaboration between multiple remote people. This is becoming increasingly important as telephones incorporate more computing power and portable computers become more like telephones.

In WearCom we use the simplest form of body-stabilized display; one with one degree of orientation to give the user the impression they are surrounded by a virtual cylinder of visual and auditory information (figure 6). We track the user's head orientation so as they as look around they can see different portions of the information space. The cylindrical display is very natural to use since most head and body motion is about the vertical axis, making it very difficult for the user to become disoriented. With this display configuration a Mixed Reality conferencing space could be created that allows remote collaborators to appear as virtual avatars distributed about the user (Figure 7).

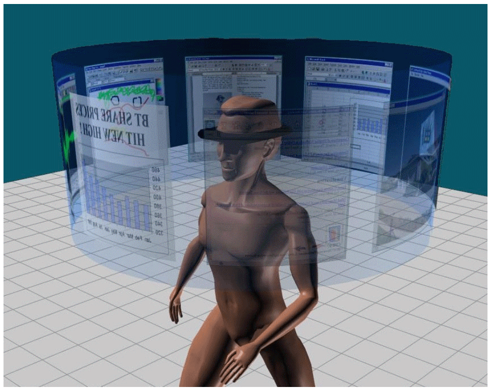

Figure 6. Our Body-Stabilized Display

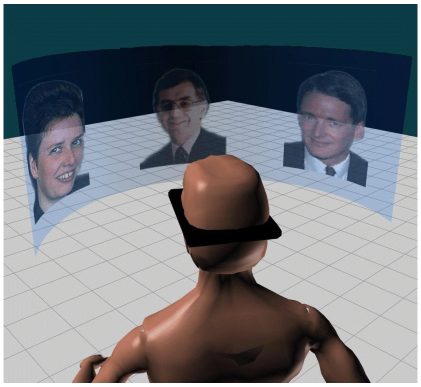

Figure 7. A MR Spatial Conferencing Space.

A wearable computer can be used to provide spatialized 3D graphics and audio cues to aid communication. The result is an augmented reality communication space with audio enabled avatars of the remote collaborators surrounding the user. The user can use natural head motions to attend to the remote collaborators, can communicate freely while being aware of other side conversations and can move through the communication space. In this way the conferencing space could support many simultaneous users. The user could also see the real world, enabling remote collaborators to help them with real world tasks.

The WearCom prototype implements the wearable communications described above. Our research is initially focused on collaboration between a single wearable computer user and several desktop PC users. The wearable computer we use is a custom built 586 PC 104 based computer with 20mb of RAM running Windows 95 (Figure 8.). A hand held Logitech wireless radio trackball with three buttons is used as the primary input device. The display is a pair of Virtual i-O iglasses! converted into a monoscopic display by the removal of the left eyepiece. This headmounted display can either be used in see-through or occluded mode, has a resolution of 262 by 230 pixels and a 26-degree field of view. They also have a sourceless two-axis inclinometer and a magnetometer used as a three degree of freedom orientation tracker. A BreezeCom wireless LAN is used to give 2mb/s Internet within 500 feet of a base station. The wearable also has a soundBlaster compatible sound board with headmounted microphone. The desktop PCs are standard Pentium class machines with Internet connectivity and sound capability.

The wearable computer has no graphics acceleration hardware and limited wireless bandwidth so the interface is deliberately kept simple. The conferencing space runs as a full screen application that is initially blank until remote users connect. When users join the conferencing space they are represented by blocks with static pictures of themselves on them (Figure 9). Although the resolution of the images is crude it is sufficient to identify who the speakers are and their spatial relationships. A radar display shows the location of the other users in the conferencing space, enabling users to find each other easily.

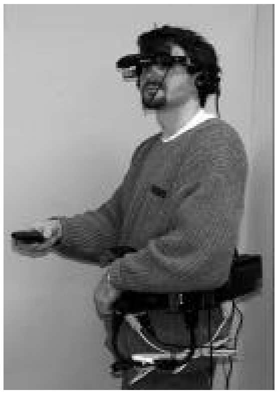

Figure 8. Wearable Hardware.

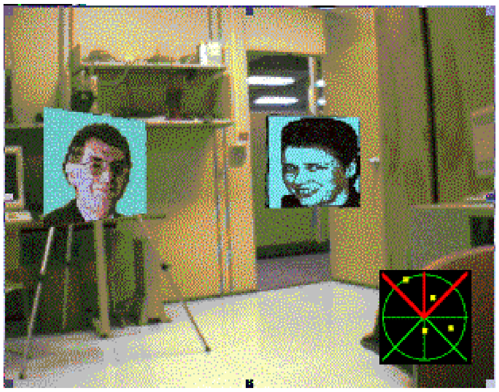

Figure 9. The User's View of the Conferencing Space.

