A comparison of optical and video see-through head-mounted displays

J. P. Rolland, R. L. Holloway, and H. Fuchs

Department of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175
{rolland,holloway,fuchs}@cs.unc.edu

Abstract

One of the most promising and challenging future uses of head-mounted displays (HMDs) is in applications where virtual environments enhance rather than replace real environments. To obtain an enhanced view of the real environment, the user wears a see-through HMD to see 3D computer-generated objects superimposed on his/her real-world view. This see-through capability can be accomplished using either an optical or a video see-through HMD. We discuss the tradeoffs between optical and video see-through HMDs with respect to technological, perceptual, and human factors issues, and discuss our experience designing, building, using, and testing these HMDs.

Keywords: augmented reality, see-through head-mounted displays, superimposition, optics, video cameras

Introduction

One of the most promising and challenging future uses of head-mounted displays (HMDs) is in applications where virtual environments enhance rather than replace real environments. This is often referred to as *augmented reality.* To obtain an enhanced view of the real environment, the user wears a see-through HMD that allows her to see 3D computer-generated objects superimposed on her real-world view. This see-through capability can be accomplished using either an optical or a video see-through HMD, as shown in Figures 1 and 2, respectively.

With optical-see-through HMDs, the real world is seen through semi-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 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.

In the same way that x-ray technology has provided us with a means to see aspects of the inner world of objects and living beings that were unseen before, see-through virtual reality technology provides us with a novel way of visualizing those aspects: the dynamic superimposition of inner worlds registered on their outer parts.

See-through HMDs have been around since the 1960s with Ivan Sutherland's first HMD which was a see-through stereo system with miniature CRTs as the display devices, a mechanical tracker to provide head position and orientation in real time, and a hand-tracking device (Sutherland, 1965).

Most see-through systems that have been designed and/or built are of the optical see-through type.

Examples are the VCASS system (Furness 1986), (Buchroeder et al., 1981), the Tilted Cat HMD (Droessler and Rotier, 1990), and the CAE Fiber-Optic HMD (Barrette, 1992). Several systems have been developed by Kaiser Electronics and McDonnell Douglas, all of them optical see-through systems (Kandebo, 1988).

See-through HMDs at the University of North Carolina at Chapel Hill (UNC-CH) have been developed since the 1980s. Applications with both video and optical see-through HMDs have been developed. The driving application for video see-through has been real-time 3D visualization of a human fetus during ultrasound echography (Bajura et al., 1992).[1] An augmented reality system displaying live ultrasound data in real time and properly registered in 3D space within a scanned subject would be a powerful and intuitive tool; it could be used for needle-guided biopsies, obstetrics, cardiology, etc. With the current system, the ultrasound images captured in real time appear to be pasted in front of the patient's body, as shown in Figure 3, rather than fixed within it. Figure 4 shows a recent near-real-time implementation where the fetus is rendered more convincingly within the body (State et al., 1994).

Another medical application of see-through HMD technology at UNC is a system developed for cranio-facial surgery planning. The eventual goal of this system is to superimpose CT skull data onto the head of the real patient, thereby giving the surgeon "x-ray vision", or the ability to "see" the patient's bone through the soft tissue. The promise of this system is that viewing the data *in situ* will allow surgeons to make better surgical plans because they will be able to see the complex relationships between the bone and soft tissue more clearly. Because of the precision required for this task, however, most of the work in this project has focused on identifying and quantifying the sources of registration error between the real and virtual objects (Holloway, 1994).

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

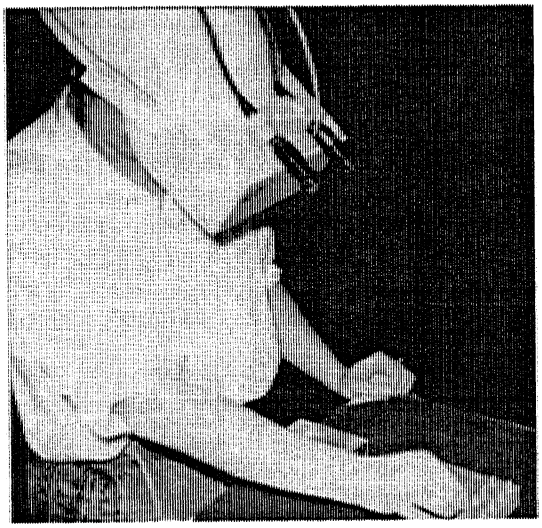

Figure 2. A video see-through head-mounted display.

