Jannick P. Rolland
and
William Gibson
Department of Computer Science,
CB 3175
University of North Carolina
Chapel Hill, North Carolina 27599

Dan Ariely
Department of Psychology, CB 3270
University of North Carolina
Chapel Hill, North Carolina 27599

Towards Quantifying Depth and Size Perception in Virtual Environments

Abstract

With the rapid advance of real-time computer graphics, head-mounted displays (HMDs) have become popular tools for 3D visualization. One of the most promising and challenging future uses of HMDs, however, is in applications where virtual environments enhance rather than replace real environments. In such applications, a virtual image is superimposed on a real image. The unique problem raised by this superimposition is the difficulty that the human visual system may have in integrating information from these two environments. As a starting point to studying the problem of information integration in see-through environments, we investigate the quantification of depth and size perception of virtual objects relative to real objects in combined real and virtual environments. This starting point leads directly to the important issue of system calibration, which must be completed before perceived depth and sizes are measured. Finally, preliminary experimental results on the perceived depth of spatially nonoverlapping real and virtual objects are presented.

1 Introduction

Head-mounted displays (HMDs) have become popular tools for 3D visualization following the rapid advance of real-time computer graphics. They provide 3D information to the user by presenting stereoscopic images to his eyes, similar to a simple slide stereoscope. The main difference is that the two images are scanned on two head-mounted miniature displays and can be updated in real time using fast computer graphics. This real-time update feature allows the user to see different stereoscopic images as he moves his head around, thus creating at each point in time new 3D views of a computer-generated, also referred to as a virtual, environment around him. An important idea behind the use of HMDs is that if this virtual environment is an exact model of a real scene, the user's experience as he moves in this virtual environment will be the same as the experience of moving in the real scene. One of the most promising and challenging future uses of HMDs, however, is in applications where virtual environments enhance rather than replace real environments. In such applications, the user wears a see-through HMD that allows him to see 3D computer-generated objects superimposed on his view of the real world.

See-through capability can be accomplished using a video or an optical-see-through HMD. In video see-through systems, the real world view is captured by two miniature video cameras mounted on head gear and the computer-generated image is added to the video representation of the real world electroni-

Presence, Vol. 4, No. 1, Winter 1995, 24–49

cally (Bajura, Fuchs, & Bubuchi, 1992; Edwards, Rolland & Keller, 1993). In optical-see-through systems, on the other hand, the real world is seen through semi-transparent mirrors placed in front of the user's eyes. These semi-transparent mirrors are also used to reflect the computer-generated images into the user's eyes. Because of their dual function, these semi-transparent mirrors are also referred to as optical combiners. Depending on the problem at hand, it may be argued that one modality is preferred over the other, however, it is in itself a subject of investigation. Because optical-see-through HMD technology provides direct visual access to the real 3D world, we chose optical-see-through HMD as the tool to study the spatial properties of combined real and virtual 3D environments. A detailed description of head-mounted display issues for both video and optical-see-through modalities is beyond the scope of this paper but is treated elsewhere (Rolland, Holloway, & Fuchs, 1994).

The problem addressed here concerns the human visual system's ability to integrate computer-generated 3D environments with the real 3D environments when presented simultaneously. When computer-generated 3D images are viewed in isolation as on a CRT display without interposing optics, the observer will see both the image on the display and the frame surrounding it, with depth cues provided by both. If the computer-generated 3D image is viewed through a see-through HMD, the observer will see a virtual image of the display, without its frame, superimposed on her view of the real environment. In this case, the intermingling of the real and computer-generated 3D worlds brings more complexity to the visualization problem. For example, the perceived spatial relationships between real and virtual objects may not follow theoretical predictions. Moreover, the spatial relationships may be unstable, virtual objects being perceived at times closer than expected and at other times further away, depending on the visual cues to depth provided (Foley, 1991). Within the complex problem of information integration in see-through HMDs, the research reported here focuses on the specific problem of how spatial relationships are altered when information from real and computer-generated images must be integrated.

One of the fundamental problems of optical-see-through HMDs is the inadequate registration of the real and virtual images. It is all the more important that the human visual system is highly sensitive to small differences in depth and lateral displacements (Gulick & Lawson, 1976; Patterson & Martin, 1992; Patterson, Moe, & Hewitt, 1992). Several serious issues, such as the time lag with which HMD displays track head movements and the inability of optical see-through devices to simulate occlusion, are equally important for the use of see-through HMDs. These topics are not being treated here but deserve investigations on their own. The research reported here addresses the problem of careful control of all relevant parameters in static displays as well as the subset of the calibration issues relevant to this research. A more detailed and general investigation of possible sources of errors in HMDs is in progress in our research group by Holloway and colleagues (1994) and will be reported elsewhere. Finally, we measured the resultant percept of depth of virtual objects as compared to real objects using our calibrated system and present experimental results.

Previous studies describing the judgments of size and distance with imaging displays have been aimed at understanding the observed misperceptions experienced by pilots flying airplanes and airplane simulators. The motivation for these studies has been the optimization of imaging displays for piloted vehicles and for pilot training simulators. In these experiments, pilots made landings by reference to panel-mounted periscope screens in airplanes and to virtual (collimated) computer-graphics images in flight simulators. They consistently misjudged the runway as being smaller and farther away than it was and consequently tended to overshoot their landings (Roscoe, Olzak, & Randle, 1966; Palmer & Cronn, 1973; Randle, Roscoe, & Petitt, 1980).

Several articles by Stanley Roscoe describe some of the problems found with both directly viewed display screens and collimated virtual images (Roscoe, 1948, 1979, 1984, 1985, 1987, 1991, 1993). For both types of displays, Roscoe and his associates have found a decrease in apparent sizes and an associated increase in apparent distances relative to the apparent sizes and distances of objects subtending the same visual angles

viewed directly in natural environments (Campbell, McEachern, & Marg, 1955; Roscoe, Hasler, & Dougherty, 1966; Hull, Gill, & Roscoe, 1982; Roscoe, 1984).

In another series of related experiments, Roscoe and colleagues measured the perceived size of a monocular collimated disk subtending a fixed visual angle (0.67°) viewed against various backgrounds at various distances (Benel, 1979; Benel & Benel, 1979; Simonelli, 1979; Hull, Gill, & Roscoe, 1982; Iavecchia, Iavecchia, & Roscoe, 1983). In all these experiments, visual accommodation (the distance at which the eyes focused) was measured, and a systematic relationship was found between the apparent size of the collimated disk and the focal distance of the eyes (r = 0.89 to 0.98 for the central tendencies). These findings have been found to apply to other tasks that differ in stimuli and viewing conditions, but an open question is whether they apply universally.

In the case of collimated head-mounted virtual displays, each of the two monocular images is collimated. From a purely optical point of view, one would expect that pilots' eyes would accommodate (focus) at optical infinity to form sharp images of the displays on their retina. Research has shown, however, that eyes do not automatically focus at optical infinity in response to collimated displays (Cogan, 1937; Mandelbaum, 1960; Leibowitz & Owens, 1975a, 1975b; Roscoe, Olzak, & Randle, 1976; Benel, 1979; Benel & Benel, 1979; Simonelli, 1980; Randle, Roscoe & Petitt, 1980; Hull, Gill, & Roscoe, 1982; Norman & Ehrlich, 1986; Iavecchia, Iavecchia, & Roscoe, 1988; Kotulak & Morse, 1994). Rather, when viewing collimated images, the eye tends to focus in the proximity of its resting point of accommodation (its dark focus), and that can be anywhere between a few inches from the eye to as much as two or three diopters beyond optical infinity, depending only slightly on the textural context of the scene (Iavecchia, Iavecchia, & Roscoe, 1988).

Roscoe (1993) offers two hypotheses, each of which may contribute to the misperceptions of size and distance with collimated imaging displays. The first is that inward accommodation from optical infinity reduces the effective angular size of the retinal image and thus the perceived size of the whole visual scene as well as individual objects within the scene. The second, advanced previously by Lockhead and Wolbarsht (1989), is that the neural signals to the eye muscles that control accommodation serve to scale the apparent size of the incoming images. In support of the latter, Lockhead and Wolbarsht found that the decrease in apparent size of distant virtual objects was still observed with visually impaired people having no accommodative range. Other researchers, notably Enright (1989), suggest that the coordinated change in vergence, accommodation, as well as changes in pupil diameter may explain the phenomenon.