Each user determines the position and orientation of their own avatar in space, which changes as they move or look about environment. The wearable user has their head tracked so they can simply turn to face the speakers they are interested in. Users can also navigate through the space; by rolling the trackball forwards or backwards their viewpoint is moved forwards or backwards along the direction they are looking. Since the virtual images are superimposed on the real world, when the user rolls the trackball it appears to they are moving the virtual space around them, rather than navigating through the space. Users are constrained to change viewpoint on the horizontal plane, just as in face-to-face conversations. The two different navigation methods (trackball motion, head tracking), match the different types of motion used in face to face communication; walking to join a join a group for conversation, and body orientation changes within a conversational group. The interface was developed using Microsoft's DirectX suite of libraries.

The wearable interface also supports 3D spatialized Internet telephony. When users connect to the conferencing space their audio is broadcast to all the other users in the space. This is spatialized according to the distance and direction between speaker and listener. As users face or move closer to different speakers the speaker volume changes due to the sound spatialisation. Since the speakers are constrained to remain in the same plane as the listener the audio spatialisation is considerably simplified. The conferencing space uses custom developed telephony libraries and the Microsoft DirectSound libraries.

Preliminary informal trials with WearCom have found that users are able to easily discriminate between three simultaneous speakers when their audio streams are spatialized, but not when non-spatialized audio is used. Participants preferred seeing a visual representation of their collaborators over just hearing their speech because it enabled them to see who is connected and the spatial relationship of the speakers. This allowed them to use some of the non-verbal cues commonly used in face-to-face communication such as gaze modulation and body motion. Lastly, users found that they could continue doing real world tasks while talking to collaborators in the conferencing space and it was possible to move the conferencing space with the trackball so that collaborators weren't blocking critical portions of the users field of view.

4.2 Co-Located Collaboration

Mixed Reality interfaces can enable co-located users to view and interact with shared virtual information spaces while viewing the real world at the same time. This preserves the rich communications bandwidth that humans enjoy in face-to-face meetings, while adding virtual images normally impossible to see. In this section we present a collaborative web browser that enables users to load and place virtual web pages around themselves in the real world and to jointly discuss and interact with them; users can see both the virtual web pages and each other, so communication is natural and intuitive.

We have developed a three-dimensional Web browser that enables multiple co-located users to collaboratively browse the World Wide Web. Users see each other and virtual web pages floating in space around them (Figure 13). The effect is a body-centered information space that the user can easily and intuitively interact with (Figure 14). The Shared Space browser supports multiple users who can communicate about the web pages shown, using natural voice and gesture.

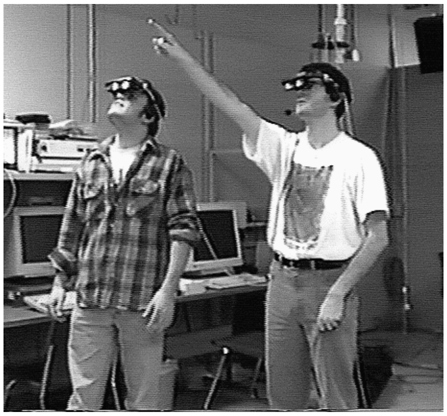

Figure 13 The Shared Space interface

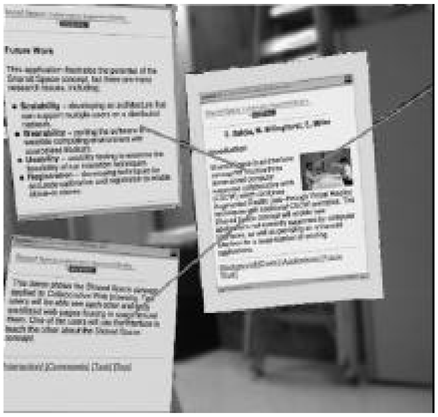

Figure 14 The Participant's View

Users wear see-through Virtual i-O stereoscopic headmounted displays, and head orientation is tracked using a Polhemus Fastrak electro-magnetic sensor. The interface is designed to be completely hands-free; pages are selected by looking at them and, once selected, can be attached to the user's viewpoint, zoomed in or out, iconified or expanded, or additional links loaded with voice commands. Speaker-independent continuous speech recognition software (Texas Instrument's DAGGER system) allows users to load and interact with web pages using vocal commands. To support this, an HMTL parser parses the web pages to extract their HTML links and assign numbers to them. In this way users can load new links with numerical commands such as "Load link one". Each time a new page is loaded a new browser object is created and a symbolic graphical link to its parent page is displayed to facilitate the visualisation of the web pages. The voice recognition software recognises 46 command phrases with greater than 90% accuracy. A switched microphone is used so participants can carry on normal conversation when not entering voice commands.