---

[1] To our knowledge, the first video see-through system with walk-around capability in addition to merging of real and synthetic information was developed at UNC-CH. The authors would be grateful for any information on other video STHMD systems.

An example in the non-medical arena is the work of Caudell and Mizell at Boeing, which uses optical see-through HMD technology to assist workers in the manufacturing assembly of airplanes (Caudell and Mizell, 1992). Combined with head-position sensing and a real-world registration system, the see-through technology in this case allows computer-generated diagrams to be superimposed and stabilized on specific locations of a real-world foamboard. The main difficulty with a complex manufacturing assembly task is the need to have sufficient registration of real and virtual information so that workers may perform their jobs without any risk of errors due to limitations in the apparatus. The successful use of the technology will enable cost reductions and efficiency improvements in several of the manual operations in aircraft manufacturing.

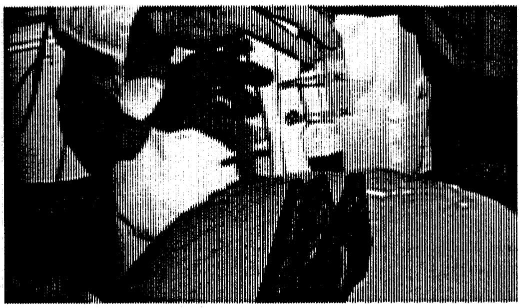

Figure 3. Real-time acquisition and superimposition of ultrasound slice images on a pregnant woman.

Another example is that of providing assistance with complex maintenance tasks (Feiner et al., 1993). The system developed at Columbia University uses a knowledge-based graphics component in order to aid the user in an end-user laser printer maintenance task. They use graphics superimposed on the laser writer to provide information on various tasks, and use tracker sensors on the printer's moving parts to reflect movements of the real-world objects in the virtual scene.

Figure 4. More convincing rendering of fetus inside abdomen.

A more recent application being developed at UNC-CH is the visualization of dynamic 3D anatomy using an optical see-through HMD. This work aims at developing a novel teaching tool to impart better understanding of bone dynamics during radiographic positioning for the radiological science student (Kancherla et al., 1994). In our first prototype, we concentrate on the positioning of the arm about the elbow joint. We use an optical see-through HMD coupled with tracking devices positioned along the arm of the patient to visualize the 3D anatomy directly superimposed on the patient as illustrated in Figure 5. We want to simulate "x-ray vision" as much possible. Therefore, we want the patient's arm to be seen as one would with unaided eyes, but enhanced with computer-simulated bones. This implies that the real scene must be very realistic, the field of view (FOV) should be sufficient to capture the whole arm, and the tracking devices and image-generation system must be fast enough to track motion in real time. The biggest challenge is dynamic registration of the real arm with the virtual bones.

As alluded to with the description of these example applications, 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 is fooled into perceiving those 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 tradeoffs. The goal of this paper is to list and discuss these tradeoffs in order to illustrate under which conditions one might choose one type of see-through system over the other.

In both systems there are two image sources: the real world and the computer-generated world; these two image sources are to be merged. Optical see-through HMDs take what might be called a "minimally invasive" 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 more invasive in the sense that they block out the real-world view in exchange for the ability to merge the two views more convincingly. The fundamental tradeoff, then, is whether the additional features afforded by the more invasive approach justify the loss of the unobstructed real-world view.

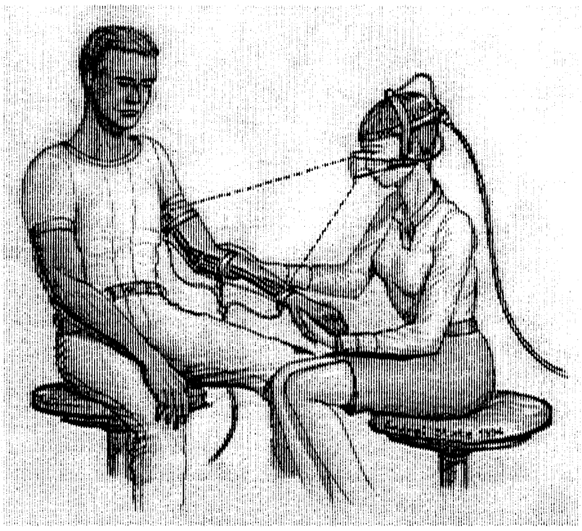

Figure 5. Teaching tool for dynamically superimposing virtual bones onto real arm (drawing by Andrei State).