Thus, Roscoe's original idea that accommodation may account for increases and decreases in apparent size and distance may not be the sole explanation for the decrease in perceived size and increase in perceived distance of virtually imaged objects. In fact, Lockhead and Wolbarsht argue that "actual accommodation of the lens is not needed, but the optical and probably also the neural cues to accommodation are needed" (Lockhead & Wolbarsht, 1989).

The results of the studies reviewed reinforce the importance of determining the effect of image collimation on perceived depths and sizes of objects in a see-through HMD. In the case of nearby objects, accommodation and convergence in the real environment are loosely coupled, whereas, they must be uncoupled for virtual 3D objects if the monocular images are collimated and the two images are to be fused. This paper addresses the question of how image collimation affects perception of depth in a see-through head-mounted display that integrates real and virtual images.

Other studies have also reported that optically generated displays presenting synthetic images of the outside world have the characteristic of producing systematic errors in size and depth judgments (Sheehy & Wilkinson, 1989; Hadani, 1991). Sheely and Wilkinson's studies show that the lateral phoria of the observer might change with the prolonged wearing of night vision goggles. Lateral phoria refers to latent tropia (eyes not pointing in the same direction), which is apparent only when all fusible contours are eliminated. Sheely and Wilkinson conclude that this phenomena can be at least partially accounted for by the inaccurate positioning of the device on the head of the user, especially the use of

incorrect interpupillary distance (IPD) for the observer. Hadani, on the other hand, points out that absolute depth estimations are strongly biased when using the device itself. Moreover, this bias is pronounced for nearby objects, which look closer than they really are. Hadani gives as an explanation of this bias the fact that the center of perspective of the goggle (defined by the objective lens) does not coincide with the user's eyes. We shall discuss how both IPD and entrance pupils of the eye are important parameters to accurate depth perception in HMDs as well.

In our research laboratories, some discrepancies between the theoretical model of some environments and its subjective perception in the HMD have been experienced. An example is the case of navigation in a walk-through experiment using a non-see-through type of HMD with about 90° binocular field of view (FOV). Several subjects reported that the floor of the virtual kitchen seemed lower than they expected and behaved as if they had to take a step down to move forward, when no steps were present or visible. This percept could have been the product of inadequate calibration because this effect was not reported systematically by all subjects. For example, no interpupillary distance adjustments were available on the system. Instead a value of 62 mm was used. It is possible that this value was correct for some of the subjects involved and not for the others.

The resulting distorted space could also be due, at least in part, to the weight and inertia of the HMD itself, which can affect subjects differently (Lackner & Dizio, 1989). Objects' characteristics such as the types of texture, illumination, and color used can also distort depth perception. Following any such observations, the question always arises as to how much of the failure of the virtual world to act as the real world is due to inadequate calibration (both intrinsic parameter setup and inaccurate positioning of the HMD on the user's head), poor head tracking, or perceptually imperfect images. The research reported here focuses on perceptual issues in HMDs because the question is whether or not a discrepancy is important to the human visual system.

One of the difficulties in measuring the accuracy of the user's percept of spatial relationships in an HMD comes from the fact that the current dynamic capability of HMDs is severely limited (Adelstein, Johnston, & Ellis, 1992; Ellis, Tyler, Kim, & Stark, 1992). A major problem with current technology is that there is a time delay (lag) in the presentation of the images to the users' eyes caused by the time that it takes to update the images with newly acquired head positions. Any alteration in the measured relationships between real and virtual objects could simply be an artifact of the system's lag. On the other hand, the capability for dynamic viewing has incredible potential because head motion parallax is known to provide strong cues to depth perception (Rogers & Graham, 1979). Until the lag limitation becomes less severe, however, the measure of veridicality of perceived depths and sizes in HMDs can be approached by looking at either static objects or self-rotating objects. Experimental results with static images are presented in this paper.

To summarize, we propose in this paper to quantify depth and size perception of virtual objects in combined real and virtual static environments. Toward this goal, important parameters within the computational model used to generate the stereo images are evaluated. The optical interface is strongly emphasized and the authors hope to bring insight into its important role. A set of optical terms is first given. Each component of the experimental setup is then outlined, potential sources of inaccuracies in displaying the stereo images that would result in some incorrect location of the objects in virtual space are identified, and the calibration procedure performed is given. Finally, the experimental paradigm used to assess perceived depth of generically shaped objects (real or virtual) in combined real and virtual environments is described and experimental results are presented.

2 Definitions

Aperture Stop: A physical constraint, often a lens retainer, that limits the diameter of the axial light bundle allowed to pass through a lens.

Aperture: An opening or hole through which radiation or matter may pass.

Apex: For a rotationally symmetric optical element,

the intersection of the optical axis with any optical surface of the element.

Chief Ray: Any ray that passes through the center of the entrance pupil, the pupil, or the exit pupil is referred to as the chief ray.

Collimated Images: Given an object and its image through a lens, the image is said to be collimated if it is at infinity. The object will thus be at the object focal point of the lens.

Entrance Pupil: In a lens or other optical system, the image of the aperture stop as seen from the object space.

Exit Pupil: In a lens or other optical system, the image of the aperture stop as seen from the image space.

Field of View (FOV): The maximum area that can be seen through a lens or an optical instrument. One must distinguish between the FOV calculated based on paraxial ray-tracing, which should be the one specified in the computational model of computer-generated images, and the FOV calculated from real ray-tracing that may be smaller due to some field stops, which is analogous to clipping planes in computer graphics. Optical distortion as well will cause the FOV derived from real ray-tracing to be smaller or larger than the paraxial FOV.

Focal Length: The distance from one of the principal points to its corresponding focal point.

Focal Point: That point on the optical axis of a lens to which an incident bundle of parallel light rays will converge.

Nodal Points: The intersection with the optical axis of the two conjugate planes of unit positive angular magnification.

Principal Planes: The two conjugate planes in an optical system of unit positive linear magnification.

Principal Points: The intersection of the principal plane and the optical axis of a lens.

Pupil: (1) In the eye, the opening in the iris that permits light to pass and be focused on the retina; (2) in a lens, the image of the aperture stop as seen from object and image space.

3 Computational Model

To generate accurate stereoscopic images for an HMD, one must consider the multiple potential sources of errors that can occur at several stages of the computation. First, the antialiased computational model assumes that the eyes of the user can be reduced to a single point, referred to as the eyepoint or center of perspective. Moreover, it further assumes that if the virtual images of the display screens are centered on the eyes of the subject and the center pixels of the two display screens are turned on, the lines joining the eyepoint to the virtual optical images of those bright pixels will be parallel. The point is optically specified to be at infinity and the model assumes that the subject will see it there if viewed binocularly. Any 3D point of light at a finite distance can be rendered by turning on the appropriate pixels on each display such that the lines of sight of the eyes passing through the virtual images of those pixels converge at that distance. This concept, very simple in nature, is actually far from simple to implement with accuracy in an HMD. A more complete description of the computational model used in HMDs can be found in Robinett and Rolland (1992). A computational model that details the appropriate transformation matrices in the most general case where the eyes may be decentered with respect to the virtual images of the display screens, and the optical axis of the viewer may be nonparallel can be found in Robinett and Holloway (1994).

Some of the sources of inaccuracies in depth location of 3D virtual objects come, in the most extreme case, from literally ignoring the optical system used to magnify the images, and in less extreme cases, underestimating or overestimating the FOV, inaccurately specifying the center of perspective, and finally ignoring the variations in interpupillary distance (IPD) between subjects. The approach of tweaking the software parameters until things look about right is sometimes all one needs to get a subjective percept of depth. Such an approach, however, is far too imprecise for virtual environment applications where there is a need to register real and virtual information (Bajura, Fuchs, & Ohbuchi, 1992; Janin, Mizell, & Caudell, 1993). In this case, an accurate calibration becomes absolutely necessary. Although accurate

calibration is especially important when virtual and real environments are combined, as with see-through HMDs, accurate calibration benefits non-see-through HMDs as well.