Two important aspects of the interface are gaze awareness and information privacy. Users need to know which page they are currently looking at as well as the pages their collaborators are looking at. This is especially difficult when there are multiple web pages close to each other. The Virtual I-O head mounted display has only a 26 degree field of view so it is tempting to overlap pages so that several can be seen at once. To address this problem, each web page highlights when a user looks at it. Each page also has gaze icons attached to it for each user that highlights to show which users are looking at the page. In this way users can tell where their collaborators are looking. When each web page is loaded it is initially visible only to the user that loaded it. Users can change page visibility from private to public with vocal commands; users can only see the public web pages and their own private objects.

The collaborative web interface uses a body-stabilized information space similar to that in the WearCom interface. However, in this case all three degrees of head orientation are used, providing a virtual sphere of information. Even though the head-mounted display has only a limited field of view, the ability to track head orientation and place objects at fixed locations relative to the body effectively creates a 360-degree circumambient display. Since the displays are wearable users can collaborate in any location, and because interface objects are not attached to real world locations the registration requirements are not as stringent.

The Mixed Reality interface facilitates a high bandwidth of communication between users as well as natural 3D manipulation of the virtual images. The key characteristic of this interface is the ability to see the real world and collaborators at the same time as the virtual web pages floating in space. This means that users can use natural speech and gesture to communicate with each other about the virtual information space. In informal trials, users found the interface intuitive and communication with the other participants seamless and natural. Collaboration could be left to normal social protocols rather than requiring mechanisms explicitly encoded in the interface. Unlike sharing a physical display, users with the wearable information space can restrict the ability of others to see information in their space. They were able to easily spatially organize web pages in a manner that facilitated rapid recall, and the distinction between public and private information was found to be useful for collaborative information presentation.

5 Computer Vision Methods for Collaborative Mixed Reality

In the previous sections we have described collaborative mixed reality interfaces which use magnetic or inertial trackers to create collaborative spaces. Wired position sensors such as these have been used effectively for immersive virtual reality systems. However, in many MR applications the need for high resolution, large scale position and orientation tracking make the use of these sensors impractical. Computer vision methods do not require any wired sensors and so play an important role in MR applications, particularly interfaces which use world-stabilized information display. In this section we describe our computer vision techniques for accurate world-stabilized image registration, and some collaborative applications that use these technique.

Computer vision techniques have previously been applied with great success in MR interfaces. A common approach is to use physical markers to aid with the registration. For example, Rekimoto proposes a registration method using square markers each with a unique 2D-matrix code [32]. Other methods using fiducial markers include those of State [40] and Kutulakos [23]. All these cases are video based MR systems in which virtual objects are superimposed on video images captured by the camera. This considerably simplifies the camera calibration requirements. In video-based Mixed Reality, only the relationship between camera coordinates and 3D world coordinates is needed. However, in optical see-through MR systems, stereoscopic views of virtual objects appear in the physical space, so the eye and HMD screen coordinate systems must also be known.

In a MR system using optical see-through HMD, stereo images have to be provided to both eyes. In order to do that, accurate measurement of position of the camera with respect to each eye is required. This is because the scene captured by a head-mounted camera is different from the view of each eye. This type of camera calibration is difficult and some systems such as AR^{2}Hockey [30] ignore the parallax between the camera and eyes entirely. However, one of the application areas we are interested in is a collaborative MR system for wearable computers. In this case, we can expect that the 3D space in which virtual objects appear will be close to the user. The parallax between the eyes and the camera therefore can not be ignored. In order to display a virtual object in a position close to the user, an accurate calibration is required.

5.1 Calibration method

We calibrate the head-mounted camera using the calibration frame shown in Figure 18. This is a simple cardboard frame with a ruled grid of lines on it that is attached to the front of the HMD as shown. By attaching the calibration frame to the head mounted display, the head position doesn’t change relative to the grid of lines, so we can find the relationship between the camera coordinate system and the screen coordinate system, by way of the calibration frame coordinate system. There are two critical transformations we need to find; that between the screen coordinate frame and calibration frame coordinate frame, and that between the calibration frame coordinate frame and camera coordinate frame. Figure 19 shows these transformations and the various coordinates systems used.

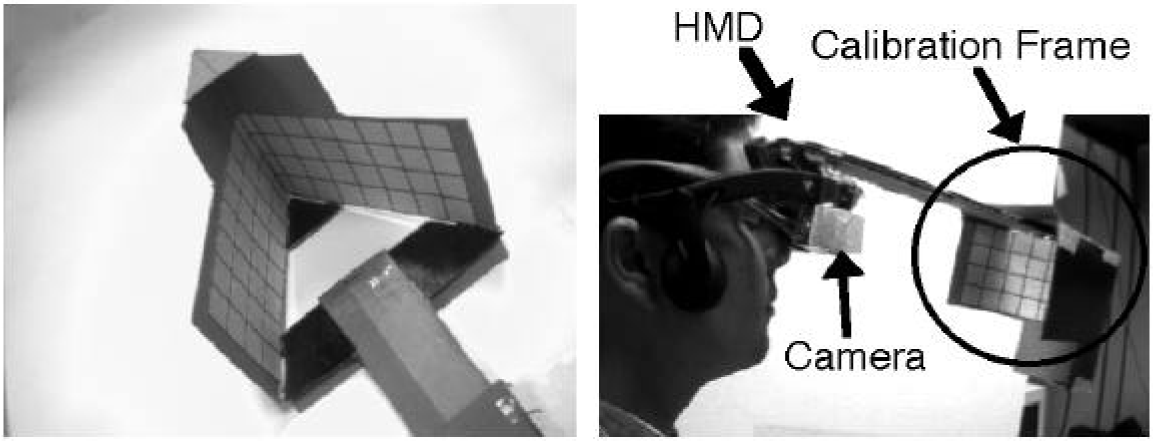

Figure 18. The Calibration Frame used in our Calibration Method.

