Jannick P. Rolland
School of Optics/CREOL, and
School of Electrical Engineering and
Computer Science
University of Central Florida
Orlando, FL 32816–2700

Henry Fuchs
Department of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175

Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization

Abstract

We compare two technological approaches to augmented reality for 3-D medical visualization: optical and video see-through devices. We provide a context to discuss the technology by reviewing several medical applications of augmented-reality research efforts driven by real needs in the medical field, both in the United States and in Europe. We then discuss the issues for each approach, optical versus video, from both a technology and human-factor point of view. Finally, we point to potentially promising future developments of such devices including eye tracking and multifocus planes capabilities, as well as hybrid optical/video technology.

1 Introduction

One of the most promising and challenging future uses of head-mounted displays (HMDs) is in applications in which virtual environments enhance rather than replace real environments. This is referred to as *augmented reality* (Bajura, Fuchs, & Ohbuchi, 1992). To obtain an enhanced view of the real environment, users wear see-through HMDs to see 3-D computer-generated objects superimposed on their real-world view. This see-through capability can be accomplished using either an optical HMD, as shown in figure 1, or a video see-through HMD, as shown in figure 2. We shall discuss the tradeoffs between optical and video see-through HMDs with respect to technological and human-factor issues, and discuss our experience designing, building, and testing these HMDs in medical visualization.

With optical see-through HMDs, the real world is seen through half-transparent mirrors placed in front of the user's eyes, as shown in figure 1. These mirrors are also used to reflect the computer-generated images into the user's eyes, thereby optically combining the real- and virtual-world views. With a video see-through HMD, the real-world view is captured with two miniature video cameras mounted on the head gear, as shown in figure 2, and the computer-generated images are electronically combined with the video representation of the real world (Edwards, Rolland, & Keller, 1993; State et al., 1994).

See-through HMDs were first developed in the 1960s. Ivan Sutherland's 1965 and 1968 optical see-through and stereo HMDs were the first computer graphics-based HMDs that used miniature CRTs for display devices, a mechanical tracker to provide head position and orientation in real time, and a hand-tracking device (Sutherland, 1965, 1968). While most of the developments in see-through HMDs aimed at military applications (Buchroeder,

Presence, *Vol. 9, No. 3, June 2000,* 287–309

Figure 1. *Optical see-through head-mounted display (Photo courtesy of KaiserElectro-Optics).*

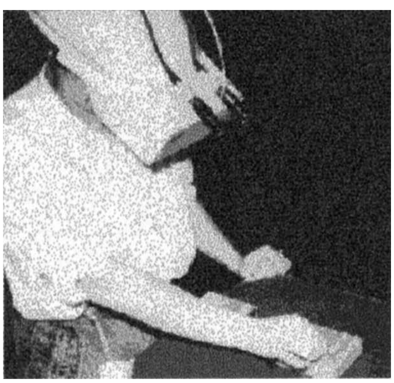


Figure 2. *A custom optics video see-through head-mounted display developed at UNC-CH. Edwards et al. (1993) designed the miniature video cameras. The viewer was a large FOV opaque HMD from Virtual Research.*

Seeley, & Vukobratovich, 1981; Furness, 1986; Droessler & Rotier, 1990; Barrette, 1992; Kandebo, 1988; Desplat, 1997), developments in 3-D scientific and medical visualization were initiated in the 1980s at the University of North Carolina at Chapel Hill (Brooks, 1992).

In this paper, we shall first review several medical visualization applications developed using optical and video see-through technologies. We shall then discuss technological and human-factors and perceptual issues related to see-through devices, some of which are employed in the various applications surveyed. Finally, we shall discuss what the technology may evolve to become.

2 Some Past and Current Applications of Optical and Video See-Through HMDs

The need for accurate visualization and diagnosis in health care is crucial. One of the main developments of medical care has been imaging. Since the discovery of X-rays in 1895 by Wilhelm Roentgen, and the first X-ray clinical application a year later by two Birmingham (UK) doctors, X-ray imaging and other medical imaging modalities (such as CT, ultrasound, and NMR) have emerged. Medical imaging allows doctors to view aspects of the interior architecture of living beings that were unseen before. With the advent of imaging technologies, opportunities for minimally invasive surgical procedures have arisen. Imaging and visualization can be used to guide needle biopsy, laparoscopic, endoscopic, and catheter procedures. Such procedures do require additional training because the physicians cannot see the natural structures that are visible in open surgery. For example, the natural eye-hand coordination is not available during laparoscopic surgery. Visualization techniques associated with see-through HMDs promise to help restore some of the lost benefits of open surgery (for example, by projecting a virtual image directly on the patient, eliminating the need for a remote monitor).

The following paragraphs briefly discuss examples of recent and current research conducted with optical see-through HMDs at the University of North Carolina at Chapel Hill (UNC-CH), the University of Central Florida (UCF), and the United Medical and Dental Schools of Guy's and Saint Thomas's Hospitals in England, video-see-through at UNC-CH, and hybrid optical-video see-through at the University of Blaise Pascal in France.

A rigorous error-analysis for an optical see-through HMD targeted toward the application of optical see-through HMD to craniofacial reconstruction was conducted at UNC-CH (Holloway, 1995). The superimposition of CT skull data onto the head of the real patient would give the surgeons "X-ray vision." The premise of

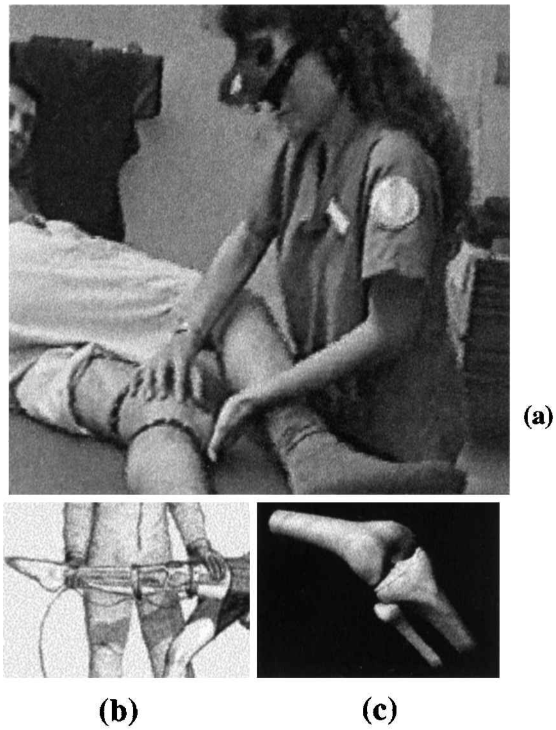


Figure 3. *(a) The VRDA tool will allow superimposition of virtual anatomy on a model patient. (b) An illustration of the view of the HMD user (Courtesy of Andrei State). (c) A rendered frame of the knee-joint bone structures animated based on a kinematic model of motion developed by Baillot and Rolland that will be integrated in the tool (1998).*

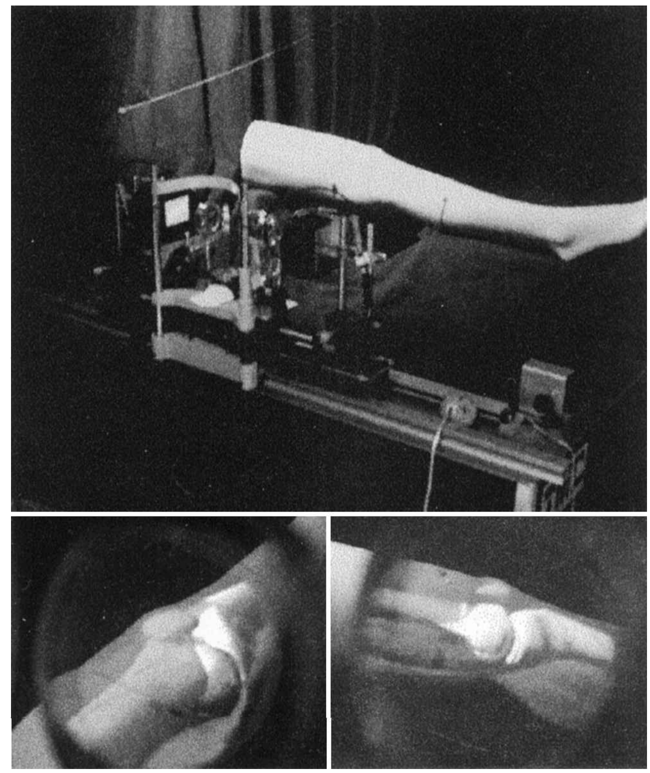


Figure 4. *First demonstration of the superimposition of a graphical knee-joint superimposed on a leg model for use in the VRDA tool: (a) a picture of the benchprototype setup; a snapshot of the superimposition through one lens of the setup in (b) a diagonal view and (c) a side view (1999).*

that system was that viewing the data in situ allows surgeons to make better surgical plans because they will be able to see the complex relationships between the bone and soft tissue more clearly. Holloway found that the largest registration error between real and virtual objects in optical see-through HMDs was caused by delays in presenting updated information associated with tracking. Extensive research in tracking has been pursued since at UNC-CH (Welch & Bishop, 1997).

One of the authors and colleagues are currently developing an augmented-reality tool for the visualization of human anatomical joints in motion (Wright et al., 1995; Kancherla et al., 1995; Rolland & Arthur, 1997; Parsons & Rolland, 1998; Baillot & Rolland, 1998; Baillot et al., 1999). An illustration of the tool using an optical see-through HMD for visualization of anatomy is shown in figure 3. In the first prototype, we have concentrated on the positioning of the leg around the knee joint. The joint is accurately tracked optically by using three infrared video cameras to locate active infrared markers placed around the joint. Figure 4 shows the results of the optical superimposition of the graphical knee joint on a leg model, seen through one of the lenses of our stereoscopic bench prototype display.

An optical see-through HMD coupled with optical tracking devices positioned along the knee joint of a model patient is used to visualize the 3-D computer-rendered anatomy directly superimposed on the real leg in motion. The user may further manipulate the joint and investigate the joint motions. From a technological aspect, the field of view (FOV) of the HMD should be sufficient to capture the knee-joint region, and the tracking devices and image-generation system must be fast enough to track typical knee-joint motions during manipulation at interactive speed. The challenge of capturing accurate knee-joint motion using optical markers located on the external surface of the joint was addressed by Rolland and Arthur (1997). The application aims at developing a more advanced tool for teaching dynamic anatomy (advanced in the sense that the tool allows combination of the senses of touch and vision). We aim this tool to specifically impart better understanding of bone motions during radiographic positioning for the radiological science (Wright et al., 1995).