Because visual information is the focus here, an HMD can be thought of as a four-component system: screens displaying visual information, optical magnifiers forming virtual images of the display screens, the eyes of the user looking at the virtual images, and the head-tracking device updating visual information in real time. The software must model the hardware components involved in creating the images presented to the eyes in order to generate the correct set of stereoscopic images on the retinas. This includes the characteristics of the frame buffer where the images are rendered, the display-screens' parameters, some specifications of the imaging optics, and the position of the optics with respect to the eyes of the observer. An important point to note is that the spatial relationship of these components relative to each other is important to the calibration of the system. If not only visual information but also auditory, tactile, force feedback, and olfactory information were considered, additional hardware components would be needed but some of the calibration issues would remain the same.

The next sections detail components of the system. First, the imaging process taking place between the miniature display screens and the user's eyes is explained. Then the impact of the display screens' alignment and positioning on several perceptual issues are described. Next, the eyepoint for the system and how to set the FOV are addressed. Finally, the importance of the user's interpupillary distance is discussed and the effect of small head and eyes movements behind the optics is explored.

3.1 Image Formation

A brief description of the image-forming process will focus attention on three important elements: (1) the model of the optics in global computations, (2) operation of the optics, and (3) the way monocular images are transformed for each eye. A detailed review of the basics of geometric optics can be found in Longhurst (1973). Some of these principles useful for understanding HMDs are reviewed here.

Imaging optical systems can be described most generally by their principal points, nodal points, and focal length. First consider a monocular system. Given the focal length f' of the system (see Fig. 1), the position and size of the virtual image with respect to the second (image) principal point as well as the linear magnification (m) of the system can be determined using the paraxial imaging equations

$$\frac{n'}{x'} = \frac{n}{x} + \frac{n'}{f'} \tag{1a}$$

and

$$m = \frac{x'}{x} \tag{1b}$$

where n and n' are the indexes of refraction in object and image space, respectively, x is the signed distance from the object principal plane to the object plane, and x' is the signed distance from the image principal plane to the image plane. With the sign convention shown on Figure 1 by the plus sign, that is a positive distance goes from left to right, x' and f' are positive, while x is negative, in Eq. (1). The display screens and their corresponding images are said to be *conjugates* of one another. In the case of a binocular system, one would apply the same imaging scheme to each eye.

A real object AB, which is a display screen in this case, is imaged as a real image $A'B'$. Now, if the object AB is moved between the focal point F and the principal plane P, the image $A'B'$ in Figure 1 will become virtual as in HMDs. This is illustrated in Figure 3. Primes are used here to refer to quantities in the image space. Moreover, in the case of an HMD, n equal n' equal 1, and the nodal points are thus coincident with the principal points. A particular imaging case is when the display screen is placed at the focal plane passing through the focal point F. In this case, its virtual image $A'B'$ is located at infinity as shown in Figure 2, and the image is said to be collimated.

The size and position of the virtual image with respect to some stationary landmark (i.e., the second principal point, which can always be related to any lens apex in the

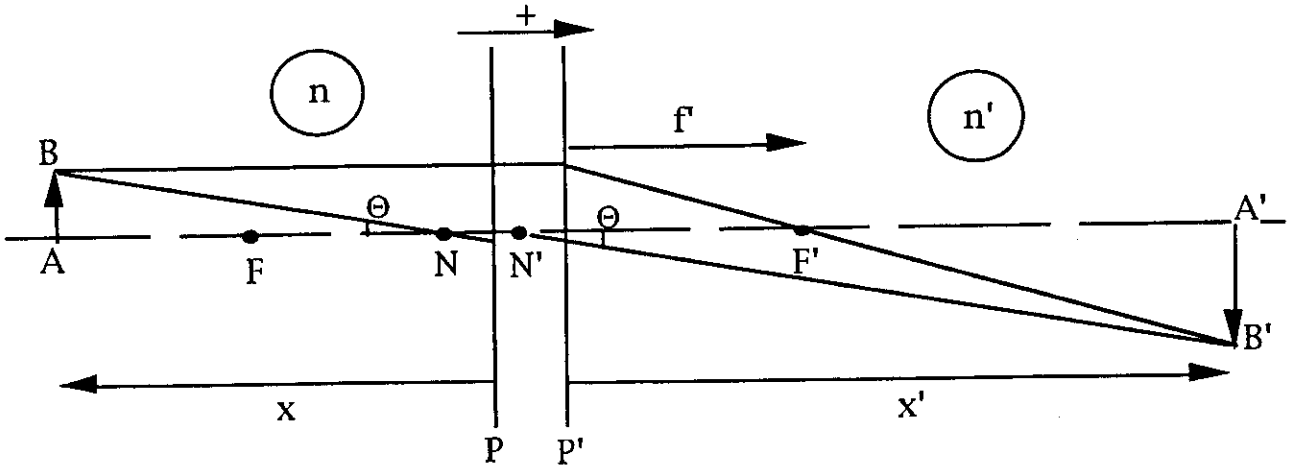

Figure 1. *The principle of image formation. An object AB is imaged through a lens to form an image A'B'. If prime denotes image space, the lens is represented by its principal planes P and P', its nodal points N and N', its index of refraction n and n', and its focal points F and F' in object and images space, respectively. Two rays, one parallel to the optical axis in object space and passing through F' in image space, and the other passing through conjugate points N and N' with angular magnification of 1, are used to graphically find the image A'B' of AB. The location x' of A'B' with respect to P' expressed as a function of x, f', n, and n' is given by Eq. (1a).*

optical system) are the parameters needed to be able to calculate the correct FOV of the system as described in Section 3.4. Calculating the FOV cannot be done by using solely the display screen size and the distance of the user's eye from the physical display screens themselves. Moreover, if the FOV is set incorrectly, objects in the 3D virtual space will be shifted forward or backward in depth.

Described here is the optical imaging between two conjugate planes, a display screen and its virtual image. It is also the case that a plane in image space characterized by x' in Eq. (1) can only be the conjugate of a single plane in object space characterized by x in Eq. (1). Some caution needs to be exercised, however, in terms of determining how these virtual images should be viewed as detailed in Section 3.3.

3.2 Display Screens Alignment and Positioning

The following discussion assumes (1) the eyes of the subject are centered on the optical axes of the lenses and (2) the eyes are reduced to one physical point that overlaps the theoretical center of projection used to calculate the stereo projections for each left and right image. The effect of the mislocation of the eyes is discussed in Sections 3.3 and 3.5. The computational depths are defined as the depth points predicted from the model, and the displayed depths are defined as the depths presented to the observer in the HMD. Note that both of these depths are different from the perceived depths.

Given the above assumptions, for the left and right images to be easily fused and computational depths to correlate accurately with the displayed depths, the displays viewport must be centered on the optical axes. This also maximizes the monocular FOVs. The two screens must be kept from unwanted rotations around the optical axis. Finally, the two screens must be magnified equally through the optics.

Any lateral misalignments of the display screens with respect to the optical axis will cause an error in depth location unless compensated for by the software by laterally shifting the images on the display by the calculated amount. An existing lateral misalignment may be due to displays physically too large to be centered on the optics, as in the VPL and Flight Helmets (Virtual Research), or mechanical limitations in mounting the displays centered on the optics with fractional-pixel precision. For the work reported here, an optical bench prototype HMD was constructed using off-the-shelf optical components.

This provided complete access to all the parameters of the system and eliminated lateral misalignments often found in commercially available systems. The 3-inch diagonal LCD displays used were too large to be mounted without optical offsets in a side-by-side geometry, as in the VPL Eyephone. As a result, an oblique geometry was adopted that allowed the displays to be centered on the optics within residual mechanical and assembly errors. A picture of the setup is shown in Figure 6. The exact amount of residual lateral mechanical misalignment cannot be determined a priori, but a calibration technique, derived from our computational model and described in detail in Section 4.2, was used to compensate for small errors.