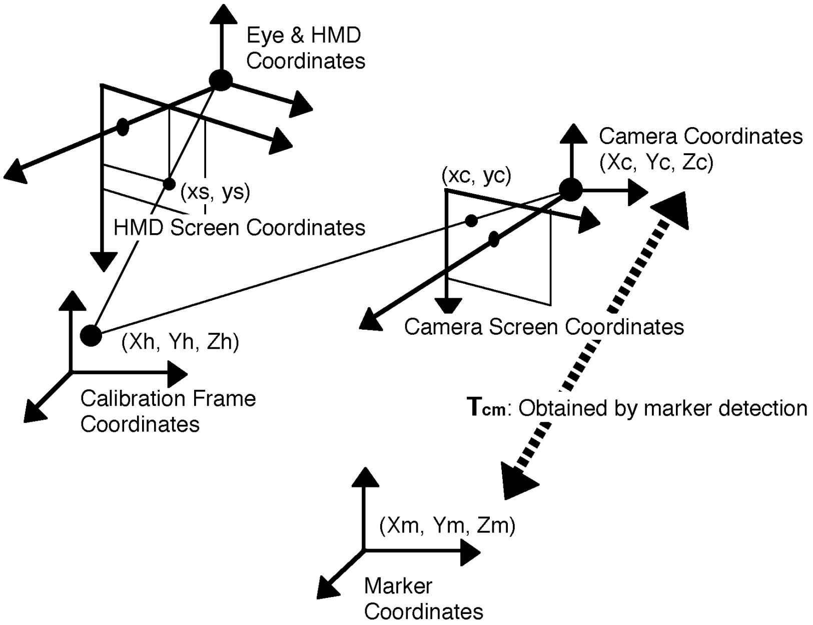
Eye & HMD Coordinates
(xs, ys)
HMD Screen Coordinates
Camera Coordinates (Xc, Yc, Zc)
(xc, yc)
Camera Screen Coordinates
(Xh, Yh, Zh)
Calibration Frame Coordinates
Tcm: Obtained by marker detection
(Xm, Ym, Zm)
Marker Coordinates

Figure 19: The coordinate systems used in our calibration procedure.

We assume all rays from a physical object reach the focal point of the eye through the HMD screen. Then the relationship between the HMD screen and the focal point of the eye makes a perspective camera model shown by the equation below. $\mathbf{T_{sh}}$ is the calibration frame to screen coordinate frame transformation matrix, while ($\mathbf{x_s}$, $\mathbf{y_s}$) is the screen coordinates of the point being transformed and ($\mathbf{X_h}$, $\mathbf{Y_h}$, $\mathbf{Z_h}$) the location of the same point expressed in the calibration frame coordinate frame. Using $\mathbf{T_{sh}}$ as the transformation between the virtual 3D space and screen coordinates enables the virtual 3D space to coincide with the physical 3D space on the HMD Screen.

$$\begin{bmatrix} hx_s \\ hy_s \\ h \\ 1 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{21} & H_{22} & H_{23} & H_{24} \\ H_{31} & H_{32} & H_{33} & H_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_h \\ Y_h \\ Z_h \\ 1 \end{bmatrix} = \mathbf{T_{sh}} \begin{bmatrix} X_h \\ Y_h \\ Z_h \\ 1 \end{bmatrix}$$

In a similar way, the relationship between the camera and calibration frame coordinate is:

$$\begin{bmatrix} hx_c \\ hy_c \\ h \\ 1 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 \\ C_{21} & C_{22} & C_{23} & 0 \\ C_{31} & C_{32} & C_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \mathbf{C} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_h \\ Y_h \\ Z_h \\ 1 \end{bmatrix} = \mathbf{CT_{ch}} \begin{bmatrix} X_h \\ Y_h \\ Z_h \\ 1 \end{bmatrix}$$

where C is a matrix representation of the inner camera parameters, and $\mathbf{T_{ch}}$ is the transformation matrix from calibration frame coordinates ($\mathbf{X_h}$, $\mathbf{Y_h}$, $\mathbf{Z_h}$) to Camera coordinates ($\mathbf{X_c}$, $\mathbf{Y_c}$, $\mathbf{Z_c}$).