To support the need for accurate motions of the knee joint in the Virtual Reality Dynamic Anatomy (VRDA) tool, an accurate kinematic model of joint motion based on the geometry of the bones and collision detection algorithms was developed (Baillot & Rolland, 1998; Baillot et al., 1999). This component of the research is described in another paper of this special issue (Baillot et al., 2000). The dynamic registration of the leg with the simulated bones is reported elsewhere (Outters et al., 1999). High-accuracy optical tracking methods, carefully designed and calibrated HMD technology, and appropriate computer graphics models for stereo pair generation play an important role in achieving accurate registration (Vaissie and Rolland, 2000; Rolland et al., 2000).

At the United Medical and Dental Schools of Guy's and Saint Thomas's Hospitals in England, researchers are projecting simple image features derived from preoperative magnetic resonance and computer-tomography images into the light path of a stereo operating microscope, with the goal of eventually allowing surgeons to visualize underlying structures during surgery. The first prototype used low-contrast color displays (Edwards et al., 1995). The current prototype uses high-contrast monochrome displays. The microscope is tracked intraoperatively, and the optics are calibrated (including zoom and focus) using a pinhole camera model. The intraoperative coordinate frame is registered using anatomical features and fiducial markers. The image features used in the display are currently segmented by hand. These include the outline of a lesion, the track of key nerves and blood vessels, and bone landmarks. This computer-guided surgery system can be said to be

Figure 5. *Real-time acquisition and superimposition of ultrasound slice images on a pregnant woman (1992).*

equivalent to an optical see-through system operating on a microscopic scale. In this case, the real scene is now seen through magnifying optics, but the eye of the observer is still the direct detecting device as in optical see-through.

One of the authors and colleagues at the UNC-CH are currently developing techniques that merge video and graphical images for augmented reality. The goal is to develop a system displaying live, real-time, ultrasound data properly registered in 3-D space on a scanned subject. This would be a powerful and intuitive visualization tool as well. The first application developed was the visualization of a human fetus during ultrasound echography. Figure 5 shows the real-time ultrasound images which appear to be pasted in front of the patient's body, rather than fixed within it (Bajura et al., 1992). Real-time imaging and visualization remains a challenge. Figure 6 shows a more recent, non-real-time implementation of the visualization in which the fetus is rendered more convincingly within the body (State et al., 1994).

Recently, knowledge from this video and ultrasound technology has also been applied to developing a visualization method for ultrasound-guided biopsies of breast lesions that were detected during mammography screening procedures (Figure 7) (State et al., 1996). This application was motivated from the challenges we observed during a biopsy procedure while collaborating on research with Etta Pisano, head of the Mammography Research Group at UNC-CH. The goal was to be able to locate any tumor within the breast as quickly and accurately as possible. The technology of video see-through

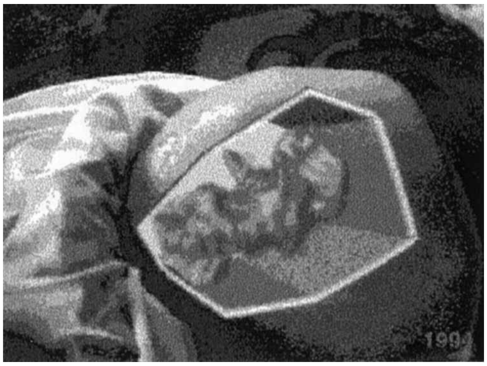


Figure 6. *Improved rendering of fetus inside the abdomen (1994).*

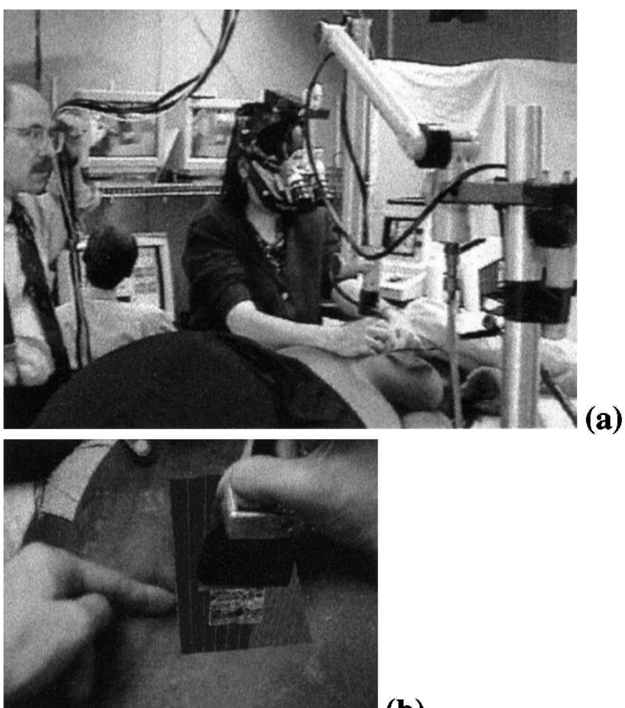


Figure 7. *Ultrasound guided biopsy (a) Laboratory setup during evaluation of the technology with Etta Pisano and Henry Fuchs (b) A view through the HMD (1996).*

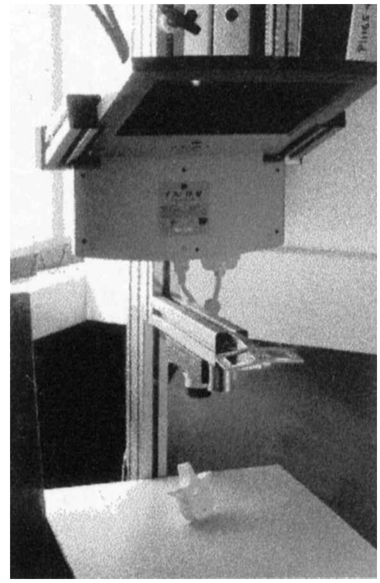


Figure 8. *Laboratory prototype of the hybrid optical/video see-through AR tool for guided scoliosis surgery developed by Peuchot at the University of Blaise Pascal, France (1995).*

already developed was thus applied to this problem. The conventional approach to biopsy is to follow the insertion of a needle in the breast tissue with a remote monitor displaying real-time, 2-D, ultrasound depth images. Such a procedure typically requires five insertions of the needle to maximize the chances of biopsy of the lesion. In the case in which the lesion is located fairly deep in the breast tissue, the procedure is difficult and can be lengthy (one to two hours is not atypical for deep lesions). Several challenges remain to be overcome before the technology developed can actually be tested in the clinic, including accurate and precise tracking and a technically reliable HMD. The technology may have applications in guiding laparoscopy, endoscopy, or catheterization as well.

At the University of Blaise Pascal in Clermont Ferrand, France, researchers developed several augmented-reality visualization tools based on hybrid optical and video see-through to assist in surgery to correct scoliosis (abnormal curvature of the spine column) (Peuchot, Tanguy, & Eude, 1994, 1995). This application was developed in collaboration with a surgeon of infantile scoliosis. The visualization system shown in figure 8 is from an optics point of view, the simplest see-through system one may conceive. It is first of all fixed on a stand, and it is designed as a viewbox positioned above the patient.

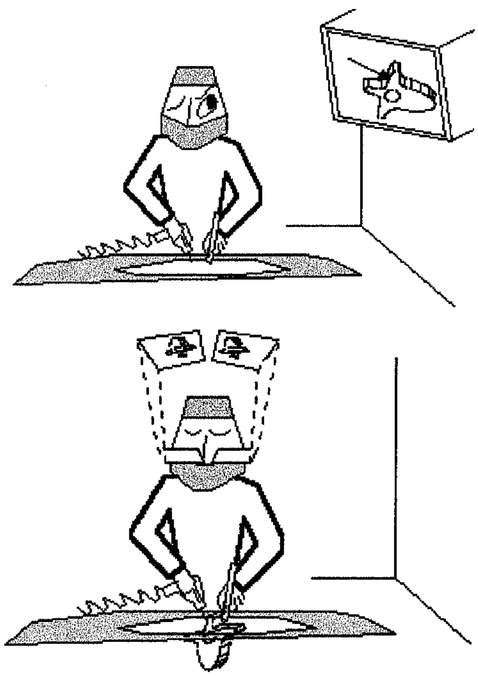


Figure 9. *Graphics illustration of current and future use of computer-guided surgery according to Bernard Peuchot.*

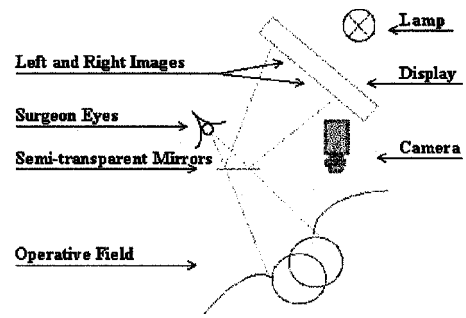


Figure 10. *Optical scheme of the hybrid optical/video see-through AR tool shown in Fig. 8.*

The surgeon positions himself above the viewbox to see the patient, and the graphical information is superimposed on the patient as illustrated in figure 9. The system includes a large monitor where a stereo pair of images is displayed, as well as half-silvered mirrors that allow the superimposition of the real and virtual objects. The monitor is optically imaged on a plane through the semi-transparent mirrors, and the spine under surgery is located within a small volume around that plane. An optical layout of the system is shown in figure 10.

In the above hybrid optical-video system, vertebrae are located in space by automatic analysis of the perspective view from a single video camera of the vertebrae. A standard algorithm such as the inverse perspective algorithm is used to extract the 3-D information from the projections observed in the detector plane (Dhome et al., 1989). The method relies heavily on accurate video tracking of vertebral displacements. High-accuracy algorithms were developed to support the application including development of subpixel detectors and calibration techniques. The method has been validated on vertebral specimens and accuracy of submillimeters in depth has been demonstrated (Peuchot, 1993, 1994).

The success of the method can be attributed to the fine calibration of the system, which, contrary to most systems, does not assume a pinhole camera model for the video camera. Moreover, having a fixed viewer with no optical magnification (contrary to typical HMDs) and a constant average plane of surgical operation reduces the complexity of problems such as registration and visualization. It can be shown, for example, that rendered depth errors are minimized when the virtual image planes through the optics (a simple semi-transparent mirror in Peuchot's case) is located in the average plane of the 3-D virtual object visualized (Rolland et al., 1995). Furthermore, the system avoids the challenging problems of tracking, optical distortion compensation, and conflicts of accommodation and convergence related to HMDs (Robinett & Rolland, 1992; Rolland & Hopkins, 1993). Some tracking and distortion issues will be further discussed in sections 3.1 and 3.2, respectively. However, good registration of real and virtual objects in a static framework is a first step to good calibration in a dynamic framework, and Peuchot's results are state of the art in this regard.