A vertical misalignment of the displays with respect to the optics will either (1) shift the objects vertically within the FOV, if the amount of offset is equal for the two eyes, or (2) affect the ability to fuse the images due to induced *dipvergence* of the eyes of the observer if the amount of offset is different for each eye. While the eyes have a wide latitude in fusing images with lateral shifts, the slightest difference in vertical position causes eye strain and possibly headaches. Moreover, if the discrepancy is more than 0.34° visual angle the time needed to fuse the images increases exponentially (Burton & Home, 1980). In the particular case of the system used in this study, the vertical dimension of the display area was 40 mm, and an angle of 0.34° at the eyes between the two displays corresponds to a vertical displacement of one display with respect to the other of about 0.5 mm. Such a displacement corresponds roughly to 6 addressable vertical lines out of a total of 442 addressable lines. Experience with the alignment process has shown that achieving better than 6 addressable lines out of 442 is well within reach of the experimenters. This was confirmed by the fact that the images were very easily fusible and no headaches were ever reported by any of the subjects.

A rotation of the display screens around the optical axis can prevent the fusion of the two images for severe angles. In any case, it introduces distortion of the 3D objects being formed through stereo fusion. A detailed study of this error in HMDs is in progress (Holloway, 1994). This error is easily minimized by good mechanical assembly and visual inspection.

Finally, the positioning of the display screens with respect to the optics must be the same for the two eyes because the displays' position determines the magnification of the monocular images. A difference in magnification for the two eyes will cause a systematic error in depth location for the objects being viewed. This problem of different image magnifications has been well studied in visual science under the name of aniseikonia (Reading, 1983). It should be noted also that small displacements of the displays with respect to the optics can result in large displacements of the virtual images in image space. High precision mechanical assembly as well as measurements of where the virtual images are with respect to the optical system can greatly minimize if not annul such errors.

3.3 Eyepoint

In computer graphics, the center of perspective projection is often referred to as the *eyepoint* because it is either assumed to be the eye location of the observer or the location of video cameras used for the acquisition of stereo images (acting as two eyes separated by an interocular distance). Although the simplest optical system, known as the *pinhole camera,* reduces to a single point that serves as the center of projection or eyepoint, this is generally not the case for most optical systems (Barrett & Swindell, 1981). Depending on the required accuracy of virtual-object registration within the real world, the eyepoint for perspective projection used in the computational model must be defined in terms of a more or less precise location within the eye. This is especially the case because a standard eye is roughly 25 mm in diameter (Levine, 1985).

There is some controversy in the literature about which point within the eye must be chosen for the center of projection. If the meaning of this point is examined from a computer graphics point of view, the correct answer can be determined. The existence of a geometric center of projection within the eyes is related to the assumption that stimulating a point-like region on the retina produces a perception of light that is projected out

into the surrounding environment through "a center of projection" in the direction defined by the centroid of energy of the stimulated point on the retina and the center of projection. This means that any point at any depth always maps as a point on the retina. This is the case, however, only if either all the points considered in the surroundings lie in a plane that is conjugate to the retinal surface or those points lie in different planes but are being focused on the retina sequentially and the eyes have enough time to refocus. According to this purely geometric projection, where there is a point to point mapping, the center of projection for the object and image point can be taken as being the first and second nodal point of the eye, respectively (Robinett & Rolland, 1992; Deering, 1992).

Controversy arises from the fact that when the 3D space is imaged onto a plane or a surface such as the retina, not all the points in the 3D space are simultaneously focused on the retina, as has been pointed out by Ogle (1962). In this case, only points that lie in the plane or surface in space conjugate to the retinal surface are in focus and all other points are imaged as a blurry spot on the retina. The important point is that the center of the blurry spot is determined by the corresponding chief ray passing through the eye, not the nodal ray. Thus, the nodal point of an optical system corresponds to the optical center of projection only for precisely conjugate planes, and not for out of focus planes. For this reason, Ogle suggests choosing the entrance pupil of the eye as the center of projection in object space. It is, however, an experimental question whether the eyes operate on the basis of a single center of projection as assumed by current computational model for 3D graphics (Foley & Van Dam, 1982).

If we assume, however, a simple geometric model with the need of an eyepoint, we shall push the above argument further by stating that it holds even when those out of focus planes appear sharp to the eyes due to the eye's depth of focus. This is the case because the planes' sharpness does not come from the facts that the planes are conjugates of one another, rather it is a result of the finite resolution of the retinal mosaic. The nodal points of any optical system are a useful geometric construct for describing imaging properties of conjugate planes in an optical system, as described in Section 3.1 but, from an optical system point of view, it becomes meaningless for comparing visual directions of points lying at different planes in depth. Because the first nodal point of the human eye is only at about 3.5 mm behind the entrance pupil in a standard eye (Ogle, 1962), only small errors in objects' locations would be caused by choosing the nodal point rather than the entrance pupil of the eye as the eyepoint. Whether those errors are within tolerance depends solely on the application and the task of the HMD user.

If the entrance pupil of the eyes is the geometric eyepoint in the real 3D world, it should also be the eyepoint in the HMD because the virtual images presented to the eyes should give us the same monocular view of a 3D scene that one would see through a similarly shaped glass window. Thus, although the blur is not rendered on the virtual images, due to the auxiliary need to track the eyes and to increase the dynamic range of the graphics engine, each point on the virtual image does in fact represent a centroid of energy of a bundle of rays passing through the eyes. This directly results from the fact that the virtual images are imaged on the retina by the HMD optics. Therefore, although the imaging taking place in HMDs deals only with plane to plane imaging, and blurry spots are represented as sharp points on the virtual images in current HMDs, the entrance pupil of the eyes should be chosen as the center of projection for viewing the virtual images. Furthermore, the entrance pupil of the eyes should coincide with the computational center of projection. Also note that the precise location of the center of projection may have some impact on the value given to the FOV, as discussed below.

3.4 Field of View

An important component of HMDs calibration is the monocular FOVs of the optical viewer. Because most viewers are not corrected for residual optical distortion, the FOV cannot be measured experimentally. The best and safest approach in determining the FOV is to determine the imaging conditions of the system such as finite or infinite conjugates (either the object or the image is at infinity), the focal length, and the location of

the principal planes. The vertical and horizontal FOVs of an optical system working at finite conjugates are generally specified by (1) the vertical (V) and horizontal (H) dimensions of the display screens, respectively, (2) the focal length f of the magnifier, (3) the distance of the display screens x from the first principal plane of the optics, and (4) the location (L) of the entrance pupils of the eyes from the virtual images of the display screens formed by the optical system. One may combine x and f to calculate the position x' of the images of the displays formed through the optics with respect to the second principal plane, as well as the magnification m of the optical system for this set of conjugate planes as given by Eq. (1). The FOV is then most generally given by

$$\text{FOV} = 2\Theta = \tan^{-1}\frac{my}{L} \tag{2}$$

where y equal $V/2$ or $H/2$ for vertical and horizontal FOVs, respectively.

There are two common optical configurations that make exception to this rule: one where the FOV is independent of the eye location, and another where the FOV is independent of the display screens' location with respect to the optics, in other words, when the FOV is independent of the system optical magnification. The setup for the FOV for either of those configurations is described here.

The FOV subtended at the eyes is independent of the position of the eyes with respect to the optics as shown in Figure 2, when L becomes infinite, that is, when the virtual images are collimated. The magnification m becomes infinite as well, and one characterizes the size of an object in the FOV by its angular subtense at the eyepoint. In this particular case, the FOV depends solely on the size of the display and the focal length of the optical system. If we denote the focal length of the optics as f' (see Fig. 2), the FOV is now given by

$$\text{FOV} = 2\Theta = \tan^{-1}\frac{y}{f'} \tag{3}$$

For a given lens size and pupil diameter, only a limited range of eyepoint locations will produce an *unvignetted* FOV. Rays of light emitted by the display screens are said to be vignetted by the lens if they intercept the lens at a height larger than the outer diameter of the lens. As the eyes get further away from the lens, the effective rays from a point on the screens to the pupils of the eyes intercept the lens closer to its outer diameter and will eventually be vignetted before being completely occluded.