We discuss the tradeoffs between optical and video see-through HMDs with respect to technological, perceptual, and human factors issues from our experience designing, building, using, and testing these HMDs. Those tradeoffs are also discussed with respect to systems built today, those buildable with today's technology, and those perhaps buildable 5 to 10 years from now. An outline of the issues discussed in the paper is illustrated in Figure 6.

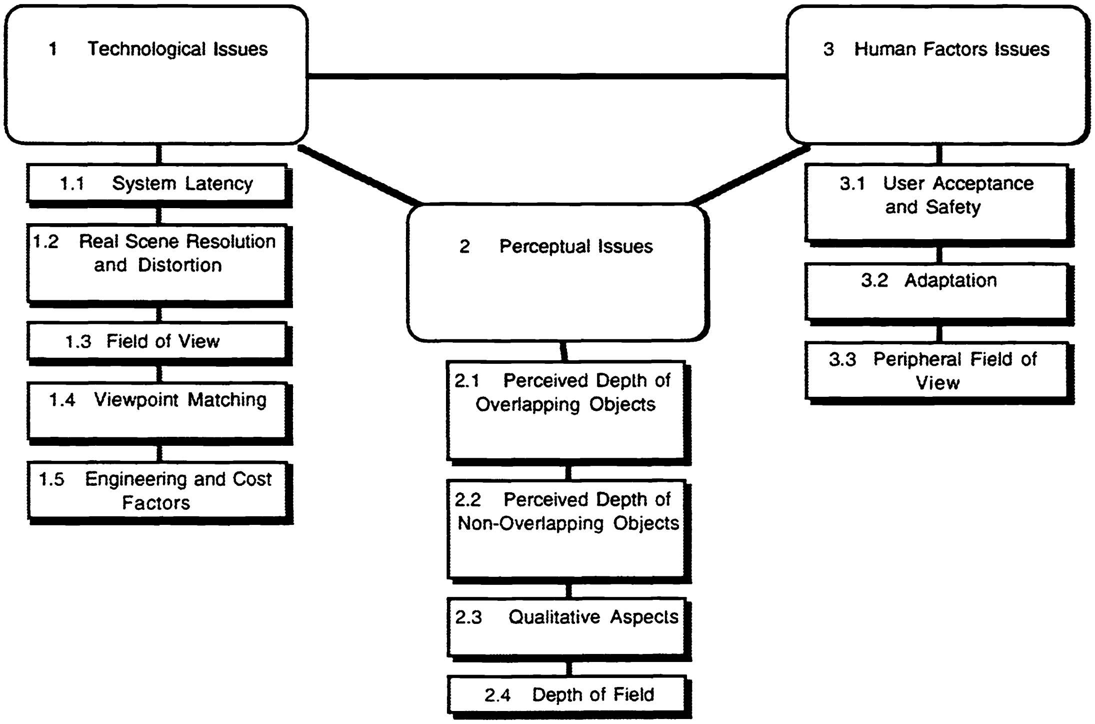

Figure 6. Paper Outline

1. Technological Issues

The technological issues discussed in this section include latency of the system, resolution and distortion of the real scene, FOV of the see-through device, and engineering and cost factors.

1.1. System Latency

An essential component of see-through HMDs is the capacity to properly register the user's surroundings and the synthetic space. Assuming geometric calibration between tracking and the HMD optics, the major impediment to achieving this 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 completed and presented to the user. Lag is the largest source of registration error in most current HMD systems (Holloway, 1994).

This lag in typical systems is between 60 and 180 ms. The user's head can move significantly during such a period of time, and the discrepancy destroys the illusion of the synthetic objects being fixed in the environment. The synthetic objects "swim" around so much, in fact, that they don't even seem to be part of the environment. Azuma and Bishop demonstrated this swimming effect and show how one can minimize it using predicted HMD positions instead of measured ones (Azuma and Bishop, 1994). They use an optical tracking system augmented with accelerometers and rate gyros and have found that their system can predict well 60 ms ahead, but not well 120 ms ahead.

One of the major advantages of video see-through HMDs is the capability of enforcing registration of the real and synthetic images. In other words, because the system has access to both the real and synthetic images, it can manipulate them in space or in time in order to register them.