The grid of the calibration frame is used to find the exact values of the matrices $\mathbf{T_{sh}}$,**C** and $\mathbf{T_{ch}}$. When a user wears this frame they see the view shown in figure 20. To calibrate the HMD, while wearing the calibration frame the user fits virtual lines drawn on the HMD screen to the corresponding line segments on the physical calibration tool. The virtual lines can be moved and rotated by keyboard operation. The positions of all of the intersections of the real line segments are known in the calibration frame coordinate system. This user operation finds the corresponding positions in the HMD screen coordinate system. By using this data, the transformation matrix $\mathbf{T_{sh}}$ can be estimated. The user carries out this process for each of eyes, generating two matrices. The resultant $\mathbf{T_{sh}}$ matrices are used as the transformation between the virtual 3D coordinate system and the HMD screen coordinate system. To find the camera calibration matrices, **C** and

$\mathbf{T_{ch}}$, the same process is used, however in this case video from the camera is displayed on a computer monitor and the user aligns the virtual lines on-screen.

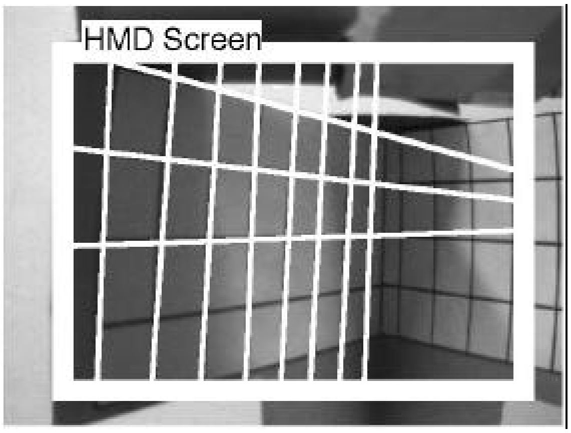

Figure 20. Simulated view of the image the user sees in the calibration frame.

5.2 Registration Method

Once the camera and eye transformation matrices have been found they can be used to register virtual images with the real world. In order to do this we use square fiducial markers attached to real world objects. These markers are detected using basic image processing techniques and virtual objects are drawn relative to the marker coordinates system.

Displaying a virtual object on markers in the physical 3D space requires that $\mathbf{T_{hm}}$, the relationship between the marker coordinate system and calibration frame coordinate system is known. If this is known then transformation between the marker coordinate system and screen coordinates can be found from $\mathbf{T_{sm}} = \mathbf{T_{sh}}.\mathbf{T_{hm}}$. In figure 19 the matrix $\mathbf{T_{cm}}$ represents the transformation between the marker coordinate system and the camera coordinate system. This transformation can be estimated by image analysis, using the same computer vision methods that Rekimoto [32] and others use to identify fiducial location in MR systems. The transformation matrix $\mathbf{T_{ch}}$, representing the relationship between the camera and the calibration frame coordinate systems, is found from the camera calibration, so the transformation matrix from marker coordinate system to the calibration frame coordinate system can be calculated easily from the following:

$$\mathbf{T_{hm}} = \mathbf{T_{ch}}^{-1} \cdot \mathbf{T_{cm}}$$

The perspective transformation matrix from calibration frame coordinate system to HMD screen coordinates system, $\mathbf{T_{sh}}$, is known from the system calibration. So the screen coordinates ($\mathbf{x_s}$, $\mathbf{y_s}$) of the virtual image that is to appear at the physical coordinates of ($\mathbf{X_m}$, $\mathbf{Y_m}$, $\mathbf{Z_m}$) in the marker coordinate frame can be found from the following equation:

$$\begin{bmatrix} hx_s \\ hy_s \\ h \\ 1 \end{bmatrix} = \mathbf{T_{sh}} \cdot \mathbf{T_{ch}}^{-1} \cdot \mathbf{T_{cm}} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix}$$

The above matrix representation is suitable for the OpenGL graphics libraries we use for developing our applications. OpenGL uses both a modelview matrix and a projection matrix to render 3D graphics, so $\mathbf{T_{sh}} \cdot \mathbf{T_{ch}}^{-1}$ can be regarded as a projection matrix and $\mathbf{T_{cm}}$ as the modelview matrix. The matrix $\mathbf{T_{sh}} \cdot \mathbf{T_{ch}}^{-1}$ does not change after the calibration so it only needs to be calculated once. However, the transformation $\mathbf{T_{cm}}$ changes each time the user moves their head (and camera) position. So a virtual object can be accurately registered in the marker coordinate system by continuously updating the modelview matrix for $\mathbf{T_{cm}}$.

5.3 Examples of vision based collaborative interfaces

The calibration and registration methods we describe above allow us to developed collaborative Mixed Reality interfaces that use world-stabilized images. In this section we briefly describe two applications that are currently under development. Many other world stabilized MR collaborative applications are possible.