It is important to note that the method developed for this application employs a hybrid optical-video technology. In this case, video is essentially used to localize real objects in the surgical field, and optical see-through is used as the visualization tool for the surgeon. While the first system developed used one video camera, the methods have been extended to include multiple cameras

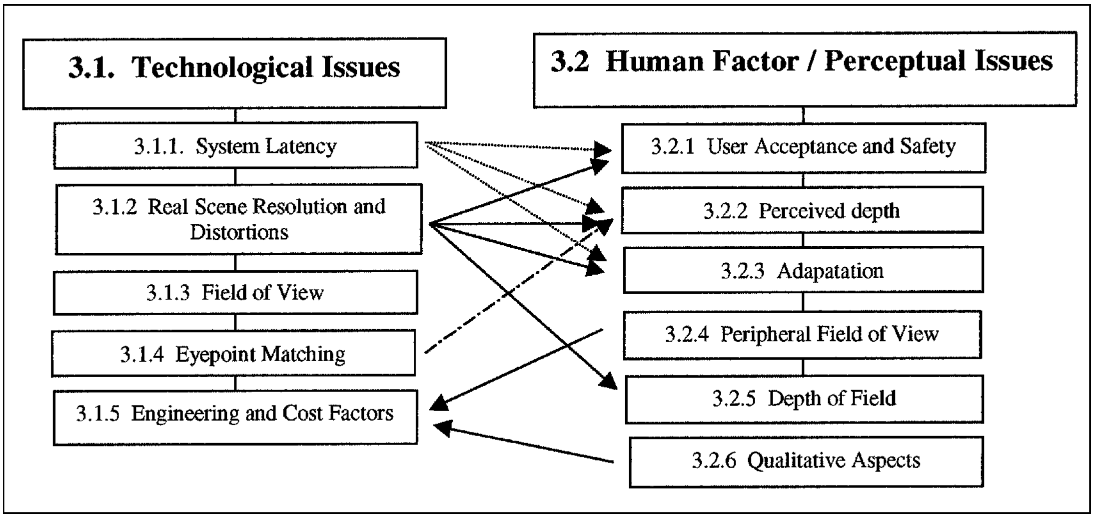


Figure 11. *Outline of sections 3.1 and 3.2 of this paper.*

with demonstrated accuracy and precision of 0.01 mm (Peuchot, 1998). Peuchot chose the hybrid system over a video see-through approach because "it allows the operator to work in his real environment with a perception space that is real." Peuchot judged this point to be critical in a medical application like surgery.

3 A Comparison of Optical and Video See-Through Technology

As suggested in the description of the applications described, the main goal of augmented-reality systems is to merge virtual objects into the view of the real scene so that the user's visual system suspends disbelief into perceiving the virtual objects as part of the real environment. Current systems are far from perfect, and system designers typically end up making a number of application-dependent trade offs. We shall list and discuss these tradeoffs in order to guide the choice of technology depending upon the type of application considered.

Both systems, optical and video, have two image sources: the real world and the computer-generated world. These image sources are to be merged. Optical see-through HMDs take what might be called a "minimally obtrusive" approach; that is, they leave the view of the real world nearly intact and attempt to augment it by merging a reflected image of the computer-generated scene into the view of the real world. Video see-through HMDs are typically more obtrusive in the sense that they block out the real-world view in exchange for the ability to merge the two views more convincingly. In recent developments, narrow fields of view in video see-through HMDs have replaced large field-of-view HMDs, thus reducing the area where the real world (captured through video) and the computer-generated images are merged into a small part of the visual scene. In any case, a fundamental consideration is whether the additional features afforded by video see-through HMDs justify the loss of the unobstructed real-world view.

Our experience indicates that there are many tradeoffs between optical and video see-through HMDs with respect to technological and human-factors issues that affect designing, building, and assessing these HMDs. The specific issues are laid out in figure 11. While most of these issues could be discussed from both a technological and human-factors-standpoint (because the two are closely interrelated in HMD systems), we have chosen to classify each issue where it is most adequately addressed at this time, given the present state of the technology. For example, delays in HMD systems are addressed under technology because technological improvements are actively being pursued to minimize delays. Delays also certainly have impact on various human-factor issues (such as the perceived location of objects in depth and user acceptance). Therefore, the multiple arrows shown in figure 11 indicate that the technological and human-factor-categories are highly interrelated.

3.1 Technological Issues

The technological issues for HMDs include latency of the system, resolution and distortion of the real scene, field of view (FOV), eyepoint matching of the see-through device, and engineering and cost factors. While we shall discuss properties of both optical and video see-through HMDs, it must be noted that, contrary to optical see-through HMDs, there are no commercially available products for video see-through HMDs. Therefore, discussions of such systems should be considered carefully as findings may be particular to only a few current systems. Nevertheless, we shall provide as much insight as possible into what we have learned with such systems as well.

3.1.1 System Latency. An essential component of see-through HMDs is the capacity to properly register a user's surroundings and the synthetic space. A geometric calibration between the tracking devices and the HMD optics must be performed. The major impediment to achieving registration is the gap in time, referred to as *lag,* between the moment when the HMD position is measured and the moment when the synthetic image for that position is fully rendered and presented to the user.

Lag is the largest source of registration error in most current HMD systems (Holloway, 1995). This lag in typical systems is between 60 ms and 180 ms. The head of a user can move during such a period of time, and the discrepancy in perceived scene and superimposed scene can destroy the illusion of the synthetic objects being fixed in the environment. The synthetic objects can "swim" around significantly in such a way that they may not even seem to be part of the real object to which they belong. For example, in the case of ultrasound-guided biopsy, the computer-generated tumor may appear to be located outside the breast while tracking the head of the user. This swimming effect has been demonstrated and minimized by predicting HMD position instead of simply measuring positions (Azuma & Bishop, 1994).

Current HMD systems are lag limited as a consequence of tracker lag, the complexity of rendering, and displaying the images. Tracker lag is often not the limiting factor in performance. If displaying the image is the limiting factor, novel display architectures supporting frameless rendering can help solve the problem (Bishop et al., 1994). Frameless rendering is a procedure for continuously updating a displayed image, as information becomes available instead of updating entire frames at a time. The tradeoffs between lag and image quality are currently being investigated (Scher-Zagier, 1997). If we assume that we are limited by the speed of rendering an image, eye-tracking capability may be useful to quickly update information only around the gaze point of the user (Thomas et al., 1989; Rolland, Yoshida, et al., 1998; Vaissie & Rolland, 1999).

One of the major advantages of video see-through HMDs is the potential capability of reducing the relative latencies between the 2-D real and synthetic images as a consequence of both types of images being digital (Jacobs et al., 1997). Manipulation of the images in space and in time is applied to register them. Three-dimensional registration is computationally intensive, if at all robust, and challenging for interactive speed. The spatial approach to forcing registration in video see-through systems is to correct registration errors by imaging landmark points in the real world and registering virtual objects with respect to them (State et al., 1996). One approach to eliminating temporal delays between the real and computer-generated images in such a case is to capture a video image and draw the graphics on top of the video image. Then the buffer is swapped, and the combined image is presented to the HMD user. In such a configuration, no delay apparently exists between the real and computer-generated images. If the actual latency of the computer-generated image is large with respect to the video image, however, it may cause sensory conflicts between vision and proprioception because the video images no longer correspond to the real-world scene. Any manual interactions with real objects could suffer as a result.

Another approach to minimizing delays in video see-through HMDs is to delay the video image until the computer-generated image is rendered. Bajura and Neumann (1995) applied chroma keying, for example, to dynamically image a pair of red LEDs placed on two real objects (one stream) and then registered two virtual objects with respect to them (second stream). By tracking more landmarks, better registration of real and virtual objects may be achieved (Tomasi and Kanade, 1991). The limitation of the approach taken is the attempt to register 3-D scenes using 2-D constraints. If the user rotates his head rapidly or if a real-world object moves, there may be no "correct" transformation for the virtual scene image. To align all the landmarks, one must either allow errors in registration of some of the landmarks or perform a nonlinear warping of the virtual scene that may create undesirable distortions of the virtual objects. The nontrivial solution to this problem is to increase the speed of the system until scene changes between frames are small and can be approximated with simple 2-D transformations.

In a similar vein, it is also important to note that the video view of the real scene will normally have some lag due to the time it takes to acquire and display the video images. Thus, the image in a video see-through HMD will normally be slightly delayed with respect to the real world, even without adding delay to match the synthetic images. This delay may increase if an image-processing step is applied to either enforce registration or perform occlusion. The key issue is whether the delay in the system is too great for the user to adapt to it (Held & Durlach, 1987).

Systems using optical see-through HMDs have no means of introducing artificial delays into the real scene. Therefore, the system may need to be optimized for low latency, perhaps less than 60 ms, where predictive tracking can be effective (Azuma & Bishop, 1994). For any remaining lag, the user may have to limit his actions to slow head motions. Applications in which speed of movement can be readily controlled, such as in the VRDA tool described earlier, can benefit from optical see-through technology (Rolland & Arthur, 1997). The advantage of having no artificial delays is that real objects will always be where they are perceived to be, and this may be crucial for a broad range of applications.

3.1.2 Real-Scene Resolution and Distortion. If real-scene resolution refers to the resolution of the real-scene object, the best real-scene resolution that a see-through device can provide is that perceived with the naked eye under unit magnification of the real scene. Certainly under microscopic observation as described by Hill (Edwards et al., 1995), the best scene resolution goes beyond that obtained with a naked eye. It is also assumed that the see-through device has no image-processing capability.