The spatial approach to forcing registration in video see-through systems is to correct registration errors by tracking landmark points in the real-world images and registering virtual objects with respect to those. This technique was successfully demonstrated by (Bajura & Neumann, 94) who tracked a pair of red light-emitting diodes (LEDs) placed on two real objects and then registered two virtual objects with respect to them. By tracking more landmarks, better registration of real and virtual objects can be achieved. The problem with this approach is that it is an attempt to register three-dimensional scenes using two-dimensional 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: one must either allow misregistration of some of the landmarks or perform a non-linear warping of the virtual scene in order to align all of the landmarks (which may induce undesirable distortions of virtual objects). The non-trivial 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 2D transformations.

Enforcing registration in the time domain can be done by introducing delay to the video real-scene images to match the delay of the computer-generated images. This may be done in conjunction with the spatial methods just mentioned. If the delay is too large, however, it may cause sensory conflicts between vision and proprioception, since video images no longer correspond to the real-world scene. Any manual interactions with real objects could suffer as a result. 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, video see-through HMDs will normally be slightly delayed with respect to the real world even without adding delay to match the synthetic images. This delay may become even worse if an image-processing step for enforcing registration or performing occlusion is added. The key issue, of course, is whether the delay in the system is too great for the user to adapt to it. This subject has been treated at length in the teleoperation literature; (Held and Durlach, 1987) discuss this problem and give an extensive list of references on the subject.

Systems using optical see-through HMDs have no way to introduce artificial delays to the real scene, so the system needs to be optimized for low latency; that is, less than 60 ms according to (Azuma and Bishop, 1994). For any remaining lag, the user can restrain himself to using slow head motions or predictive tracking can be used as described earlier. The advantage, however, of not introducing artificial delays is that real objects will always be where they are perceived to be, and this may be crucial for certain applications.

1.2. Real-Scene Resolution and Distortion

The best real-scene resolution a see-through-device can provide is that reached with the naked eye. This assumes 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 an optical see-through HMD because the optical interface to the real world is simply a thin parallel 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 diameter 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. Such an aberration is negligible compared to the ability the human eye to resolve a visual angle of 1 minute of arc. Similarly, planar plates introduce very little distortion of the real scene, typically below 1%. There is no distortion for those chief rays that pass the plate parallel to the plate's normal.[2]

In the case of video see-through HMDs, real-scene images are digitized by miniature cameras and converted to an analog signal which is fed to the HMD displays. These images are then magnified by the HMD viewing optics which typically use eyepiece design. The perceived resolution of the real scene can thus be limited by the resolution of either of the three components of the system: the video cameras, the HMD displays, or the HMD viewing optics.

Currently available miniature video cameras typically have a resolution of 720x480, which is also near the resolution limit of the miniature displays currently used in HMDs[3]. Both the miniature displays and the video cameras currently limit the resolution of most systems.

In addition, eyepiece designs have also been recognized for many decades to be extremely limited in optical quality due to their fairly large FOVs and their need for an exit pupil size large enough to accommodate the size of the pupils of the 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.

A new technology, referred to as *tiling*, may overcome the current limitations of conventional eyepiece design (Kaiser, 94). 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 now have a fairly narrow FOV, high image quality can be achieved. The challenge is in the assembly process and in rendering seamless views from multiple displays.

Theoretically, distortion is not a problem in video see-through systems since 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 increases the complexity of the synthetic image generation significantly (State et al., 1994). An alternative is to use ultra-low distortion cameras, merge unprocessed real scenes with virtual scenes, and warp the merged images to compensate for the distortion of the HMD viewing optics as a last step.

---

[2] The chief ray is defined as a ray that passes through a point in the FOV and through the center of the pupils of the sytem. The exit pupil in an HMD is the entrance pupil of the human eye.

[3] The number of physical elements is typically 720 x 240, and one uses signal processing to interpolate between lines to get 720 x 480.

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 any loss in real-scene resolution. Current video see-through systems are seriously limited in terms of resolution, but with new tiling technologies and the growing availability of high-resolution flat-panel displays, video see-through HMDs may match the resolution of the human visual system five years from now.

1.3. Field of View

Another challenging issue is that of providing the user with a large FOV. What we refer to here is the monocular FOV at a given time, head position and orientation, unless otherwise specified. The binocular FOV is a function of the monocular FOV's overlap and can be calculated for each system based on its geometry. For 100% overlap, the binocular FOV is equal to the monocular FOV, while for no overlap, it is twice the monocular FOV. A decrease in binocular overlap, however, results in a decrease of the binocular fusion zone where stereo vision is possible. For most applications, having a larger binocular FOV means that fewer head movements are required to perceive an equivalently large scene. However, in many cases, one would prefer to have a large binocular FOV without trading off the amount of binocular overlap that is necessary for stereo vision. In those cases, the monocular FOV itself must be optimized. 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. We will use the term *overlay FOV* to describe the merged virtual and real FOV.