Our first application is a prototype of a video conferencing system for wearable computers. This extends the WearCom work by allowing the virtual images of remote participants to be attached to real world locations. In this case unique fiducial markers are used to represent each of the people a user may wish to call. When the user places a marker in view of the head mounted camera the system initiates an audio and visual connection to the remote user and attaches a virtual representation of them to the marker. Figure 21a shows the view through the user's head mounted display when they are conferencing with two other people. The user can arrange the layout of participants as they want and can continue to do a real world task without the interference of video display.

The interesting aspect of this approach is that it is the opposite of traditional video conferencing. Our goal is to put a virtual representation of the remote user into the local user's real world location, enabling them to have a videoconference regardless of where they are. In contrast, current video conferencing requires the user to move to a desktop computer or videoconferencing suite, often removing them from their workplace. As with WearCom, the interface also restores the spatial cues lost in traditional videoconferencing. However, having remote users represented as objects in the physical environment means that their virtual avatars can also gesture and interact visually with other objects in the user's space. This is potentially a powerful new interaction technique for collaborative MR interfaces.

A second area we are looking at is how our vision technique can be used to aid co-located MR collaboration on a wearable computer platform. For wearable computers, there are many problems remaining with the input interface. We have made a prototype of input interface using a paper tablet and a pen that enables a user to draw virtual lines on the tablet surface. Figure 21b shows a tablet with 6 fiducial markers and virtual annotations aligned with the markers. The user can take this interface anywhere, and more importantly other people with them that have wearable computers can also see the virtual annotations, so this can be used as a powerful tool for co-located collaboration. Since some markers may be occluded by a hand or a pen and missed by the camera, the 3D position of the tablet is estimated from whatever markers are visible. This ensures robust tracking of the tablet interface.

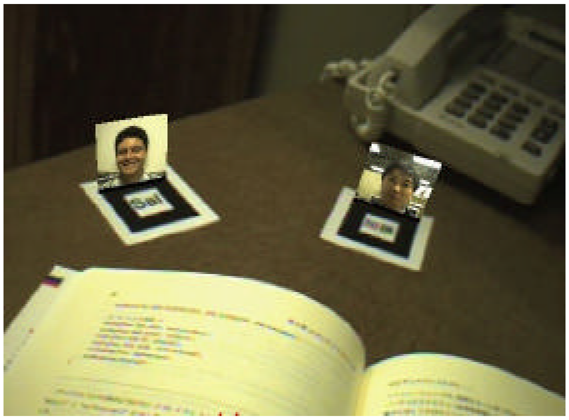

Fig. 21a. MR World Stabilized Video Conferencing

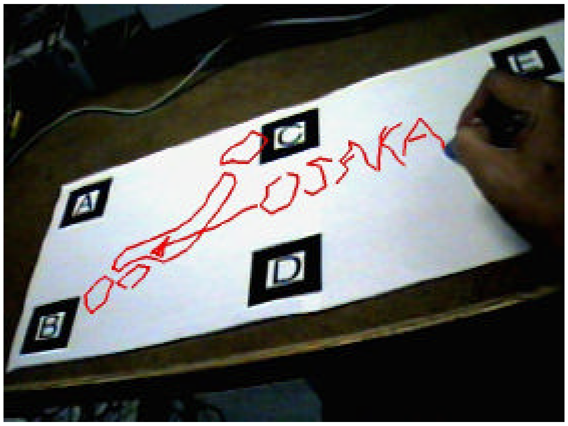

Fig. 21b. MR Tablet Writing Interface

6 Conclusions

Mixed Reality interfaces enable the development of innovative CSCW applications that are seamless and enhance face-to-face and remote collaboration. They are seamless because they allow the users to use traditional tools and workplace practices while overlaying virtual images onto the real world. Thus MR interfaces enhance the real world rather than replacing it entirely as do immersive VR environments. These MR enhancements can be used to support face-to-face and remote collaboration in ways otherwise impossible, enabling users to go "beyond being there".

In this chapter we have shown several examples of collaborative body and world-stabilized information spaces. User experiences with these interfaces have shown that they facilitate collaboration in a natural manner, enabling people to use normal gestures and non-verbal behavior in face-to-face collaboration, and to have access to their traditional tools and workplace in both face-to-face and remote collaboration.

Although these results are promising, they just scratch the surface of possible applications and there are several important research directions that need to be addressed in the future. First, rigorous user studies must be conducted to identify the unique characteristics of collaborative MR interfaces. The few papers that have been published in the field have shown some of the possible benefits of using these types of collaborative tools, but these benefits will only be realized when applications are developed based on solid interface design principles and user studies. These studies should compare user performance in MR interfaces with the equivalent immersive VR interfaces to examine how the seamlessness between the real and virtual world affects performance. They should also measure the effects of registration errors and system latency on collaborative performance, and provide guidelines for the acceptable latency for a range of tasks. Many of these types of studies have been conducted for single user MR interfaces and need to be replicated for collaborative applications.