A resolution extremely close to that obtained with the naked eye is easily achieved with a nonmicroscopic optical see-through HMD, because the optical interface to the real world is simply a thin parallel plate (such as a glass plate) positioned between the eyes and the real scene. Such an interface typically introduces only very small amounts of optical aberrations to the real scene: For example, for a real-point object seen through a 2 mm planar parallel plate placed in front of a 4 mm dia. eye pupil, the diffusion spot due to spherical aberration would subtend a 2 10^{-7} arc-minute visual angle for a point object located 500 mm away. Spherical aberration is one of the most common and simple aberrations in optical systems that lead to blurring of the images. Such a degradation of image quality is negligible compared to the ability of the human eye to resolve a visual angle of 1 minute of arc. Similarly, planar plates introduce low distortion of the real scene, typically below 1%. There is no distortion only for the chief rays that pass the plate parallel to its normal.[1]

In the case of a video see-through HMD, real-scene images are digitized by miniature cameras (Edwards et al., 1993) and converted into an analog signal that is fed to the HMD. The images are then viewed through the HMD viewing optics that typically use an eyepiece design. The perceived resolution of the real scene can thus be limited by the resolution of the video cameras or the HMD viewing optics. Currently available miniature

1. A chief ray is defined as a ray that emanates from a point in the FOV and passes through the center of the pupils of the system. The exit pupil in an HMD is the entrance pupil of the human eye.

video cameras typically have a resolution of 640×480, which is also near the resolution limit of the miniature displays currently used in HMDs.[2] Depending upon the magnification and the field of view of the viewing optics, various effective visual resolutions may be reached. While the miniature displays and the video cameras seem to currently limit the resolution of most systems, such performance may improve with higher-resolution detectors and displays.

In assessing video see-through systems, one must distinguish between narrow and wide FOV devices. Large-FOV ($\geq$50 deg.) eyepiece designs are known to be extremely limited in optical quality as a consequence of factors such as optical aberrations that accompany large FOVs, pixelization that may become more apparent under large magnification, and the exit pupil size that must accommodate the size of the pupils of a person's eyes. Thus, even with higher-resolution cameras and displays, video see-through HMDs may remain limited in their ability to provide a real-scene view of high resolution if conventional eyepiece designs continue to be used. In the case of small to moderate FOV (10 deg. to 20 deg.) video see-through HMDs, the resolution is still typically much less than the resolving power of the human eye.

A new technology, referred to as *tiling,* may overcome some of the current limitations of conventional eyepiece design for large FOVs (Kaiser, 1994). The idea is to use multiple narrow-FOV eyepieces coupled with miniature displays to completely cover (or tile) the user's FOV. Because the individual eyepieces have a fairly narrow FOV, higher resolution (nevertheless currently less than the human visual system) can be achieved. One of the few demonstrations of high-resolution, large-FOV displays is the tiled displays. A challenge is the minimization of seams in assembling the tiles, and the rendering of multiple images at interactive speed. The tiled displays certainly bring new practical and computational challenges that need to be confronted. If a see-through capability is desired (for example, to display virtual furniture in an empty room), it is currently unclear whether the technical problems associated with providing overlay can be solved.

Theoretically, distortion is not a problem in video see-through systems because the cameras can be designed to compensate for the distortion of the optical viewer, as demonstrated by Edwards et al. (1993). However, if the goal is to merge real and virtual information, as in ultrasound echography, having a warped real scene significantly increases the complexity of the synthetic-image generation (State et al., 1994). Real-time video correction can be used at the expense of an additional delay in the image-generation sequence. An alternative is to use low-distortion video cameras at the expense of a narrower FOV, merge unprocessed real scenes with virtual scenes, and warp the merged images. Warping can be done using (for example) real-time texture mapping to compensate for the distortion of the HMD viewing optics as a last step (Rolland & Hopkins, 1993; Watson & Hodges, 1995).

The need for high, real-scene resolution is highly task dependent. Demanding tasks such as surgery or engineering training, for example, may not be able to tolerate much loss in real-scene resolution. Because the large-FOV video see-through systems that we have experienced are seriously limited in terms of resolution, narrow-FOV video see-through HMDs are currently preferred. Independently of resolution, an additional critical issue in aiming towards narrow-FOV video see-through HMDs is the need to match the viewpoint of the video cameras with the viewpoint of the user. Matching is challenging with large-FOV systems. Also, methods for matching video and real scenes for large-FOV tiled displays must be developed. At this time, considering the growing availability of high-resolution flat-panel displays, we foresee that the resolution of see-though HMDs could gradually increase for both small- and large-FOV systems. The development and marketing of miniature high-resolution technology must be undertaken to achieve resolutions that match that of human vision.

3.1.3 Field of View.

A generally challenging issue of HMDs is providing the user with an adequate FOV for a given application. For most applications, increasing

2. The number of physical elements is typically 640×480. One can use signal processing to interpolate between lines to get higher resolutions.

the binocular FOV means that fewer head movements are required to perceive an equivalently large scene. We believe that a large FOV is especially important for tasks that require grabbing and moving objects and that it provides increased situation awareness when compared to narrow-FOV devices (Slater & Wilbur, 1997). The situation with see-through devices is somewhat different from that of fully opaque HMDs in that the aim of using the technology is different from that of immersing the user in a virtual environment.

3.1.3.1 Overlay and Peripheral FOV. The term *overlay FOV* is defined as the region of the FOV where graphical information and real information are superimposed. The peripheral FOV is the real-world FOV beyond the overlay FOV. For immersive opaque HMDs, no such distinction is made; one refers simply to the FOV. It is important to note that the overlay FOV may need to be narrow only for certain augmented-reality applications. For example, in a visualization tool such as the VRDA tool, only the knee-joint region is needed in the overlay FOV. In the case of video HMD-guided breast biopsy, the overlay FOV could be as narrow as the synthesized tumor. The real scene need not necessarily be synthesized. The available peripheral FOV, however, is critical for situation awareness and is most often required for various applications whether it is provided as part of the overlay or around the overlay. If provided around the overlay, the transition from real to virtual imagery must be made as seamless as possible. This is an investigation that has not yet been addressed in video see-through HMDs.

Optical see-through HMDs typically provide from 20 deg. to 60 deg. overlay FOV via the half-transparent mirrors placed in front of the eyes, a characteristic that may seem somewhat limited but promising for a variety of medical applications whose working visualization distance is within arm reach. Larger FOVs have been obtained, up to 82.5×67 deg., at the expense of reduced brightness, increased complexity, and massive, expensive technology (Welch & Shenker, 1984). Such FOVs may have been required for performing navigation tasks in real and virtual environments but are likely not required in most augmented-reality applications. Optical see-through HMDs, however, whether or not they have a large overlay FOV, have been typically designed open enough that users can use their peripheral vision around the device, thus increasing the total real-world FOV to closely match one's natural FOV. An annulus of obstruction usually results from the mounts of the thin see-through mirror similar to the way that our vision may be partially occluded by a frame when wearing eyeglasses.

In the design of video see-through HMDs, a difficult engineering task is matching the frustum of the eye with that of the camera (as we shall discuss in section 3.1.4). While such matching is not so critical for far-field viewing, it is important for near-field visualization as in various medical visualizations. This difficult matching problem has lead to the consideration of narrower-FOV systems. A compact, 40×30 deg. FOV design, designed for optical see-through HMD but adaptable to video see-through, was proposed by Manhart, Malcolm, & Frazee (1993). Video see-through HMDs, on the other hand, can provide (in terms of a see-through FOV) the FOV displayed with the opaque type viewing optics that typically range from 20 deg. to 90 deg. In such systems where the peripheral FOV of the user is occluded, the effective real-world FOV is often smaller than in optical see-through systems. When using a video see-through HMD in a hand-eye coordination task, we found in a recent human-factor study that users needed to perform larger head movements to scan an active field of vision than when performing the task with the unaided eye (Biocca & Rolland, 1998). We predict that the need to make larger head movements would not arise as much with see-through HMDs with equivalent overlay FOVs but larger peripheral FOVs, because users are provided with increased peripheral vision, and thus additional information, to more naturally perform the task.

3.1.3.2 Increasing Peripheral FOV in Video See-Through HMDs. An increase in peripheral FOV in video see-through systems can be accomplished in two ways: in a folded optical design, as used for optical see-through HMDs, however with an opaque mirror instead of a half-transparent mirror, or in a nonfolded design but with nonenclosed mounts. The latter calls for innovative optomechanical design because heavier optics have to be supported than in either optical or folded video see-through. Folded systems require only a thin mirror in front of the eyes, and the heavier optical components are placed around the head. However, the tradeoff with folded systems is a significant reduction in the overlay FOV.

3.1.3.3 Tradeoff Resolution and FOV. While the resolution of a display in an HMD is defined in the graphics community by the number of pixels, the relevant measure of resolution is the number of pixels per angular FOV, which is referred to as *angular resolution*. Indeed, what is of importance for usability is the angular subtends of a pixel at the eye of the HMD user. Most current high-resolution HMDs achieve higher resolution at the expense of a reduced FOV. That is, they use the same miniature, high-resolution CRTs but with optics of less magnification in order to achieve higher angular resolution. This results in a FOV that is often narrow. The approach that employs large high-resolution displays, or light valves, and transports the high-resolution images to the eyes by imaging optics coupled to a bundle of optical fibers achieves high resolution at fairly large FOVs (Thomas et al., 1989). The current proposed solutions that improve resolution without trading FOV are either tiling techniques, high-resolution inset displays (Fernie, 1995; Rolland, Yoshida, et al., 1998), or projection HMDs (Hua et al., 2000).

Projective HMDs differ from conventional HMDs in that projection optics are used instead of eyepiece optics to project real images of miniature displays in the environment. A screen placed in the environment reflects the images back to the eyes of the user. Projective HMDs have been designed and demonstrated, for example, by Kijima and Ojika (1997) and Parsons and Rolland (1998). Kijima used a conventional projection screen in his prototype. Parsons and Rolland developed a first-prototype projection HMD system to demonstrate that an undistorted virtual 3-D image could be rendered when projecting a stereo pair of images on a bent sheet of microretroreflector cubes. The first proof-of-concept system is shown in figure 12. A comprehensive investigation of the optical characteristics of projective HMDs is given by Hua et al. (2000). We are also developing the next-generation prototypes of the technology using custom-made miniature lightweight optics. The system presents various advantages over conventional HMDs, including distortion-free images, occluded virtual objects from real-objects interposition, no image cross-talks for multiuser participants in the virtual world, and the potential for a wide FOV (up to 120 deg.).