Optical see-through HMDs typically provide from 20° to 60° overlay FOV via the semi-transparent mirror placed in front of the eyes, a value which may appear somewhat limited. Larger FOVs have been obtained, up to 82.5 x 67 degrees (Welch and Shenker, 1984), at the expense of reduced brightness, increased complexity, and massive, expensive technology. Optical see-through HMDs, however, have been designed open enough that the user can use his/her peripheral vision around the device, thus increasing the total real-world FOV to numbers that match closely 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 eye- or sun-glasses.

Video see-through HMDs, on the other hand, can provide as large a see-through FOV as can be displayed with the viewing optics. Typical values range from 20° to 90°. However, because current HMDs used in video see-through are of the opaque type where the peripheral FOV of the user is occluded, the effective real-world FOV is often smaller than in optical see-through systems. We found in a recent human-factors study using a video see-through system that users needed to perform larger head movements to scan an active field of vision required for a task than when they used the unaided eye (Rolland at al., 1994). We predict that the need to make larger head movements would not arise as much with see-through HMDs with large peripheral FOVs.

An increase in peripheral FOV in video see-through systems can be accomplished in two ways: 1) in a folded optical design as used for optical see-through HMDs but with an opaque mirror instead of a semi-transparent mirror; 2) in a non-folded design but with non-enclosed mounts, which calls for innovative opto-mechanical design since heavier optics have to be supported than in either optical or folded video see-through. Folded systems only require a thin mirror in front of the eyes, and the heavier optical components are placed around the head. The tradeoff with folded systems, however, is a significant reduction in the overlay FOV (as experienced with optical see-through HMDs).

Most current high-resolution HMDs achieve their higher resolution at the expense of a reduced FOV. That is, they use smaller, high-resolution CRTs and optics with less magnification in order to achieve higher angular resolution, but the resulting FOV may be too narrow for many applications. A solution to this problem is again, perhaps, to move towards tiling techniques to improve resolution without trading FOV.

These techniques also bring new practical and computational challenges that need to be confronted. In particular, for a tiled video see-through system, one either needs multiple, correlated cameras for each eye, or an ultra-wide-FOV camera for each eye whose image can then be segmented somehow. Thus, while tiling is a promising approach for wide-FOV, high-resolution HMDs, it also introduces some challenging technical problems for use in video see-through HMD systems.

1.4 Viewpoint Matching

The viewpoint of a camera or eye is equivalent to the center of projection used in the computer graphics model and is taken here to be the center of the entrance pupil of the eye or camera. A problem with video see-through HMDs is that the straightforward method of mounting the cameras introduces an error in the viewpoints for the real-world images.

The straightforward method of mounting the cameras on a video see-through HMD is to separate them by the appropriate interpupillary distance (IPD) and mount them on the front or top of the HMD, or to mount them on the side of the HMD, in which case they will be separated by more than the IPD. The problem with this approach is that the viewpoint of the camera does not correspond to the viewpoint of the eye, which introduces a shift in the perceived scene for each eye, which may lead to perceptual anomalies (see the Human Factors section for more on this).

Since the camera cannot be physically placed at the actual eyepoint, one alternative is to use mirrors to fold the optical path (much like a periscope) in order to make the camera viewpoint correspond to the real eyepoint. The following figure illustrates this idea.

While this approach solves the viewpoint-shift problem, it increases the optical path length 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 FOV for correct real-world viewpoints or require the user to adapt to the shifted viewpoints (more on this later).

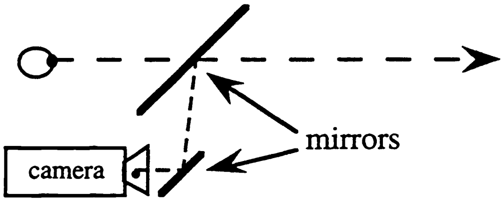

Figure 7. Folded path to make viewpoints for real and virtual scenes optically identical

1.5. Engineering and Cost Factors

Simple HMD designs suffer from low resolution, limited FOV, poor ergonomic design, and heavy weight. To overcome any of these limitations, one must face new challenges and further tradeoffs. A good ergonomic design implies that either the HMD is light enough that it does not weigh much more than a pair of eyeglasses or the HMD folds around the user's head so that the center of gravity of the device falls more or less near the center of rotation of the head (Rolland, 1994). This assures maximum comfort and usability. Reasonably light HMD designs currently suffer very narrow FOVs, on the order of 20 degrees. As far as we know, there are currently no large FOV see-through HMDs of any type that are comparable in weight to a pair of eyeglasses.