The other exception to the rule is the case where the FOV is independent of the position of the display screens with respect to the optics. This occurs when the eyepoints are located at the focal points of the lens as shown also in Figure 3. In this case, the FOV can also be calculated using Eq. (3).

Another important parameter of the system calibration that is related to correctly specifying the FOV is the *overscan* of the frame buffer. The computed images are rendered into a frame buffer before being scanned onto the display screens. The default assumption of the graphics library is that the frame buffer (512×640 pixels in our case) maps exactly to the vertical and horizontal dimensions of the screens. In practice, this is often not the case, and the exact overscan of the frame buffer must be measured. In this case, overscan refers to that region of the frame buffer that does not get drawn onto the display area. This overscan can be accounted for by defining a subviewport within the frame buffer that corresponds to the display capability of the screens. This subviewport must be centered on the display screens. The FOV can then be specified as the FOV corresponding to this subviewport, which indeed corresponds to the screens' size.

A last detail has to do with the implementation of specifying the FOV. A common approach is to specify the vertical FOV and a pixel aspect ratio instead of the horizontal FOV. The pixel aspect ratio can be calculated using the physical dimensions of the screens and the number of vertical and horizontal scan lines of the frame buffer that are actually visible on the screens.

3.5 Interpupillary Distance and Eyes Position

Three types of distances should be distinguished: (1) the human subject interpupillary distance (IPD), (2) the lateral separation of the optical axes of the monocular optical systems, referred to as the optics baseline, and

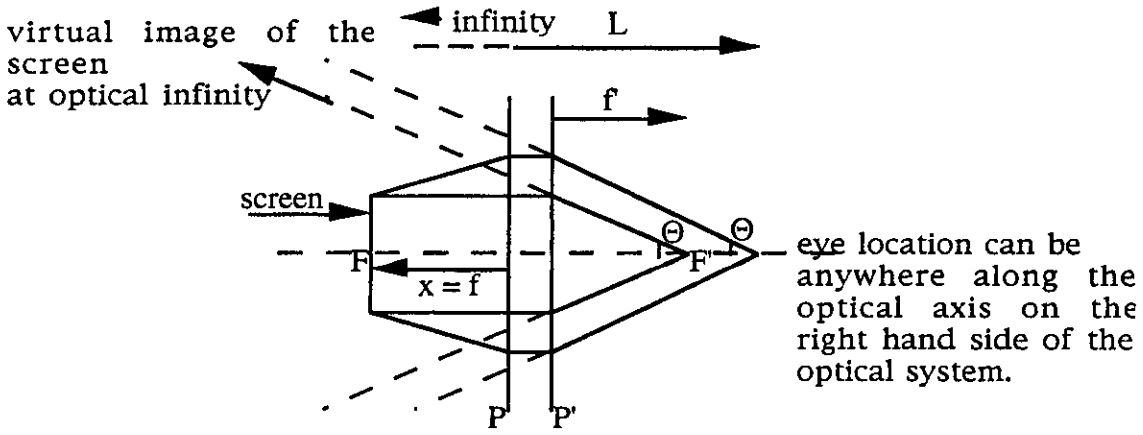

Figure 2. *Model of a lens by its principal planes P and P' and its focal points F and F'. For the display screen at F, the virtual image of the display is at infinity and is subtending a visual angle of 2Θ, independently of the position of the eye.*

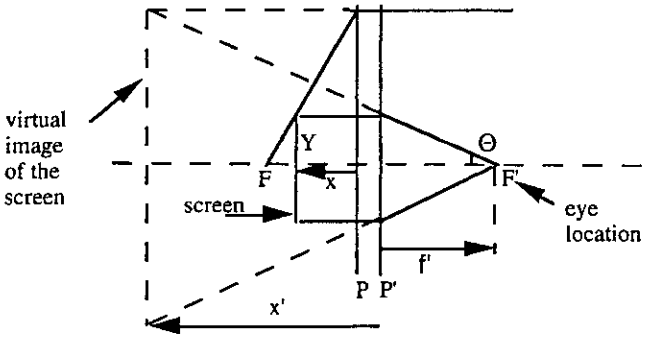

Figure 3. *Model of a lens by its principal planes P and P' and its focal points F and F'. For the eye position at F', the FOV is given by 2Θ, independently of the position of the display screen with respect to the lens.*

(3) the lateral separation of the computational eyepoints, referred to as the computational baseline because it is used to compute objects positions and shapes on the screens.

In most systems, the optics baseline is fixed, the variation in subject's IPD is not accounted for, and the computational baseline is set to a value of about 62 mm to allow "most users" to fuse the images. The difficulty in fusing images remains the problem of narrow IPD observers, however, because 62 mm is still above both the 5th percentile of male and female IPDs, which are 57.7 and 35.8 mm, respectively (Woodson, 1992). Thus, a fixed value for the computational baseline makes it difficult for a significant part of the population to fuse the images. Moreover, it creates shifts in depth perception to almost everyone because none of the parameters— IPD, optical baseline, and computational baseline—actually match. In a calibrated setup, the optics baseline and the computational baseline should be set to the IPD of the subject.

Section 3.4 showed that the exact position of the eyes behind the optics need not necessarily be known to determine the FOV of the system. For example, in the case of collimated images, the FOV subtended at the eyes is independent of the location of the eyes. The absolute position of the user's eyes, when considered as a rigid pair behind the viewer, however, is always important to the perceived location of virtual objects relative to real objects in combined real and virtual environments. We could as well speak of head position, here, because "eyes as a rigid pair" refers to the position of the head behind the see-through combiners. When the eye-pair moves laterally (perpendicularly to the optical axes of the viewer), or longitudinally (along the optical axes of the viewer) the 3D virtual objects are not necessarily perceived to move by the same amount as would be preferred. Virtual objects can be shown to move by the same amount as the eye only for the case of collimated images. Thus, although the eyes' position need not necessarily be known for the calculation of the FOV, the position of the eyes needs to be known to measure the

expected depth location of virtual objects relative to real objects.

An expression has been derived for the change in 3D location of virtual objects relative to real objects due to a lateral or longitudinal displacement of the eyes in the general case of noncollimated images. The case of collimated images will be treated as a particular case. A mismatch in IPD is not included in this expression, the eyes being considered as a rigid pair. Errors in eye location can be expressed as lateral and longitudinal components. The lateral component corresponds to the case where both eyes are to the left or to the right of the nominal position. The longitudinal component corresponds to the eyes being moved forward or backward from the nominal position. The change in 3D location of a virtual object due to a lateral displacement of the eyes is given by

$$x_{p'} = x_p + \Delta x - \Delta x \frac{D}{L} \tag{4}$$

where x_p and $x_{p'}$ are the x-locations of point P for the nominal and perturbed locations, respectively, Δx is the signed lateral displacement of the pair of eyes, D is the distance of P along the z-axis, and L is the distance of the monocular virtual images from the nominal eye position (Fig. 4). If E_L and E_R are the location of the left and right eyepoints, respectively, the derivation of Eq. (4) assumes E_L as the origin, i.e., E_L (0, 0, 0), the x-axis going from left to right, and the z-axis going from the eyepoint toward the virtual images. In this case, E_R has for coordinates E_R (IPD, 0, 0), and P and P', $P(x, 0, D)$ and $P'(x', 0, D')$, respectively. It can be shown that D equals D' for lateral movements. Moreover, if L equals D, that is the point P is located in the plane of the monocular virtual images, the position of P according to the computational model is unchanged from nominal value because the second and third terms of Eq. (4) cancel out ($x_{p'} = x_p$). If L becomes infinite, that is the case of collimated images, the point P moves exactly by the same amount as the eyes ($x_{p'} = x_p + \Delta x$). The virtual point seems to follow the eyes. In the last case where L is neither equal to D nor infinite, however, the location of P is not only a function of Δx, but also of D and L. If D is smaller than L, the point P moves in the same direction as the eyes but by a lesser amount. If D is greater than L, the point P moves in the opposite direction. A virtual object or space that is made of an ensemble of points P is thus distorted.