A second area for future work is exploration into the types of unique interface metaphors that collaborative MR interfaces make possible. MR interfaces allow users to use real world objects to interact with virtual images, and enhance existing real world objects. However it remains an unanswered question as how to best use these capabilities. In order to explore this there needs to be better supporting technologies developed such as improved techniques for image registration and object and body tracking.

Finally, promising application areas need to be identified. One possibility is applications in the field of wearable computing. This is an area where MR head-mounted displays are commonly used and current user interface needs are not being well met. The traditional WIMP interface is not appropriate for the wearable computing platform, both because of the inherent assumptions that it makes about the input and output devices and the nature of the tasks that wearable computers are being used for. However the ability of MR interfaces to enhance rather than replace the users real world task suggest that MR techniques could be used to develop a more appropriate interface metaphor. Wearable MR interfaces for collaboration is a largely unexplored field that promises to change the way people collaborate with wearable computers.

7 References

1. Amselen, D. A Window on Shared Virtual Environments. *Presence,* 1995, Vol. 4(2), pp. 130-145.

2. Bajura, M., Fuchs, H., Ohbuchi, R. Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery Within the Patient. In *Proceedings of SIGGRAPH '92,* 1992, New York: ACM Press, pp. 203-210.

3. Benford, S., Greenhalgh, C., Lloyd, D. Crowded Collaborative Virtual Environments. In *Proceedings of CHI '97*, Atlanta, Georgia. March 1997, New York: ACM Press, pp.59-66

4. Blundell, B.G., and Schwarz, A.J. A Graphics Hierarchy for the Visualization of 3D Images by Means of a Volumetric Display System. In *Proceedings of the IEEE Region 10's Ninth Annual International Conference,* Singapore, Aug. 22-26, 1994, pp. 1-5. Vol. 1. IEEE New York, NY.

5. Carlson, C., and Hagsand, O. (1993) DIVE - A Platform for Multi-User Virtual Environments. Computers and Graphics. Nov/Dec 1993, Vol. 17(6), pp. 663-669.

6. Caudell, T.P., and Mizell, D.W. Augmented Reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on Systems Science,* Kauai, Hawaii, 7th-10th Jan. 1992, Vol. 2, pp. 659-669.

7. Cruz-Neira, C., Sandin, D. J., Defanti, T. A., Kentyon, R. V., and Hart, J. C. The CAVE: Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM,* 1992, Vol. 35 (6), pp. 65.

8. Draft, R.L., Lengel, R.H. Organizational Information requirements, media richness, and structural design. *Management Science*, Vol. 32, 1991, pp. 554-571.

9. Feiner, S., MacIntyre, B., and Seligmann, D. Knowledge-Based Augmented Reality. *Communications of the ACM*, Vol. 36(7), 1993, pp. 53-62.

10. Fish, R.S., Kraut, R.E., Root, R.W., Rice, R. *Evaluating Video as a Technology for Informal Communication.* Bellcore Technical Memorandum, TM-ARH017505, 1991.

11. Gaver, W. 1992. The Affordances of Media Spaces for Collaboration, in *Proc. CSCW'92*, Toronto, Canada, Oct.31-Nov.4, 1992, New York: ACM Press, pp. 17-24.

12. Grudin, J. Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. In *Proceedings of CSCW '88*, Portland, Oregon, 1988, New York: ACM Press, pp. 85-93.

13. Heath, C., Luff, P. Disembodied Conduct: Communication Through Video in a Multimedia Environment. In *Proceedings of CHI '91,* 1991, New York, NY: ACM Press, pp. 99-103.

14. Hindus, D., Ackerman, M., Mainwaring, S., Starr, B. Thunderwire: A Field study of an Audio-Only Media Space. In *Proceedings of CSCW '96,* Nov. 16th -20th, 1996, New York, NY: ACM Press.

15. Hollan, J., Stornetta, S. Beyond Being There. In *Proceedings of CHI '92*, 1992, New York: ACM Press, pp.119-125.

16. Ishii, H., Miyake, N., Toward an Open WorkSpace: Computer and Video Fusion Approach of TeamWorksation. *Communications of the ACM*, Dec 1991, Vol 34, No. 12, pp. 37-50.

17. Ishii, H., Kobayashi, M., Grudin, J., Integration of Inter-Personal Space and Shared Workspace: ClearBoard Design and Experiments. In *Proceedings of CSCW '92*, 1992, pp.33-42.

18. Ishii, H., Kobayashi, M., Arita, K., Iterative Design of Seamless Collaboration Media. *Communications of the ACM*, Vol 37, No. 8, August 1994, pp. 83-97.

19. Klaus, A., Kramer, A., Breen, D., Chevalier, P., Crampton, C., Rose, E., Tuceryan, M., Whitaker, R., Greer, D. (1995) Distributed Augmented Reality for Collaborative Design Applications. In *Proceedings of Eurographics '95*. pp. C-03-C-14, September 1995.