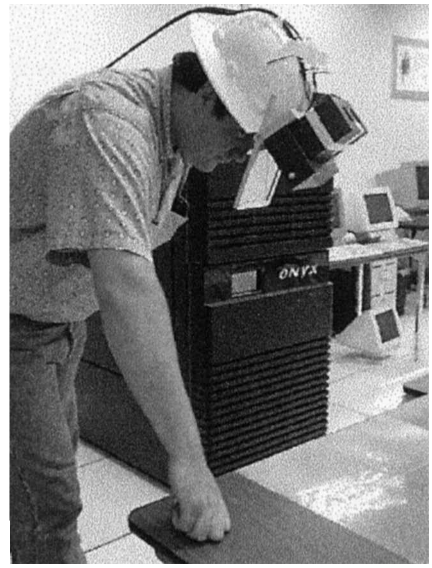
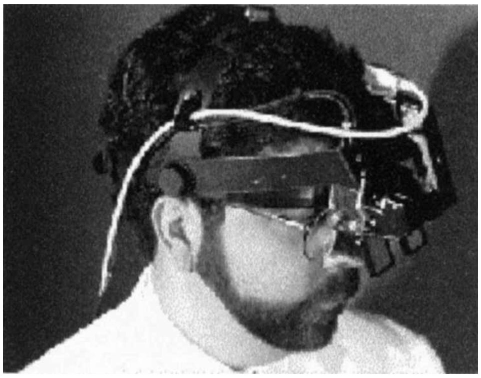


Figure 12. *Proof of concept prototype of a projection head-mounted display with microreflector sheeting (1998).*

3.1.4 Viewpoint Matching. In video see-through HMDs, the camera viewpoint (that is, the entrance pupil) must be matched to the viewpoint of the observer (the entrance pupil of the eye). The viewpoint of a camera or eye is equivalent to the center of projection used in the computer graphics model that computes the stereo images and is taken here to be the center of the entrance pupil of the eye or camera (Vaissie & Rolland, 2000). In earlier video see-through designs, Edwards et al. (1993) investigated ways to mount the cameras to minimize errors in viewpoint matching. The error minimization versus exact matching was a consequence of working with wide-FOV systems. If the viewpoints of the cameras do not match the viewpoints of the eyes, the user experiences a spatial shift in the perceived scene that may lead to perceptual anomalies (as further



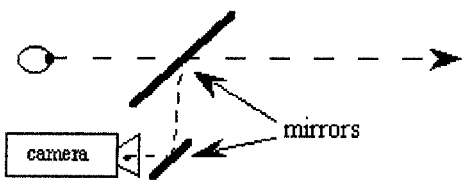


Figure 13. *A 10 degree FOV video see-through HMD: Dglasses developed at UNC-CH. Lipstick cameras and a double fold mirror arrangement was used to match the viewpoints of the camera and user (1997).*

discussed under human-factors issues (Biocca & Rolland, 1998). Error analysis should then be conducted in such a case to match the need of the application.

For cases in which the FOV is small (less than approximately 20 deg.), exact matching in viewpoints is possible. Because the cameras cannot be physically placed at the actual eyepoints, mirrors can be used to fold the optical path (much like a periscope) to make the cameras' viewpoints correspond to the real eyepoints as shown in figure 13 (Edwards et al., 1993). While such geometry solves the problem of the shift in viewpoint, it increases the length of the optical path, which reduces the field of view, for the same reason that optical see-through HMDs tend to have smaller fields of view. Thus, video see-through HMDs must either trade their large FOVs for correct real-world viewpoints or require the user to adapt to the shifted viewpoints as further discussed in section 3.2.3.

Finally, correctly mounting the video cameras in a video see-through HMD requires that the HMD has an interpupillary distance (IPD) adjustment. Given the IPD of a user, the lateral separation of the video cameras must then be adjusted to that value in order for the views obtained by the video cameras to match those that would have been obtained with naked eyes. If one were to account for eye movements in video see-through HMDs, the level of complexity in slaving the camera viewpoint to the user viewpoint would be highly increased. To our knowledge, such complexity has not yet been considered.

3.1.5 Engineering and Cost Factors. HMD designs often suffer from fairly low resolution, limited FOV, poor ergonomic designs, and excessive weight. A good ergonomic design requires an HMD whose weight is similar to a pair of eyeglasses, or which folds around the user's head so the device's center of gravity falls near the center of rotation of the head (Rolland, 1994). The goal here is maximum comfort and usability. Reasonably lightweight HMD designs currently suffer narrow FOVs, on the order of 20 deg. To our knowledge, at present, no large-FOV stereo see-through HMDs of any type are comparable in weight to a pair of eyeglasses. Rolland predicts that it could be achieved with some emerging technology of projection HMDs (Rolland, Parsons, et al., 1998). However, it must be noted that such technology may not be well suited to all visualization schemes as it requires a projection screen somewhere in front of the user that is not necessarily attached to the user's head.

With optical see-through HMDs, the folding can be accomplished with either an on-axis or an off-axis design. Off-axis designs are more elegant and also far more attractive because they elimate the ghost images that currently plague users of on-axis HMDs (Rolland, 2000). Off-axis designs are not commercially available because very few prototypes have been built (and those that have been built are classified) (Shenker, 1998). Moreover, off-axis systems are difficult to design and are thus expensive to build as a result of off-axis components (Shenker, 1994). A nonclassified, off-axis design has been designed by Rolland (1994, 2000). Several factors (including cost) have also hindered the construction of a first prototype as well. New generations of computer-controlled fabrication and testing are expected to change this trend.

Since their beginning, high-resolution HMDs have been CRT based. Early systems were even monochrome, but color CRTs using color wheels or frame-sequential color have been fabricated and incorporated into HMDs (Allen, 1993). Five years ago, we may have thought that, today, high-resolution, color, flat-panel displays would be the first choice for HMDs. While this is slowly happening, miniature CRTs are not fully obsolete. The current optimism, however, is prompted by new technologies such as reflective LCDs, microelectromechanical systems (MEMS)-based displays, laser-based displays, and nanotechnology-based displays.

3.2 Human-Factor and Perceptual Issues

Assuming that many of the technological challenges described have been addressed and high-performance HMDs can be built, a key human-factor issue for see-through HMDs is that of user acceptance and safety, which will be discussed first. We shall then discuss the technicalities of perception in such displays. The ultimate see-through display is one that provides quantitative and qualitative visual representations of scenes that conform to a predictive model (for example, conform to that given by the real world if that is the intention). Issues include the accuracy and precision of the rendered and perceived locations of objects in depth, the accuracy and precision of the rendered and perceived sizes of real and virtual objects in a scene, and the need of an unobstructed peripheral FOV (which is important for many tasks that require situation awareness and the simple manipulation of objects and accessories.

3.2.1 User Acceptance and Safety. A fair question for either type of technology is "will anyone actually wear one of these devices for extended periods?" The answer will doubtless be specific to the application and the technology included, but it will probably center upon whether the advanced capabilities afforded by the technology offset the problems induced by the encumbrance and sensory conflicts that are associated with it.

In particular, one of us thinks that video see-through HMDs may be met with resistance in the workplace because they remove the direct, real-world view in order to augment it. This issue of trust may be difficult to overcome for some users. If wide-angle FOV video see-through HMDs are used, this problem is exacerbated in safety-critical applications. A key difference in such applications may turn out to be the failure mode of each technology. A technology failure in the case of optical see-through HMDs may leave the subject without any computer-generated images but still with the real-world view. In the case of video see-through, it may leave the user with the complete suppression of the real-world view, as well as the computer-generated view.

However, it may be that the issue has been greatly lessened because the video view occupies such a small fraction (approximately 10 deg. visual angle) of the scene in recent developments of the technology. It is especially true of flip-up and flip-down devices such as that developed at UNC-CH and shown in figure 13.

Image quality and its tradeoffs are definitely critical issues related to user acceptance for all types of technology. In a personal communication, Martin Shenker, a senior optical engineer with more than twenty years of experience designing HMDs, pointed out that there are currently no standards of image quality and technology specifications for the design, calibration, and maintenance of HMDs. This is a current concern at a time when the technology may be adopted in various medical visualizations.

3.2.2 Perceived Depth. *3.2.2.1 Occlusion.* The ability to perform occlusion in see-through HMDs is an important issue of comparison between optical and video see-through HMDs. One of the most important differences between these two technologies is how they handle the depth cue known as *occlusion* (or interposition). In real life, an opaque object can block the view of another object so that part or all of it is not visible. While there is no problem in making computer-generated objects occlude each other in either system, it is considerably more difficult to make real objects occlude virtual objects (and vice versa) unless the real world for an application is predefined and has been modeled in the computer. Even then, one would need to know the exact location of a user with respect to that real environment. This is not the case in most augmented-reality applications, in which the real world is constantly changing and on-the-fly acquisition is all the information one will ever have of the real world. Occlusion is a strong monocular cue to depth perception and may be required in certain applications (Cutting & Vishton, 1995).

In both systems, computing occlusion between the real and virtual scenes requires a depth map of both scenes. A depth map of the virtual scene is usually available (for *z*-buffered image generators), but a depth map of the real scene is a much more difficult problem. While one could create a depth map in advance from a static real environment, many applications require on-the-fly image acquisition of the real scene. Assuming the system has a depth map of the real environment, video see-through HMDs are perfectly positioned to take advantage of this information. They can, on a pixel-by-pixel basis, selectively block the view of either scene or even blend them to minimize edge artifacts. One of the chief advantages of video see-through HMDs is that they handle this problem so well.

The situation for optical see-through HMDs can be more complex. Existing optical see-through HMDs blend the two images with beam splitters, which blend the real and virtual images uniformly throughout the FOV. Normally, the only control the designer has is the amount of reflectance versus transmittance of the beam splitter, which can be chosen to match the brightness of the displays with the expected light levels in the real-world environment. If the system has a model of the real environment, it is possible to have real objects occlude virtual ones by simply not drawing the occluded parts of the virtual objects. The only light will then be from the real objects, giving the illusion that they are occluding the virtual ones. Such an effect requires a darkened room with light directed where it is needed. This technique has been used by CAE Electronics in their flight simulator. When the pilots look out the window, they see computer-generated objects. If they look inside the cockpit, however, the appropriate pixels of the computer-generated image are masked so that they can see the real instruments. The room is kept fairly dark so that this technique will work (Barrette, 1992). David Mizell (from Boeing Seattle) and Tom Caudell (University of New Mexico) are also using this technique; they refer to it as "fused reality" (Mizell, 1998).

While optical see-through HMDs can allow real objects to occlude virtual objects, the reverse is even more challenging because normal beam splitters have no way of selectively blocking out the real environment. This problem has at least two possible partial solutions. The first solution is to spatially control the light levels in the real environment and to use displays that are bright enough so that the virtual objects mask the real ones by reason of contrast. (This approach is used in the flight simulator just mentioned for creating the virtual instruments.) This may be a solution for a few applications. A possible second solution would be to locally attenuate the real-world view by using an addressable filter device placed on the see-through mirror. It is possible to generate partial occlusion in this manner because the effective beam of light entering the eye from some point in the scene covers only a small area of the beam splitter, the eye pupil being typically 2mm to 4mm in photopic vision. A problem with this approach is that the user does not focus on the beam splitter, but rather somewhere in the scene. A point in the scene maps to a disk on the beam splitter, and various points in the scene map to overlapping disks on the beam splitter. Thus, any blocking done at the beam splitter may occlude more of the scene than expected, which might lead to odd visual effects. A final possibility is that some applications may work acceptably without properly rendered occlusion cues. That is, in some cases, the user may be able to use other depth cues, such as head-motion parallax, to resolve the ambiguity caused by the lack of occlusion cues.