With optical see-through, the folding can be accomplished with either an on-axis or an off-axis design. Off-axis designs are not only more elegant but also by far more attractive since they free the user from seeing ghost images that plague current on-axis designs. The reason we cannot buy off-axis designs is that very few prototypes have been built and those that have been built have been classified. Moreover, off-axis systems are difficult to design and have turned out to be hard to build (Shenker, 1994). Moreover, they are usually more expensive to design and build due to the cost of off-axis and/or non spherical optical elements that come with such systems. A non-classified off-axis design was designed by Rolland (1994) at UNC-CH. Multiple factors, including cost, have so far prohibited us from building the first prototype.

Video see-through systems usually do not use a folded design and therefore suffer from a lack of peripheral FOV, since they use rigid helmets to hold the optics and displays in front of the user's eyes. With tiling techniques, no folding is required since the large FOV that drove the need for complex optics, and thus that of folding it, is now created by multiple narrow FOV devices that are a lot simpler optically. It is an open question whether such devices can be successfully built. Moreover, the cost of such a device may be significantly higher due to the need for multiple miniature high-resolution displays, precision alignment of the displays, and image-generation hardware to drive them.

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). In the next five to ten years, we predict that high-resolution color flat-panel displays will be the first choice for HMDs. It has been predicted for years that CRTs would become obsolete, yet they are still used today. The reason for renewed optimism is the support given to the development of flat panel technologies by government agencies such as ARPA and the existence of a market for high-resolution miniature displays in video camcorders.

2. Perceptual Issues

The ultimate system is one that provides quantitative and qualitative visual representation of scenes that conforms to that given by the real world. This includes 1. accuracy and precision of depth, 2. accuracy and precision of size, of real and virtual objects in a scene, and 3. an unobstructed peripheral FOV which is important for many tasks from situation awareness to simple manipulation of objects and accessories. In this section, perceived depth and size, qualitative aspects of the real scene, and depth of field issues will be discussed. Peripheral FOV issues will be discussed in the human factors section.

Perceived depth and size of objects are important because they determine the 3D space they occupy. Perceived depth will be accurate and precise if objects appear on average at the location predicted by the computational model used to position virtual objects. Perceived depth will be precise if objects appear within a small spatial zone around that average location. A strong component of rendering depth accurately is occlusion of overlapping objects. Because the ability to perform occlusion is an important issue of comparison between optical and video see-through HMDs, we discuss it first.

2.1. Perceived Depth of Overlapping Objects

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. Moreover, occlusion is a strong monocular cue to depth perception and may be required in certain applications.

In both systems, any attempt to do 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 progress in this area is being made (Tomasi & Kanade 1991),(Laveau & Faugeras 1994), the problem is far from solved. Thus, occlusion cues for either type of display will be limited by the state of the art in this area. We can now move on to a discussion of the tradeoffs with respect to occlusion for each type of see-through HMD.

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 out 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 is 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 cause real objects to occlude virtual ones simply by 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.

This technique is currently used by CAE Electronics in their flight simulator: When the pilot looks out the window, s/he sees computer-generated objects. If s/he looks inside the cockpit, however, the appropriate pixels of the computer-generated image are masked so s/he can see the real instruments. They keep the room fairly dark so that this technique will work (Barrette, 1992). Tom Caudell and David Mizell from Boeing Seattle are also using this technique; they refer to it as "fused reality" (Mizell, 1994).

While optical see-through HMDs can allow real objects to occlude virtual objects, the reverse is not easy to do since normal beam splitters have no way of selectively blocking out the real environment. There are at least two possible solutions to this problem, neither of them perfect. The first solution is to 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 certain applications, but probably not for all. 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. The problem with this approach is that the user does not focus on the beam splitter, but rather on the virtual image it creates. Thus, any blocking done at the beam splitter will be out of focus, 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.

2.2. Perceived Depth of Non-Overlapping Objects

In the case of non-overlapping objects, one may resort to depth cues other than occlusion. These include familiar sizes, stereopsis, perspective, texture, and motion parallax. A psychophysical investigation of perceived depth using stereopsis and perspective as the visual cues to depth in a virtual environment is given in (Rolland et al., 1994). The apparatus consisted of an optical see-through bench prototype HMD. Some of the main results were that a systematic, but relatively small shift, in perceived depth from predicted values was found, but that the precision of the measures varied significantly across subjects. Data were only reported for three subjects, and further data need to be acquired to confirm those findings. Some measures of perceived size were conducted by Roscoe and colleagues in see-through HMDs and a main result was that objects seemed to be perceived smaller than they actually were (Roscoe, 1984, 1991). It is an experimental question whether current computational models correctly predict perceived depth and sizes in virtual environments.