A similar equation can be derived for the longitudinal case. The change in depth location due to a longitudinal displacement of the eyes is given by

$$z_{p'} = z_p + \Delta z - \Delta z \frac{D}{L} \tag{5}$$

where z_p and $z_{p'}$ are the z-locations of point P for the nominal and perturbed locations, respectively, and Δz is the signed longitudinal displacement of the pair of eyes. In this case, a displayed object is being compressed for D smaller than L and stretched for D greater than L. A virtual object or space made of an assemble of points P is thus distorted in all directions.

This simple calculation brings insight into the tradeoffs between different HMD setups for any specific task. If the task is to judge the depth of virtual and real objects located within a few meters from the observer in an optical-see-through environment, it would be better for the user to have the monocular images positioned a few meters away as well so that accommodation and convergence are more naturally coupled. The above calculation tells us that, in this case, there will be only a finite volume where the virtual space will not be significantly distorted by potential small eye-pair lateral or longitudinal movements. This volume can be determined according to the required accuracy of the task at hand. For virtual objects positioned at larger distances, however, collimation may be a better choice.

4 Apparatus Setup and Calibration

4.1 Apparatus Setup

All studies described here were carried out on an optical bench prototype HMD designed and built using off-the-shelf optical components. The layout of the optical setup for one eye is given in Figure 5, while the setup is shown in Figure 6. Building our own setup was necessary to have complete access to and control over all system parameters. The focal length of the magnifier was

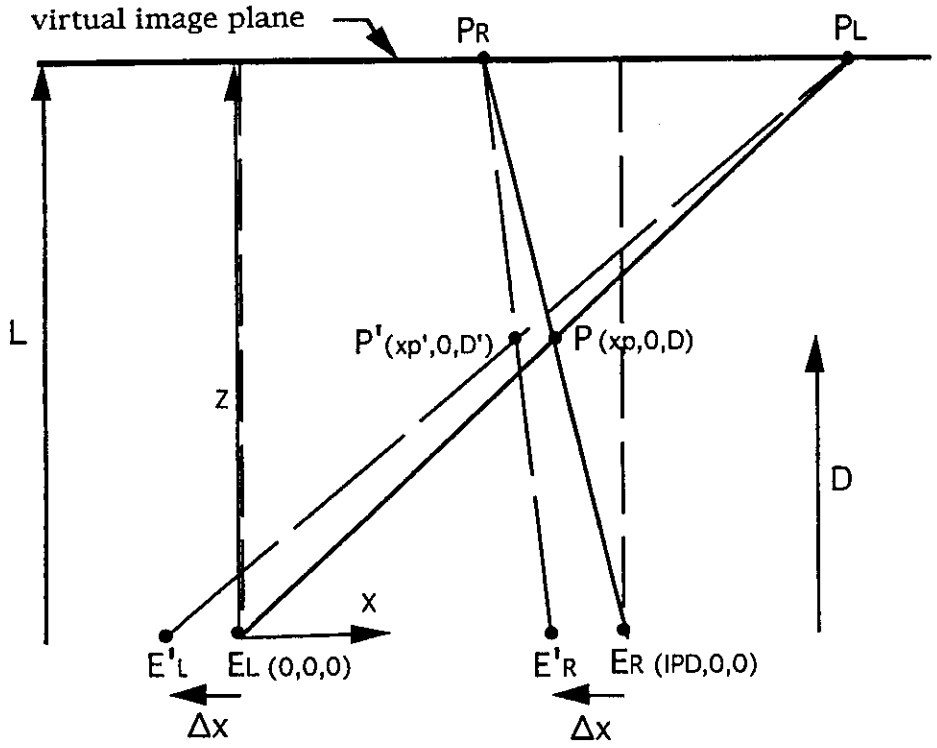

Figure 4. *This figure illustrates the effect of lateral eye movements on the perceived location of a point P in 3D space. If the eyes move as a rigid pair by an amount Δx, the point P is now perceived as P' and $x_{p'}$ is related to x_p by Eq. (4).*

86.24 mm and in all the experiments described here, the images were collimated. The computer graphics were generated by Pixel Plane 5, a massively parallel graphics engine developed at UNC-CH under the direction of Henry Fuchs and John Poulton (Fuchs et al., 1989).

The displays used for the study were passive matrix Memorex LCD color displays with vertical and horizontal dimensions of 40 by 52.65 mm, respectively. The number of vertical and horizontal addressable scan lines was measured to be 442 × 593, respectively. The vertical FOV is 26.11° as calculated using Eq. (3). A pixel ratio of 1.02 can be calculated from the physical dimensions of the display area and the number of vertical and horizontal addressable lines. Neglecting the overscan of the frame buffer would result in an error of about 5° in the specified FOV. This would create objects that look larger and closer than theoretically intended. Such an error could be detected by the calibration procedure described below.

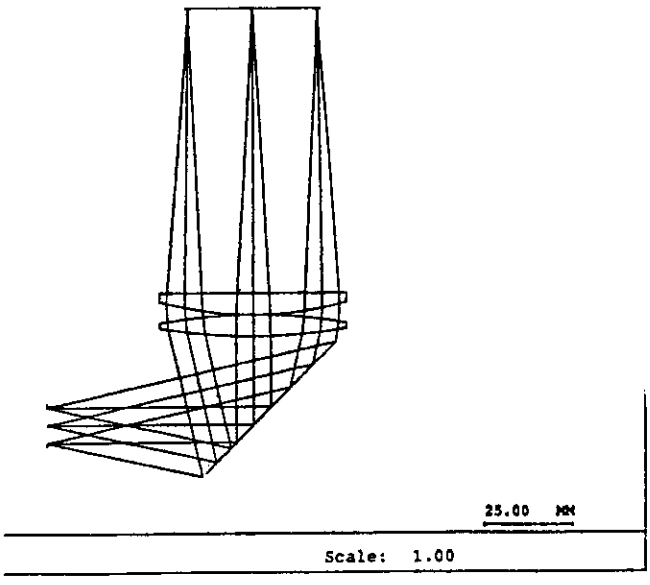

Figure 5. *Optical layout of the magnifier for one eye.*

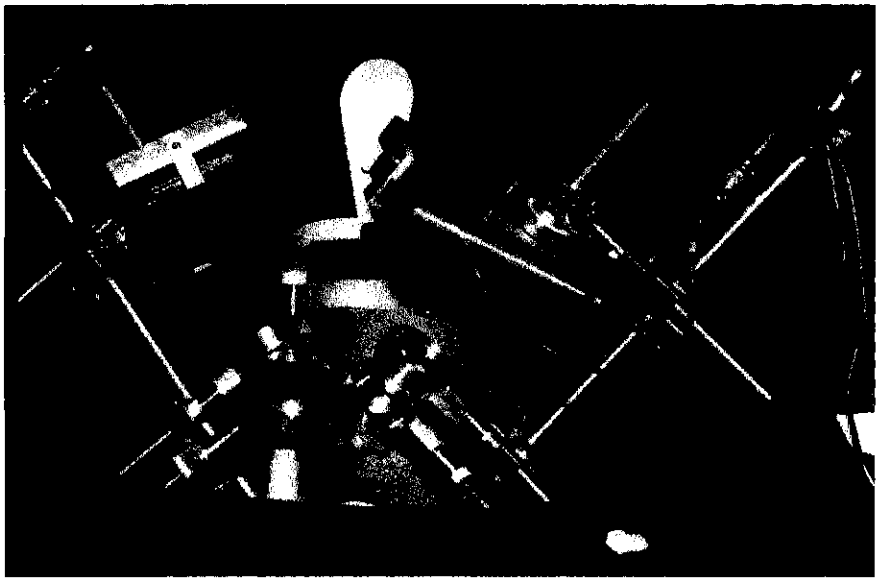

Figure 6. *Bench prototype optical setup designed and built by J. P. Rolland.*

The geometry adopted for the layout of the optical system eliminated gross lateral misalignments of the display screens with respect to the optics that are often found in commercially available systems. As for any engineering system, there are inevitable residual misalignments due to mechanical assembly. The exact amount of residual lateral misalignment cannot be easily measured, however, without introducing supplementary measuring devices. A quick and reliable calibration technique based on the computational model was derived and described below.