20. Kleinke, C.L. Gaze and eye contact: a research review. *Psychological Bulletin,* 1986, 100, pp.78-100.

21. Kraut, R., Miller, M., Siegal, J. Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication. In *Proceedings of CSCW '96*, Nov. 16th-20th, 1996, New York, NY: ACM Press.

22. Kruger, W., Bohn, C., Frohlich, B., Schuth, H., Strauss, W., Wesche, G. The Responsive Workbench: A Virtual Work Environment. *IEEE Computer,* Vol. 28(7), 1995, pp.42-48.

23. Kutulakos, K., Vallino, J. Calibration-Free Augmented Reality. *Transaction of visualization and computer graphics,* vol.4, No.1, 1998, pp.1-20.

24. Li-Shu, Flowers, W. Teledesign: Groupware User Experiments in Three-Dimensional Computer Aided Design. *Collaborative Computing,* Vol. 1(1), 1994, pp. 1-14.

25. Mandeville, J., Davidson, J., Campbell, D., Dahl, A., Schwartz, P., and Furness, T. A Shared Virtual Environment for Architectural Design Review. In *CVE '96 Workshop Proceedings,* 19-20th September 1996, Nottingham, Great Britain.

26. Mann, S. Smart Clothing: The Wearable Computer and WearCam. *Personal Technologies,* Vol. 1, No. 1, March 1997, Springer-Verlag.

27. Milgram, P., and Kishino, F. A taxonomy of mixed reality visual displays, *IEICE Transactions on Information and Systems*, *Special issue on Networked Reality,* Dec. 1994.

28. Nakanishi, H., Yoshida, C., Nishimura, T., Ishida, T. FreeWalk: Supporting Casual Meetings in a Network. In *Proceedings of CSCW '96*, Nov. 16th -20th, New York, NY: ACM Press, pp. 308-314.

29. O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., Bruce, V. Comparison of face-to-face and video-mediated interaction. *Interacting with Computers,* Vol. 8 No. 2, 1996, pp. 177-192.

30. Ohshima, T., Sato, K., Yamamoto, H., Tamura, H. AR^{2}Hockey: A case study of collaborative augmented reality, In *Proceedings of VRAIS'98*, 1998, IEEE Press: Los Alamitos, pp.268-295.

31. Reichlen, B. SparcChair: One Hundred Million Pixel Display. In *Proceedings IEEE VRAIS '93.* Seattle WA, September 18-22, 1993, IEEE Press: Los Alamitos, pp. 300-307.

32. Rekimoto, J. Matrix: A Realtime Object Identification and Registration Method for Augmented Reality. In *Proceedings of Asia Pacific Computer Human Interaction 1998 (APCHI'98),* Japan, 1998.

33. Rekimoto, J. Transvision: A Hand-held Augmented Reality System for Collaborative Design. In *Proceeding of Virtual Systems and Multimedia '96* (VSMM '96), Gifu, Japan, 18-20 Sept., 1996.

34. Schmalsteig, D., Fuhrmann, A., Szalavari, Z., Gervautz, M., Studierstube - An Environment for Collaboration in Augmented Reality. In *CVE '96 Workshop Proceedings,* 19-20th September 1996, Nottingham, Great Britain.

35. Schmandt, C., Mullins, A. AudioStreamer: Exploiting Simultaneity for Listening. In *Proceedings of CHI 95 Conference Companion,* May 7-11, Denver Colorado, 1995, ACM: New York pp. 218-219.

36. Sellen, A. Remote Conversations: The effects of mediating talk with technology. *Human Computer Interaction,* 1995, Vol. 10, No. 4, pp. 401-444.

37. Sellen, A. Speech Patterns in Video-Mediated Conversations. In *Proceedings CHI '92,* May 3-7, 1992, New York, ACM Press, pp. 49-59.

38. Short, J., Williams, E., Christie. B. *The Social Psychology of Telecommunications.* London, Wiley 1976.

39. Soltan, P., Trias, J., Dahlke, W., Lasher, M., McDonald, M. Laser-Based 3D Volumetric Display System: Second Generation. In *Interactive Technology and the New Paradigm for Technology,* IOP Press, 1995, pp. 349-358.

40. State, A., Hirota, G., Chen, D., Garrett, W., Livingston, M. Superior Augmented-Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. In *Proceedings of SIGGRAPH 96,* pp.429-438.

41. Wexelblat, A. The Reality of Cooperation: Virtual Reality and CSCW, in *Virtual Reality: Applications and Explorations.* Edited by A. Wexelblat. Boston, Academic Publishers, 1993.

42. Whittaker, S., O'Connaill, B. The Role of Vision in Face-to-Face and Mediated Communication. In *Video-Mediated Communication,* Eds. Finn, K., Sellen, A., Wilbur, S. Lawerance Erlbaum, New Jersey, 1997, pp. 23-49.