3.2.2.2 Rendered Locations of Objects in Depth. We shall distinguish between errors in the rendered and perceived locations of objects in depth. The former yields the latter. One can conceive, however, that errors in the perceived location of objects in depth can also occur even in the absence of errors in rendered depths as a result of an incorrect computational model for stereo pair generation or a suboptimal presentation of the stereo images. This is true both for optical and video see-through HMDs. Indeed, if the technology is adequate to support a computational model, and the model accounts for required technology and corresponding parameters, the rendered locations of objects in depth—as well as the resulting perceived locations of objects in depth—will follow expectations. Vaissie recently showed some limitations of the choice of a static eyepoint in computational models for stereo pair generation for virtual environments that yield errors in rendered and thus perceived location of objects in depths (Vaissie and Rolland, 2000). The ultimate goal is to derive a computational model and develop the required technology that yield the desired perceived location of objects in depth. Errors in rendered depth typically result from inaccurate display calibration and parameter determination such as the FOV, the frame-buffer overscan, the eyepoints' locations, conflicting or noncompatible cues to depth, and remaining optical aberrations including residual optical distortions.

3.2.2.3 FOV and Frame-Buffer Overscan. Inaccuracies of a few degrees in FOV are easily made if no calibration is conducted. Such inaccuracies can lead to significant errors in rendered depths depending on the imaging geometry. For some medical and computer-guided surgery applications, for example, errors of several millimeters are likely to be unacceptable. The FOV and the overscan of the frame buffer that must be measured and accounted for to yield accurate rendered depths are critical parameters for stereo pair generation in HMDs (Rolland et al., 1995). These parameters must be set correctly regardless of whether the technology is optical or video see-through.

3.2.2.4 Specification of Eyepoint Location. The specification of the locations of the user's eyepoints (which are used to render the stereo images from the correct viewpoints) must be specified for accurate rendered depth. This applies to both optical and video see-through HMDs. In addition, for video see-through HMDs, the real-scene video images must be acquired from the correct viewpoint (Biocca & Rolland, 1998).

For the computer graphics-generation component, three choices of eyepoint locations within the human eye have been proposed: the nodal point of the eye[3] (Robinett & Rolland, 1992; Deering, 1992), the entrance pupil of the eye (Rolland, 1994; Rolland et al., 1995), and the center of rotation of the eye (Holloway, 1995). Rolland (1995) discusses that the choice of the nodal point would in fact yield errors in rendered depth in all cases whether the eyes are tracked or not. For a device with eye-tracking capability, the entrance pupil of the eye should be taken as the eyepoint. If eye movements are ignored, meaning that the computer-graphics eyepoints are fixed, then it was proposed that it is best to select the center of rotation of the eye as the eyepoint (Fry, 1969; Holloway, 1995). An in-depth analysis of this issue reveals that while the center of rotation yields higher accuracy in position, the center of the entrance pupil yields in fact higher angular accuracy (Vaissie & Rolland, 2000). Therefore, depending on the task involved, and whether angular accuracy or position accuracy is most important, the centers of rotation or the centers of the entrance pupil may be selected as best eyepoints location in HMDs.

3.2.2.5 Residual Optical Distortions. Optical distortion is one of the few optical aberrations that do not affect image sharpness; rather, it introduces warping of the image. It occurs only for optics that include lenses. If the optics include only plane mirrors, there are no distortions (Peuchot, 1994). Warping of the images leads to errors in rendered depths. Distortion results from the locations of the user's pupils away from the nodal points of the optics. Moreover, it varies as a function of where the user looks through the optics. However, if the optics are well calibrated to account for the user's IPD, distortion will be fairly constant for typical eye movements behind the optics. Prewarping of the computer-generated image can thus be conducted to compensate for the

3. Nodal points are conjugate points in an optical system that satisfy an angular magnification of 1. Two points are considered to conjugate of each other if they are images of each other.

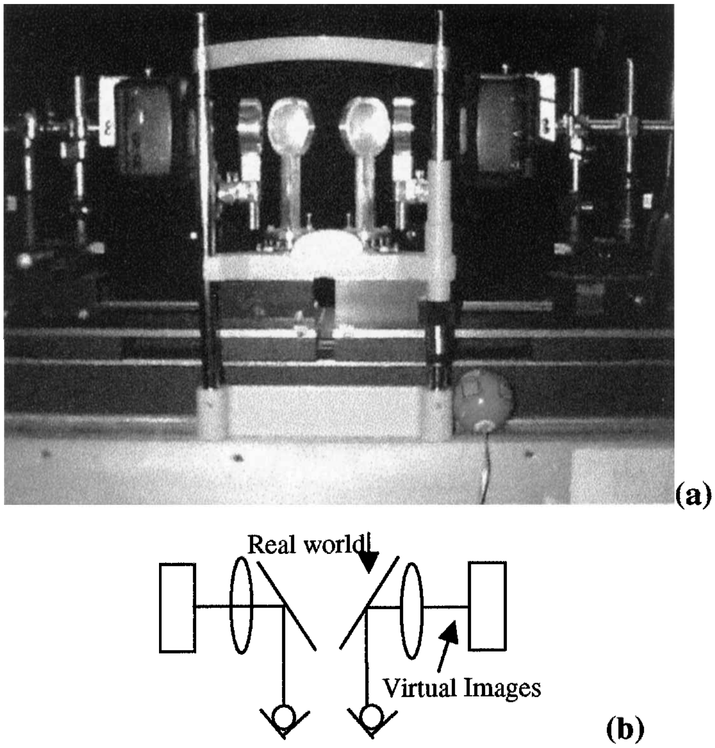


Figure 14. *(a) Bench prototype head-mounted display with head-motion parallax developed in the VGILab at UCF (1997). (b) Schematic of the optical imaging from a top view of the setup.*

optical residual distortions (Robinett & Rolland, 1992; Rolland & Hopkins, 1993; Watson & Hodges, 1995).

3.2.2.6 Perceived Location of Objects in Depth. Once depths are accurately rendered according to a given computational model and the stereo images are presented according to the computational model, the perceived location and size of objects in depth become an important issue in the assessment of the technology and the model. Accuracy and precision can be defined only statistically. Given an ensemble of measured perceived location of objects in depths, the depth percept will be accurate if objects appear in average at the location predicted by the computational model. The perceived location of objects in depth will be precise if objects appear within a small spatial zone around that average location. We shall distinguish between overlapping and nonoverlapping objects.

In the case of nonoverlapping objects, one may resort to depth cues other than occlusion. These include familiar sizes, stereopsis, perspective, texture, and motion parallax. A psychophysical investigation of the perceived location of objects in depth in an optical see-through HMD using stereopsis and perspective as the visual cues to depth is given in Rolland et al., (1995), and Rolland et al. (1997). The HMD shown in figure 14 is mounted on a bench for calibration purpose and flexibility in various parameter settings.

In a first investigation, a systematic shift of 50 mm in the perceived location of objects in depth versus predicted values was found in this first set of study (Rolland et al., 1995). Moreover, the precision of the measures varied significantly across subjects. As we learn more about the interface optics and computational model used in the generation of the stereo image pairs and improve on the technology, we have demonstrated errors on the order of 2 mm. The technology is now ready to deploy for extensive testing in specific applications, and the VRDA tool is one of the applications we are currently pursuing.

Studies of the perceived location of objects in depth for overlapping objects in an optical see-through HMD have been conducted by Ellis and Buchler (1994). They showed that the perceived location of virtual objects can be affected by the presence of a nearby opaque physical object. When a physical object was positioned in front of (or at) the initial perceived location of a 3-D virtual object, the virtual object appeared to move closer to the observer. In the case in which the opaque physical object was positioned substantially in front of the virtual object, human subjects often perceived the opaque object to be transparent. In current investigations with the VRDA tool, the opaque leg model appears transparent when a virtual knee model is projected on the leg as seen in figure 4. The virtual anatomy subjectively appears to be inside the leg model (Baillot, 1999; Outters et al., 1999; Baillot et al., 2000).

3.2.3 Adaptation. When a system does not offer what the user ultimately wants, two paths may be taken: improving on the current technology, or first studying the ability of the human system to adapt to an imperfect technological unit and then developing adaptation training when appropriate. This is possible because of the astonishing ability of the human visual and proprioceptive systems to adapt to new environments, as has been shown in studies on adaptation (Rock, 1966, for example).

Biocca and Rolland (1998) conducted a study of adaptation to visual displacement using a large-FOV video see-through HMD. Users see the real world through two cameras that are located 62 mm higher than and 165 mm forward from their natural eyepoints as shown in figure 2. Subjects showed evidence of perceptual adaptation to sensory disarrangement during the course of the study. This revealed itself as improvement in performance over time while wearing the see-through HMD and as negative aftereffects once they removed it. More precisely, the negative aftereffect manifested itself clearly as a large overshoot in a depth-pointing task, as well as an upward translation in a lateral pointing task after wearing the HMD. Moreover, some participants experienced some early signs of cybersickness.

The presence of negative aftereffects has some potentially disturbing practical implications for the diffusion of large-FOV video see-through HMDs (Kennedy & Stanney, 1997). Some of the intended earlier users of these HMDs are surgeons and other individuals in the medical profession. Hand-eye sensory recalibration for highly skilled users (such as surgeons) could have potentially disturbing consequences if the surgeon were to enter surgery within some period after using an HMD. It is an empirical question how long the negative aftereffects might persist and whether a program of gradual adaptation (Welch, 1994) or dual adaptation (Welch, 1993) might minimize the effect altogether. In any case, any shift in the camera eyepoints need to be minimized as much as possible to facilitate the adaptation process that is taking place. As we learn more about these issues, we will build devices with less error and more similarity between using these systems and a pair of eyeglasses (so that adaptation takes less time and aftereffects decrease as well).

A remaining issue is the conflict between accommodation and convergence in such displays. The issue can be solved at some cost (Rolland, et al., 2000). For lower-end systems, a question to investigate is how users adapt to various settings of the technology. For high-end systems, much research is still needed to understand the importance of perceptual conflicts and how to best minimize them.

3.2.4 Peripheral FOV. Given that peripheral vision can be provided in both optical and video see-through systems, the next question is whether it is used effectively for both systems. In optical see-through, there is almost no transition or discrepancy between the real scene captured by the see-through device and the peripheral vision seen on the side of the device.