A useful feature of the optical setup was the ability it gave us to adjust the optics and computational baselines to match the subjects IPD down to 60 mm. All test subjects had IPDs between 61 mm and 68 mm measured with an Essilor PRC pupillometer set for infinity viewing.

4.2 Calibration Technique

Once the FOV has been set, and the optics and computational baselines have been matched to the IPD of the subject, a quick calibration of the system was performed to determine if computational offsets of the displayed images were necessary. Residual lateral displacements of the displays due to mechanical limitations and residual alignment errors were adjusted with respect to the optical axis. For a carefully aligned system, the magnitude of the offsets should be very small, on the order of a few pixels, assuming the parameters within the computational model were set correctly.

Calibrating the system consisted of looking monocularly at a symmetrical pattern drawn on the displays in such a way that, if seen binocularly, its projection was at a very large distance. For this purpose, a distance greater or equal to 200 m is recommended. Although 6 m is often referred to as optical infinity, this latter distance seemed insufficient based on geometric calculations using our model. A symmetrical pattern consisting of a white circle at the center and alternating black and white circles at the periphery was used because of the importance of symmetry in the alignment process. Subjects checked for alignment of each eye with the optical axis of the magnifier by closing one eye at a time and checking that the frame of the magnifier was centered in their FOV. The experimenter draws on a sheet of paper two crosses separated by the IPD of the observer, as well as the mid-point between the two crosses. As the observer was looking through the displays monocularly, the experimenter positioned the two crosses at any depth in

space such that, while looking with the right eye, let's say, the right cross fell at the center of the virtual pattern displayed in the right eye. As the observer switched eyes, the center of the left virtual pattern and the left cross should exactly superimpose. If the superimposition for each eye does not hold, offsets for the images being displayed must be set in the software until a perfect match occurs. This is best accomplished using some systematic procedure to offset the images.

Once the calibration for infinity is completed, the experimenter can perform a consistency check by looking at the virtual pattern that is now projected at a nearby distance. This is useful because, if done correctly, the calibration should hold at any depth. The two physical crosses are now positioned at the same depth from the eyepoints than the virtual pattern. While closing alternatively one eye, the observer should perceive the center of the virtual pattern to fall on the mid point between the two crosses. This calibration technique was derived directly from the basic concept of the computational model of stereo images illustrated in Figure 7. Given a cross at 2 m, the figure shows where the intersection of the projection of the cross with the virtual image is for both the right and left eyes. This location on the virtual image corresponds to a location on the display screens according to the optical imaging equation given by Eq. (1). Therefore, if the optical path starting at the display screens is reversed and if the correct point on the display screens is marked, its virtual image will be correctly marked as well. Thus, the impact of the light with a screen placed at 2 m will coincide with the physical cross at 2 m. Note that the calibration procedure never used both eyes simultaneously to avoid confusing calibration issues with perception issues. The two eyes are only used sequentially by the subject.

This consistency check is valid, however, only if the optical distortion of the system has been compensated for (Robinett & Rolland, 1992; Rolland & Hopkins, 1993) and if the change in IPD of the observer when focusing at a nearby distance is negligible. The calibration using the optical infinity condition, on the other hand, is always valid, regardless of optical distortion as long as the center of the symmetrical pattern coincides with the center of optical distortion within a few pixels, which is the case in this experiment.

More precisely, the optical and computational baselines should in principle be set to the new IPD of the observer for the distance where the depth judgment is made. In our experiment, an estimation of the maximum change in IPD between its nominal value at infinity viewing and its value at the nearest distance, 0.8 m in our case, was 2.5 mm as measured using the first corneal reflection. We estimate that IPD changes due to convergence are more in the order of 1.5 mm when measured at the entrance pupil of the eyes, which is roughly 3 mm behind the cornea. In our setup, the IPD of the observer was taken to be its value for infinity viewing in the experiments reported here. We shall discuss this point in more detail in the discussion section.

A variant of the above technique could also be used to verify the FOV of the system (Ellis, 1993). It consists of comparing the size of a virtual object with that of a real identical object, again under monocular viewing. Caution must be exercised, however, if optical distortion is present in the display because monocular images will appear larger in the case of pincushion distortion, which often limits HMD systems. This method, however, can be used for distortion-free systems.

5 Experimental Studies

5.1 Rationale

Real world experience suggests that looking at nearby objects triggers two processes acting simultaneously: the eyes converge by rotating inward and the crystalline lenses accommodate to the location of the object. Accommodation is needed to bring into focus details of the objects. It may seem natural that the actions of accommodation and convergence are definite, and that for any specific accommodation, there must be an exactly corresponding convergence.

This is not, however, how vision works in most HMDs. In HMDs, one needs to distinguish between monocular and stereo aspects of vision. With respect to each eye separately, the virtual images of the two LCD

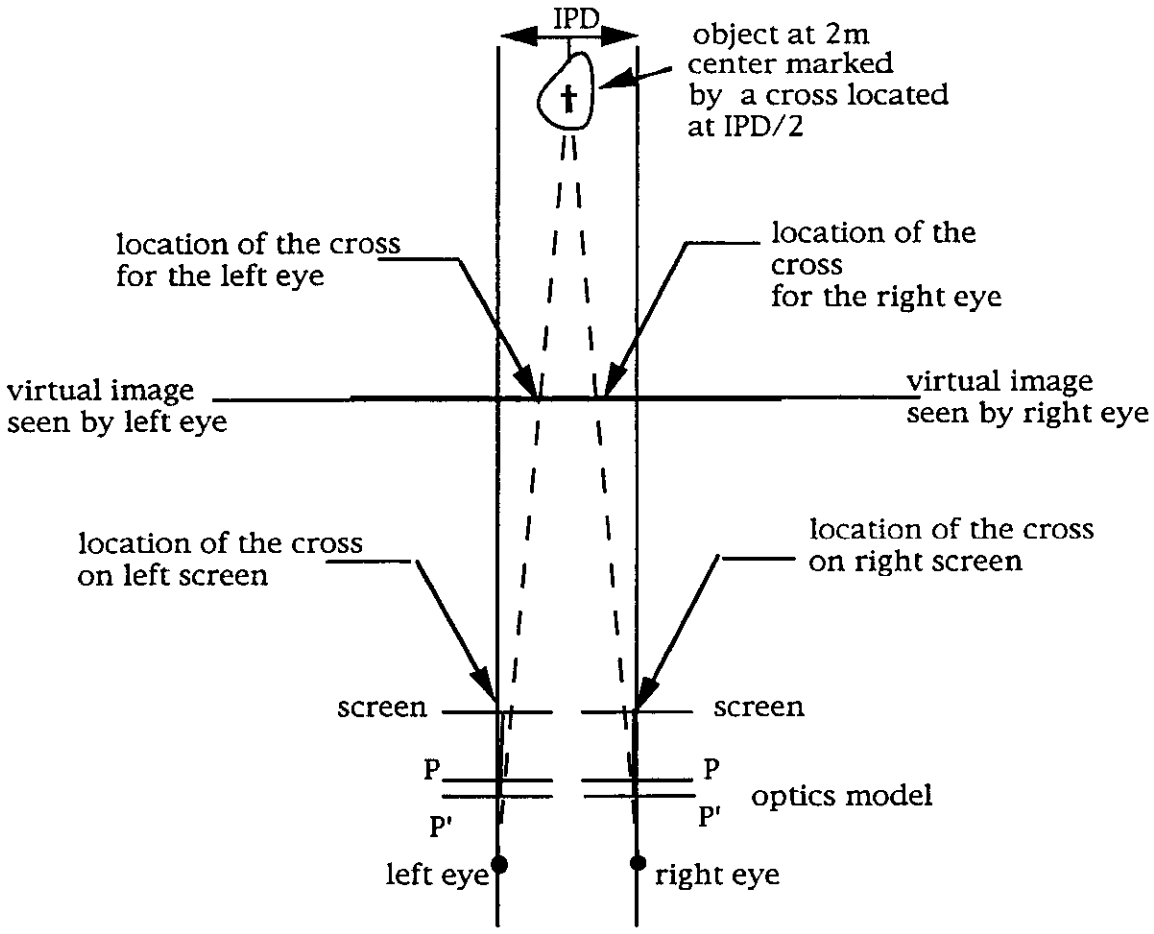

Figure 7. *Computational model of stereo pairs and its direct application to calibration.*

displays mounted on the head are formed by the optics at a definite plane in space. The position of that plane depends on the distance of the LCD displays to the optics. That distance usually remains fixed once it has been set to a desired value.