For video see-through HMDs, the problems with viewpoint and FOV matching between the displays and the cameras can introduce additional errors in perceived depth. Although it is possible in theory to eliminate such mismatches, real systems are often limited by the available technology and therefore leave it to the user to adapt to the discrepancy. The effect of viewpoint mismatches will be discussed in the section on human factors.

2.3. 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, the real objects are seen basically unaltered since they are only perceived through an ultra-thin plate of glass, as mentioned earlier.

Other aspects of objects' representation besides brightness attenuation and distortion may be altered in video see-through HMDs. Our experience with at least one system is that real objects' color and brightness are altered along with the loss in texture and levels of detail due to the limited resolution (this includes spatial, luminance, and color resolution) of the miniature video cameras and wide angle optical viewer present in the system (Rolland et al., 1994). This is perhaps resolvable with improved technology but currently limits the ability of the HMD user to perceive real objects as they would appear with unaided eyes. We do not foresee a breakthrough in the next five years, but perhaps in the next ten years.

For optical see-through systems, folding the optical path by using a semi-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 shift in depth of real objects 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 of the scene at the edges of the FOV, as discussed earlier.

2.4. Depth of Field

One important property of the human visual system not yet addressed is *depth of field*. Depth of field refers to the range of distances from the eye in which an object appears to be in focus without the need for a change in eye accommodation. If an object is accurately focused monocularly, other objects somewhat nearer and further away are also seen clearly without any change in accommodation. Still nearer or further away objects are blurred. Depth of field reduces the necessity for precise accommodation, but 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 depth of field is +/- 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 to 1.06 meter for an object 1 meter away, and to 11 meters to 33 meters for an object 17 meters away (Campbell, 1957; Moses, 1970). The important point to note is that accommodation may play an important role only at close working distance where depth of field is narrow.

With video see-through systems, the miniature cameras used for acquiring 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 semi-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 will be slightly less than with unaided eyes. This is only a problem if the user is working with nearby objects and the virtual images are focused outside of the depth of field required for nearby objects. With current displays, the virtual images suffer so much from blur due to the poor resolution of the displays themselves that the need to accommodate at different planes, for real and virtual objects to be seen clearly, may not even come into play.

Since retinal images are never truly sharp, 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 extends the apparent depth of field considerably, so that the eye may be as much as +/-0.25 diopters out of focus without stimulating accommodative change (Moses, 1970). From this observation, convergence and accommodation in a HMD may not be as decoupled as one theoretically would predict. This would occur at the expense of more blurry images that may be properly handled by the visual system. Whether these factors affect depth perception is worth further investigation.

3. Human Factors Issues

Some of the issues discussed in this section currently have few quantitative proofs because too few human factors studies have been conducted with VR technology in general, and even less with see-through VR technology. We shall therefore offer our opinions based on our experience.

3.1. 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 application- and technology-specific, but will probably boil down to the issue of whether the advanced capabilities afforded by the technology offset the problems induced by the encumbrance and sensory conflicts associated with it. In particular, video see-through HMDs may meet with resistance in the workplace since they take away the direct real-world view in order to augment it; there is an issue of trust that may be difficult to overcome for some users. Moreover, current problems with low resolution and camera/eye viewpoint mismatches will tend to push the acceptance threshold further into the future for video see-through systems.

The 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 failing in technology in the case of optical see-through may leave the subject without any computer-generated images but still with the real world view. In the case of video see-through, a failing in technology may leave the user with complete suppression of the real-world view, as well as the computer-generated view. Such a loss may be unacceptable in many applications for reasons of safety and liability.

3.2. Adaptation

When a system does not offer what the user ultimately wants, two paths may be taken: 1) improving on the current technology, or 2) studying the ability of the human system to adapt to an imperfect technological unit and 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 multiple studies on adaptation (Rock, 1966).

We recently conducted a study of adaptation to visual displacement in see-through HMDs (Rolland & Biocca et al., 1994). Users see the real world through two cameras which are located 62 mm higher, and 165 mm forward from their natural eyepoints. 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.