For video see-through, the peripheral FOV has been provided by letting the user see around the device, as with optical see-through. However, it remains to be seen whether the difference in presentation of the superimposed real scene and the peripheral real scene will cause discomfort or provide conflicting cues to the user. The issue is that the virtual displays call for a different accommodation for the user than the real scene in various cases.

3.2.5 Depth of Field. One important property of optical systems, including the visual system, is depth of field. (Depth of field refers to the range of distances from the detector (such as the eye) in which an object appears to be in focus without the need for a change in the optics focus (such as eye accommodation). For the human visual system example, if an object is accurately focused monocularly, other objects somewhat nearer and farther away are also seen clearly without any change in accommodation. Still nearer or farther objects are blurred. Depth of field reduces the necessity for precise accommodation and is markedly influenced by the diameter of the pupil. The larger the pupil, the smaller the depth of field. For a 2 mm and 4 mm pupil, the depths of field are ± 0.06 and ± 0.03 diopters, respectively. For a 4 mm pupil, for example, such a depth of field translates as a clear focus from 0.94 m to 1.06 m for an object 1 m away, and from 11 m to 33 m for an object 17 m away (Campbell, 1957; Moses, 1970). An important point is that accommodation plays an important role only at close working distances, where depth of field is narrow.

With video see-through systems, the miniature cameras that acquire the real-scene images must provide a depth of field equivalent to the required working distance for a task. For a large range of working distances, the camera may need to be focused at the middle working distance. For closer distances, the small depth of field may require an autofocus instead of a fixed-focus camera.

With optical see-through systems, the available depth of field for the real scene is essentially that of the human visual system, but for a larger pupil than would be accessible with unaided eyes. This can be explained by the brightness attenuation of the real scene by the half-transparent mirror. As a result, the pupils are dilated (we assume here that the real and virtual scenes are matched in brightness). Therefore, the effective depth of field is slightly less than with unaided eyes. This is a problem only if the user is working with nearby objects and the virtual images are focused outside of the depth of field that is required for nearby objects. For the virtual images and no autofocus capability for the 2-D virtual images, the depth of field is imposed by the human visual system around the location of the displayed virtual images.

When the retinal images are not sharp following some discrepancy in accommodation, the visual system is constantly processing somewhat blurred images and tends to tolerate blur up to the point at which essential detail is obscured. This tolerance for blur considerably extends the apparent depth of field so that the eye may be as much as ± 0.25 diopters out of focus without stimulating accommodative change (Moses, 1970).

3.2.6 Qualitative Aspects. The representation of virtual objects, and in some cases of real objects, is altered by see-through devices. Aspects of perceptual representation include the shape of objects, their color, brightness, contrast, shading, texture, and level of detail. In the case of optical see-through HMDs, folding the optical path by using a half-transparent mirror is necessary because it is the only configuration that leaves the real scene almost unaltered. A thin, folding mirror will introduce a small apparent shift in depth of real objects precisely equal to $e(n - 1)/n$, where e is the thickness of the plate and n is its index of refraction. This is in addition to a small amount of distortion ($< 1\%$) of the scene at the edges of a 60 deg. FOV. Consequently, real objects are seen basically unaltered.

Virtual objects, on the other hand, are formed from the fusion of stereo images formed through magnifying optics. Each optical virtual image formed of the display associated with each eye is typically optically aberrated. For large-FOV optics such as HMDs, astigmatism and chromatic aberrations are often the limiting factors. Custom-designed HMD optics can be analyzed from a visual performance point of view (Shenker, 1994; Rolland, 2000). Such analysis allows the prediction of the expected visual performance of HMD users.

It must be noted that real and virtual objects in such systems may be seen sharply by accommodating in different planes under most visualization settings. This yields conflicts in accommodation for real and virtual imagery. For applications in which the virtual objects are presented in a small working volume around some mean display distance (such as arm-length visualization), the 2-D optical images of the miniature displays can be located at that same distance to minimize conflicts in accommodation and convergence between real and virtual objects. Another approach to minimizing conflicts in accommodation and convergence is multifocal planes technology as described in Rolland et al., 2000).

Beside brightness attenuation and distortion, other aspects of object representation are altered in video see-through HMDs. The authors' experience with at least one system is that the color and brightness of real objects are altered along with the loss in texture and levels of detail due to the limited resolution of the miniature video cameras and the wide-angle optical viewer. This alteration includes spatial, luminance, and color resolution. This is perhaps resolvable with improved technology, but it currently limits the ability of the HMD user to perceive real objects as they would appear with unaided eyes. In wide-FOV video see-through HMDs, both real and virtual objects call for the same accommodation; however, conflicts of accommodation and convergence are also present. As with optical see-through HMDs, these conflicts can be minimized if objects are perceived at a relatively constant depth near the plane of the optical images. In narrow-FOV systems in which the real scene is seen in large part outside the overlay imagery, conflicts in accommodation can also result between the real and computer-generated scene.

For both technologies, a solution to these various conflicts in accommodation may be to allow autofocus of the 2-D virtual images as a function of the location of the user gaze point in the virtual environment, or to implement multifocal planes (Rolland et al., 2000). Given eye-tracking capability, autofocus could be provided because small displacements of the miniature display near the focal plane of the optics would yield large axial displacements of the 2-D virtual images in the projected virtual space. The 2-D virtual images would move in depth according to the user gaze point. Multifocal planes also allow autofocusing but with no need for eye tracking.

4 Conclusion

We have discussed issues involving optical and video see-through HMDs. The most important issues are system latency, occlusion, the fidelity of the real-world view, and user acceptance. Optical see-through systems offer an essentially unhindered view of the real environment; they also provide an instantaneous real-world view that assures that visual and proprioception information is synchronized. Video systems forfeit the unhindered view in return for improved ability to see real and synthetic imagery simultaneously.

Some of us working with optical see-through devices strongly feel that providing the real scene through optical means is important for applications such as medical visualization in which human lives are at stake. Others, working with video see-through devices feel that a flip-up view is adequate for the safety of the patient. Also, how to render occlusion of the real scene at given spatial locations may be important. Video see-through systems can also guarantee registration of the real and virtual scenes at the expense of a mismatch between vision and proprioception. This may or may not be perceived as a penalty if the human observer is able to adapt to such a mismatch. Hybrid solutions, such as that developed by Peuchot (1994), including optical see-through technology for visualization and video technology for tracking objects in the real environment, may play a key role in future developments of technology for 3-D medical visualization.

Clearly, there is no "right" system for all applications: Each of the tradeoffs discussed in this paper must be examined with respect to specific applications and available technology to determine which type of system is most appropriate. Furthermore, additional HMD features such as multiplane focusing and eye tracking are currently investigated at various research and development sites and may provide solutions to current perceptual conflicts in HMDs. A shared concern among scientists developing further technology is the lack of standards not only in the design but also most importantly in the calibration and maintenance of HMD systems.

Acknowledgments

This review was expanded from an earlier paper in a SPIE proceeding by Rolland, Holloway, and Fuchs (1994), and the authors would like to thank Rich Holloway for his earlier contribution to this work. We thank Myron Krueger from Artificial Reality Corp. for stimulating discussions on various aspects of the technology, as well as Martin Shenker from M.S.O.D. and Brian Welch from CAE Electronics for discussions on current optical technologies. Finally, we thank Bernard Peuchot, Derek Hill, and Andrei State for providing information about their research that has significantly contributed to the improvement of this paper. We deeply thank our various sponsors not only for their financial support that has greatly facilitated our research in see-through devices but also for the stimulating discussions they have provided over the years. Contracts and grants include ARPA DABT 63-93-C-0048, NSF Cooperative Agreement ASC-8920219; Science and Technology Center for Computer Graphics and Scientific Visualization, ONR N00014-86-K-0680, ONR N00014-94-1-0503, ONR N000149710654, NIH 5-R24-RR-02170, NIH 1-R29LM06322-O1A1, and DAAH04-96-C-0086.

References

Allen, D. (1993). A high resolution field sequential display for head-mounted applications. *Proc. IEEE virtual reality annual international symposium* (VRAIS'93) (pp. 364–370).

Azuma, R., & Bishop, G. (1994). Improving static and dynamic registration in an optical see-through HMD. *Computer Graphics: Proceedings of SIGGRAPH '94* (pp. 197–204).

Bajura, M., Fuchs, H., & Ohbuchi R. (1992). Merging virtual objects with the real world, *Computer Graphics, 26,* 203–210.

Bajura, M., & Newmann H. (1995). Dynamic registration correction in video-based augmented reality systems. *IEEE Computer Graphics and Applications, 15*(5), 52–60.

Baillot, Y., & Rolland, J. P. (1998). Modeling of a knee joint for the VRDA tool. *Proc. of Medicine Meets Virtual Reality* (pp. 366–367).

Baillot, Y., Rolland, J. P., & Wright, D. (1999). Kinematic modeling of knee-joint motion for a virtual reality tool. In *Proc. of Medicine Meets Virtual Reality* (pp. 30–35).

Baillot, Y., Rolland, J. P., Lin, K., & Wright, D. L. (2000). Automatic modeling of knee-joint motion for the Virtual Reality Dynamic Anatomy (VRDA) Tool. *Presence: Teleoperators and Virtual Environments, 9*(3), 223–235.

Baillot, Y. (1999). First implementation of the Virtual Reality Dynamic Anatomy (VRDA) tool. Unpublished *master's thesis, University of Central Florida.*

Barrette, R. E. (1992). Wide field of view, full-color, high-resolution, helmet-mounted display. *Proc. of SID '92 Symposium* (pp. 69–72).

Biocca, F. A., & Rolland, J. P. (1998). Virtual eyes can rearrange your body: Adaptation to virtual-eye location in see-thru head-mounted displays. *Presence: Teleoperators and Virtual Environments, 7*(3), 262–277.

Bishop, G., Fuchs, H., McMillan, L., & Scher-Zagier, E. J. (1994). Frameless rendering: Double buffering considered harmful. *Proc. of SIGGRAPH'94* (pp. 175–176).

Brooks, F. P. (1992). Walkthrough project: Final technical report to National Science Foundation Computer and Information Science and Engineering. Technical Report TR92-026, University of North Carolina at Chapel Hill.

Buchroeder, R. A., Seeley, G. W., & Vukobratovich, D. (1981). Design of a catadioptric VCASS helmet-mounted display. Optical Sciences Center, University of Arizona, under contract to U.S. Air Force Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio, AFAMRL-TR-81-133.

Campbell, F. W. (1957). The depth of field of the human eye. *Optica Acta, 4,* 157–164.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving the layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of Space and Motion* (pp. 69–117). Academic Press.