The presence of negative after-effects has some potentially disturbing practical implications for the diffusion of see-through HMDs. 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 like surgeons could have potentially disturbing consequences if the surgeon were to enter surgery within some period after use of a 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 any adaptation process that is taking place.

3.3. Peripheral FOV

Given that peripheral vision can be provided for both optical and video see-through systems, the next question is whether it is used as 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 can be also captured by video cameras or may be provided by letting the user see around the device, as with optical see-through. Especially in the latter case, 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.

Conclusion

We have presented a comparison of optical and video see-through head-mounted displays. In our opinion, the most important issues are system latency, occlusion, and the fidelity of the real-world view. Optical see-through systems offer an essentially unhindered view of the real environment; they also provide an instantaneous real-world view which assures that visual and proprioception information is synchronized. Video systems give up the unhindered view in return for improved ability to render occlusion cues; the issue of how to really perform occlusion is far from solved, however, and remains an active area of research. Video see-through systems can also guarantee registration of the real and virtual scenes at the expense of a mismatch between vision and proprioception, which may or may not be perceived as a penalty if the human observer is able to adapt to such a mismatch. Clearly, there is no "right" system for all applications: each of the tradeoffs discussed in this paper must be examined with respect to the planned application and available technology to determine which type of system is most appropriate.

Acknowledgments

The authors would like to thank David Mizell and Donna Wright for stimulating discussions about this work.

This research was supported by the following grants: ARPA #DABT 63-93-C-0048, NSF Cooperative Agreement #ASC-8920219, and ARPA: "Science and Technology Center for Computer Graphics and Scientific Visualization", ONR #N00014-86-K-0680, #N00014-94-1-0503, and NIH #5-R24-RR-02170.

References

Allen, D. (1993)., "A 1" high resolution field sequential display for head-mounted applications. " Proc IEEE Virtual reality annual international symposium (VRAIS).

Azuma, R., and G. Bishop (1994), "Improving static and dynamic registration in an optical see-through HMD." *Computer Graphics: Proceedings of SIGGRAPH '94.* (Orlando, July 24-29, pp. 197-204).

Bajura, M., H. Fuchs, and R. Ohbuchi (1992) “Merging Virtual Objects with the Real World,” Computer Graphics, number 26, pp. 203 -10.

Barrette, R.E. (1992), “Wide Field of View, Full-Color, High-Resolution, Helmet-Mounted Display,” SID ’92 Symposium, pp. 69-72.

Buchroeder, R.A., Seeley, G.W., and Vukobradatovich, 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.

Caudell, T.P., and Mizell, D.W. (1992), “Augmented Reality: An application of heads-up display technology to manual manufacturing processes,” Proc. of the 1992 IEEE Hawaii International Conference on Systems Sciences, pp. 659-669.

Droessler, J.G., and Rotier, D.J. (1990), “Tilted cat helmet-mounted display,” Optical Engineering, 29 (8).

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

Feiner, S., B. Macintyre, D. Seligmann (1993), “Knowledge-based augmented reality,” Communications of the ACM, 36 (7), 53-62.

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

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

Holloway, Richard. (1994), “An Analysis of Registration Errors in a See-Through Head-Mounted Display System for Craniofacial Surgery Planning”, Ph.D. dissertation, University of North Carolina at Chapel Hill. In preparation.

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 World-Wide Web home page: http://esto.sysplan.com/ESTO/.

Kancherla, A., J.P. Rolland, D.L. Wright, and G. Burdea (1995), “A novel virtual reality tool for teaching dynamic 3D anatomy”. In press.

Kandebo, S.W. (1988), “Navy to evaluate Agile Eye helmet-mounted display system,” Aviation week & space technology, August 15, 94-99.

Laveau, S., and O. Faugeras. (1994), “3-D scene representations as a collection of images and fundamental matrices,” Institut National de Recherche en Informatique et en Automatique (INREA) Report 2205. February.

Mine, Mark. (1993), “Characterization of end-to-end delays in head-mounted display systems”, Technical Report TR93-001, University of North Carolina at Chapel Hill.

Rolland, J.P. (1994), “Head-mounted displays for virtual environments: the optical interface,” presented at the International Optical Design Conference 94, Proc. OSA 22 (in press).

Rolland, J.P., F. Biocca, T. Barlow, and A. Kancherla (1994), “ Quantification of Perceptual Adaptation to Visual Displacement in See-Through Head-Mounted Displays,” forthcoming.

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

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.

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

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

Tomasi, C., and T. Kanade. (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.

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

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

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