Deering, M. (1992). High resolution virtual reality. *Proc. of SIGGRAPH'92, Computer Graphics, 26*(2), 195–201.

Desplat, S. (1997). Characterization des elements actuals et future de loculometre Metrovision-Sextant. Technical Report, Ecole Nationale Superieure de Physique de Marseilles, November.

Dhome, M., Richetin, M., Lapreste, J. P., & Rives, G. (1989). Determination of the attitude of 3-D objects from a single perspective view. *IEEE Trans. Pattern Analysis and Machine Intelligence, 11*(12), 1265–1278.

Droessler, J. G., & Rotier, D. J. (1990). Tilted cat helmet-mounted display. *Optical Engineering, 29*(8), 849–854.

Edwards, E. K., Rolland, J. P., & Keller, K. P. (1993). Video see-through design for merging of real and virtual environments. *Proc. of IEEE VRAIS'93* (pp. 223–233).

Edwards, P. J., Hawkes, D. J., Hill, D. L. G., Jewell, D., Spink, R., Strong, A., & Gleeson, M. (1995). Augmentation of reality using an operating microscope for otolaryngology and neurosurgical guidance. *J. Image Guided Surgery, 1*(3), 172–178.

Ellis, S. R., & Bucher, U. J. (1994). Distance perception of stereoscopically presented virtual objects optically superimposed on physical objects in a head-mounted see-through display. *Proc. of the Human Factors and Ergonomic Society, Nashville.*

Fernie, A. (1995). A 3.1 Helmet-mounted Display with Dual Resolution. CAE Electronics, Ltd, Montreal, Canada, SID 95 *Applications Digest,* 37–40.

Fernie, A. (1995). Improvements in area of interest helmet-mounted displays. In *Proc. of SID95.*

Fry, G. A. (1969). *Geometrical Optics.* Chilton Book Company.

Furness, T. A. (1986). The super cockpit and its human factors challenges. *Proceedings of the Human Factors Society 30,* 48–52.

Held, R., & Durlach, N. (1987). Telepresence, time delay and adaptation. *NASA Conference publication 10032.*

Holloway, R. (1995). An analysis of registration errors in a see-through head-mounted display system for craniofacial surgery planning. *Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.*

Hua, H., Girardot, A., Gao, C., and Rolland, J.P. (2000). Engineering of head-mounted projective displays. *Applied Optics* (in press).

Jacobs, M. C., Livingston, M. A., & State, A. (1997). Managing latency in complex augmented reality systems. *Proc. of 1997 Symposium on Interactive 3D Graphics, ACM SIGGRAPH,* 235–240.

Kaiser Electro-Optics. (1994). Personal communication from Frank Hepburn of KEO, Carlsbad, CA. General description of Kaiser's VIM system ("Full immersion head-mounted display system") is available via ARPA's ESTO home page: http://esto.sysplan.com/ESTO/.

Kancherla, A., Rolland, J. P., Wright, D., & Burdea, G. (1995). A novel virtual reality tool for teaching dynamic 3D anatomy. *Proc. of CVRMed '95,* 163–169.

Kandebo, S. W. (1988). Navy to evaluate Agile Eye helmet-mounted display system. *Aviation Week & Space Technology,* August 15, 94–99.

Kennedy, R. S., & Stanney, K. M. (1997). Aftereffects in virtual environment exposure: Psychometric issues. In M. J. Smith, G. Salvendy, & R. J. Koubek (Eds.), *Design of Computing.*

Kijima, R., & Ojika, T. (1997). Transition between virtual environment and workstation environment with projective head-mounted display. *Proc. of VRAIS '97,* 130–137.

Manhart, P. K., Malcom R. J., & Frazee, J. G. (1993). Augeye: A compact, solid Schmidt optical relay for helmet mounted displays. *Proc. of IEEE VRAIS '93,* 234–245.

Mizell, D. (1998). Personnal communication.

Moses, R. A. (1970). *Adlers Physiology of the Eye.* St. Louis, MO: Mosby.

Oster, P. J., & Stern, J. A. (1980). "Measurement of Eye Movement," in I. Martin & P. H. Venables (Eds.), *Techniques of Psychophysiology,* New York: John Wiley & Sons.

Outters, V., Argotti, Y., & Rolland, J. P. (1999). Knee motion capture and representation in augmented reality. Technical Report TR99-006, University of Central Florida.

Parsons, J., & Rolland, J. P. (1998). A non-intrusive display technique for providing real-time data within a surgeons critical area of interest. *Proc. of Medicine Meets Virtual Reality,* 246–251.

Peuchot, B. (1993). Camera virtual equivalent model: 0.01 pixel detectors. *Special issue on 3D Advanced Image Processing in Medicine in Computerized Medical Imaging and Graphics, 17*(4/5), 289–294.

———. (1994). Utilization de detecteurs subpixels dans la modelisation d'une camera—verification de l'hypothese stenope, *9e Congres AFCET, reconnaissance des formes et intelligence artificielle, Paris.*

———. (1998). Personal communication.

Peuchot, B., Tanguy, A., & Eude, M. (1994). Dispositif optique pour la visualization d'une image virtuelle tridimensionelle en superimposition avec an object notamment pour des applications chirurgicales. *Depot CNRS #94106623,* May 31.

———. (1995). Virtual reality as an operative tool during scoliosis surgery. *Proc. of CVRMed '95,* 549–554.

Robinett, W., & Rolland, J. P. (1992). A computational model for the stereoscopic optics of a headmounted display. *Presence: Teleoperators and Virtual Environments, 1*(1), 45–62.

Rock, I. (1966). *The nature of perceptual adaptation*. New York: Basic Books.

Rolland, J. P. (1994). Head-mounted displays for virtual environments: The optical interface. *International Optical Design Conference 94, Proc. OSA, 22,* 329–333.

Rolland, J. P. (1998). Mounted displays, *Optics and Photonics News, 9*(11), 26–30.

Rolland, J. P. (2000). Wide angle, off-axis, see-through head-mounted display. *Optical Engineering—Special Issue on Pushing the Envelop in Optical Design Software, 39*(7), (in press).

Rolland, J. P., Ariely, D., & Gibson, W. (1995). Towards quantifying depth and size perception in virtual environments. *Presence: Teleoperators and Virtual Environments, 4*(1), 24–49.

Rolland, J. P. & Arthur, K. (1997). Study of depth judgments in a see-through mounted display, *Proceeding of SPIE 3058, AEROSENSE,* 66–75.

Rolland, J. P., Holloway, R. L., & Fuchs, H. (1994). A comparison of optical and video see-through head-mounted displays. *Proceedings of SPIE* 2351, 293–307.

Rolland, J. P., & Hopkins, T. (1993). A method for computational correction of optical distortion in head-mounted displays. Technical Report TR93-045, University of North Carolina at Chapel Hill.

Rolland, J. P., Krueger, M., and Goon, A. (2000). Multifocus planes in head-mounted displays, *Applied Optics, 39*(19), in press.

Rolland, J. P., Parsons, J., Poizat, D., & Hancock, D. (1998). Conformal optics for 3D visualization, *Proc. of the International Lens Design Conference,* Hawaii (June), 760–764.

Rolland, J. P., Yoshida, A., Davis, L., & Reif, J. H. (1998). High-resolution inset head-mounted display, *Applied Optics, 37*(19), 4183–4193.

Roscoe, S. N. (1984). Judgments of size and distance with imaging displays, *Human Factors, 26*(6), 617–629.

Roscoe, S. N. (1991). The eyes prefer real images, in *Pictorial Communication in Virtual and Real Environments,* Ed. Stephen R. Ellis, Taylor and Francis.

Scher-Zagier, E. J. (1997). A human's eye view: motion, blur, and frameless rendering. *ACM Crosswords 97.*

Shenker, M. (1994). Image quality considerations for head-mounted displays. *Proc. of the OSA: International Lens Design Conference, 22,* 334–338.

Shenker, M. (1998). *Personal Communication.*

Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments, *Presence: Teleoperators and Virtual Environments, 6*(6), 603–616.

State, A., Chen, D., Tector, C., Brandt, C., Chen, H., Ohbuchi, R., Bajura, M., and Fuchs, H. (1994). Case study: Observing a volume rendered fetus within a pregnant patient, *Proceedings of Visualization '94. Washington, DC.,* 364–373.

State, A., Hirota, G., Chen, D. T., Garrett, W. E., and Livingston, M. (1996). Superior augmented-reality registration by integrating landmark tracking and magnetic tracking. *Proc. of SIGGRAPH 1996, ACM SIGGRAPH,* 429–438.

Sutherland, I. (1965). The ultimate display, *Information Processing 1965: Proc. of IFIP Congress 65,* 506–508.

Sutherland, I. E. (1968). A head-mounted three-dimensional display, *Fall Joint Computer Conference, AFIPS Conference Proceedings 33,* 757–764.

Thomas, M. L., Siegmund, W. P., Antos, S. E., and Robinson, R. M. (1989). Fiber optic development for use on the fiber optic helmet-mounted display, in *Helmet-Mounted Displays,* J. T. Colloro, Ed., Proc. of SPIE 1116, 90–101.

Tomasi, C., & Kanade, T. (1991). Shape and motion from image streams: a factorization method-Part 3: Detection and tracking of point features, *Carnegie Mellon technical report CMU-CS-91-132.*

Vaissie, L., & Rolland, J. P. (1999). Eye-tracking integration in head-mounted displays, Patent Filed January.

Vaissie, L., Rolland, J. P. (2000). Alberlían errors in head-mounted displays: choice of eyejoints location. *Technical Report TR 2000-001,* University of Central Florida.

Watson, B., & Hodges, L. F. (1995). Using texture maps to correct for optical distortion in head-mounted displays, *Proc. of VRAIS'95,* 172–178.

Welch, B., & Shenker, M. (1984). The fiber-optic Helmet-Mounted Display, *Image III,* 345–361.

Welch, R. B. (1993). Alternating prism exposure causes dual adaptation and generalization to a novel displacement. *Perception and Psychophysics, 54*(2), 195–204.

Welch, R. B. (1994). Adapting to virtual environments and teleoperators, *Unpublished manuscript, NASA-Ames Research Center, Moffett Field, CA.*

Welch, G., & Bishop, G. (1997). SCAAT: Incremental tracking with incomplete information, *Proc. of ACM SIGGRAPH,* (pp. 333–344).

Wright, D. L., Rolland, J. P., & Kancherla, A. R. (1995). Using virtual reality to teach radiographic positioning, *Radiologic Technology, 66*(4), 167–172.