A particular setting is the one of infinity viewing (collimated images) or zero accommodation where the displays are positioned at the focal plane of the lenses. Such a setting is often preferred because it sets the eyes to a relaxed state of accommodation. It can also be shown that what is perceived in the HMD is less dependent on the exact position of the eyes of the observer behind the optics as shown in Section 3.5. Indeed, most stereoscopes are so adjusted that the viewing lenses act as collimators (McKay, 1948).

With respect to the two eyes combined, vision in HMDs is similar to vision in the real world in that convergence is driven by the location of the 3D objects perceived from the disparate monocular images. For both real and virtual images, the two eyes can fuse the stereoscopic images of a 3D object only if they converge on the 3D object itself within a tolerance area known as the Panum area (Gulick & Lawson, 1976). For convergence of the eyes outside of Panum's area, the observer will perceive *diplopic* images, that is double images.

The fact that accommodation and convergence are divorced from one another in most HMD settings brings up the question of whether the distance at which the monocular images are optically formed has any impact on the perceived depths and sizes of virtual objects

embedded in the physical environment and/or creates distortions in depth and sizes of virtual spaces. This is especially important in the case of see-through HMDs, where the observer has access to both the simulated and the physical world simultaneously.

This issue is investigated here by studying the case of generically shaped, nonoverlapping, static objects positioned within 1.5 m of the observer. The impact of the uncoupling between convergence and accommodation is examined by looking at the most extreme case of uncoupling, that is when 3D real and virtual objects are placed at nearby distances while the 2D virtual monocular images are optically collimated.

5.2 Subjects

Three subjects participated in the experiments; two of them are authors of this paper. Two of the subjects were males in their late 20s, and one was female in her early 30s. Two of the subjects had large exposure and experience with HMDs, while the third subject was a novice. Subjects IPDs were 61, 64, and 65 mm. All observers were tested for stereoscopic vision using random dot stereograms.

5.3 Stimuli and Experimental Design

Each presentation was composed of a pair of objects presented simultaneously. One of the objects was a cube, the other was a cylinder, and the two objects were presented with a lateral separation of 110 mm. Lateral separation was such that, given the subject viewpoint and the range of distances, no overlap of the two objects ever occurred during the experiment. Moreover, subjects reported that such a lateral separation made it easier for them to use the center of gravity of the objects as the basis of comparison and prevented them from basing their judgments solely upon the objects' inner edges. Those edges were ill-defined for the cylinder while the inner edge was at a different depth than the center of gravity for the cube.

Generic objects of different shapes were chosen for comparison, in this case a cube and a cylinder. This was done to prevent subjects from assessing depth perception solely on the basis of the perceived size of one object relative to the other. For example, if the two objects were of similar shapes but of different sizes, one may be biased to always see the smaller cube farther away than the larger one. On the other hand, if two objects of identical size were chosen, then depth assessment could be done solely upon the assessment of their respective sizes.

The real cube was 38 cubic mm in size and the cylinder was 13 mm wide by 235 mm long. In all presentations, the cylinder was seen with its longer axis vertical while the cube was tilted forward and slightly rotated to the left so that three faces were visible. Both virtual and real objects were uniformly white, with the virtual objects modeled after the real objects. The real objects were supported by black poles that were not visible to the observer. Lighting source for the real objects were connected to the supporting poles so that movement of the real objects kept the relationship between the light source and the real object constant. The light source itself was outside of the viewer's FOV. Shading in the virtual environment was set to visually match the shading of the physical objects as close as possible. Lighting in the virtual environment was modeled using both ambient light and directional light. The luminances of the displays were about 140 and 100 cd/m^2 for the objects and background, respectively, yielding objects' contrast of about 40%. The table supporting the objects was covered by a black cloth so no background was present.

Two viewing distances were used, 0.8 and 1.2 m. For each of those two distances, three different setups were employed, making it a 3 × 2 design. The three setup conditions describe the different relationships between the pair of objects in a single presentation. In one of the conditions both objects were physically presented to the observer; this condition was called real/real. In the case where both objects were presented graphically through the HMD the condition was called virtual/virtual; finally the mixed condition in which one object was presented graphically and the other was presented in the physical environment was called real/virtual. In all cases, viewing conditions were equated and the see-through display was used for viewing the scene whether or not the objects were virtual. The object presented on the left was

always fixed in location, while the one on the right was moved in depth between trials. Motion in depth between trials was invisible to the observer, in the case of the virtual images the screens were blanked for the intertrial interval. For the real objects a blacking cloth was used in front of the subject's eyes until the object reached its new location and only then was removed.

The virtual/virtual, real/virtual, and real/real experiments were conducted in that order, while the depth presentations were counterbalanced within each of those three conditions. An improvement to the experimental design would be to counterbalance the order of the three conditions as well. Viewing duration was unlimited and terminated on a decision key press, which also initiated the next trial. Subjects used a chin rest to maintain a fixed viewing distance.

5.4 Task

Subjects were asked to judge the relative proximity in depth of two objects presented simultaneously. Subjects performed a 2-alternative force choice task in which they were asked to answer if the object on the right was closer or farther from them relative to the object on the left. The observer used a two button device for their responses and the object on the right was the one translated in depth between trials.

5.5 Psychophysical Method

A method of constant stimuli (MOC) was used to present the stimuli. The depth values presented were chosen with a staircase procedure, before each run, to ensure measurements on both the slope and the shoulders of the psychometric curve for each observer in each of the experimental conditions. Our standard was such that it ensured that at least four of the depth values chosen fell along the steepest part of the psychometric curve. Each of the three experiments consisted of 31 blocks with 10 trials in each block. Presentation of the different depths within each block was done in a random order. Data were analyzed using probit analysis on the last 30 blocks; the first block was always discarded.

The resulting psychometric function was plotted for each subject and after each experiment to ensure that at least four points lay on the linear part of the psychometric function. If less than four points lay on the slope of the curve, results were discarded and the experiment was repeated with a more suitable range. However, this never happened past the pilots experiments. Points on the shoulders of the psychometric function were used as anchor points for the subject. Anchor points are defined as depth values for the comparison stimuli that were easily discriminable from the reference. The use of such easily discriminable values was used after noticing that subjects tended to be inconsistent in their depth judgments and reported to loose depth information when presented with depth values that were only around threshold.

Veridicality of perceived depth is defined as the difference between the nominal depth as specified in the computational model and the measured value (nominal − measured). We shall denote it as ΔPSE because it represents the departure from the expected point of subjective equality. Precision of perceived depth is defined as the smallest difference in perceived depth that can be detected reliably at the 84% level.

5.6 Results

Results are shown in Figures 8 and 9 for veridicality and precision of perceived depth, respectively. Figure 8a, b, and c shows the veridicality of perceived depths for the three basic conditions, real/real, virtual/virtual, and real/virtual, respectively. Similarly, Figure 9a, b, and c shows the precision of perceived depth plotted this time on a log scale, for the three basic conditions. Each of those three graphs shows the performance of each of the three subjects separately.

A summary plot is shown in Figures 8d and 9d, for veridicality and precision of perceived depth, respectively, where the performances per condition were averaged over the three subjects.

Figure 8a, b, and c shows that there is, on average, no mean error in perceived depth if both objects are real or virtual, even though the variance associated with the mean values in the virtual/virtual condition is significantly higher. The upward shift of the data points in Fig-

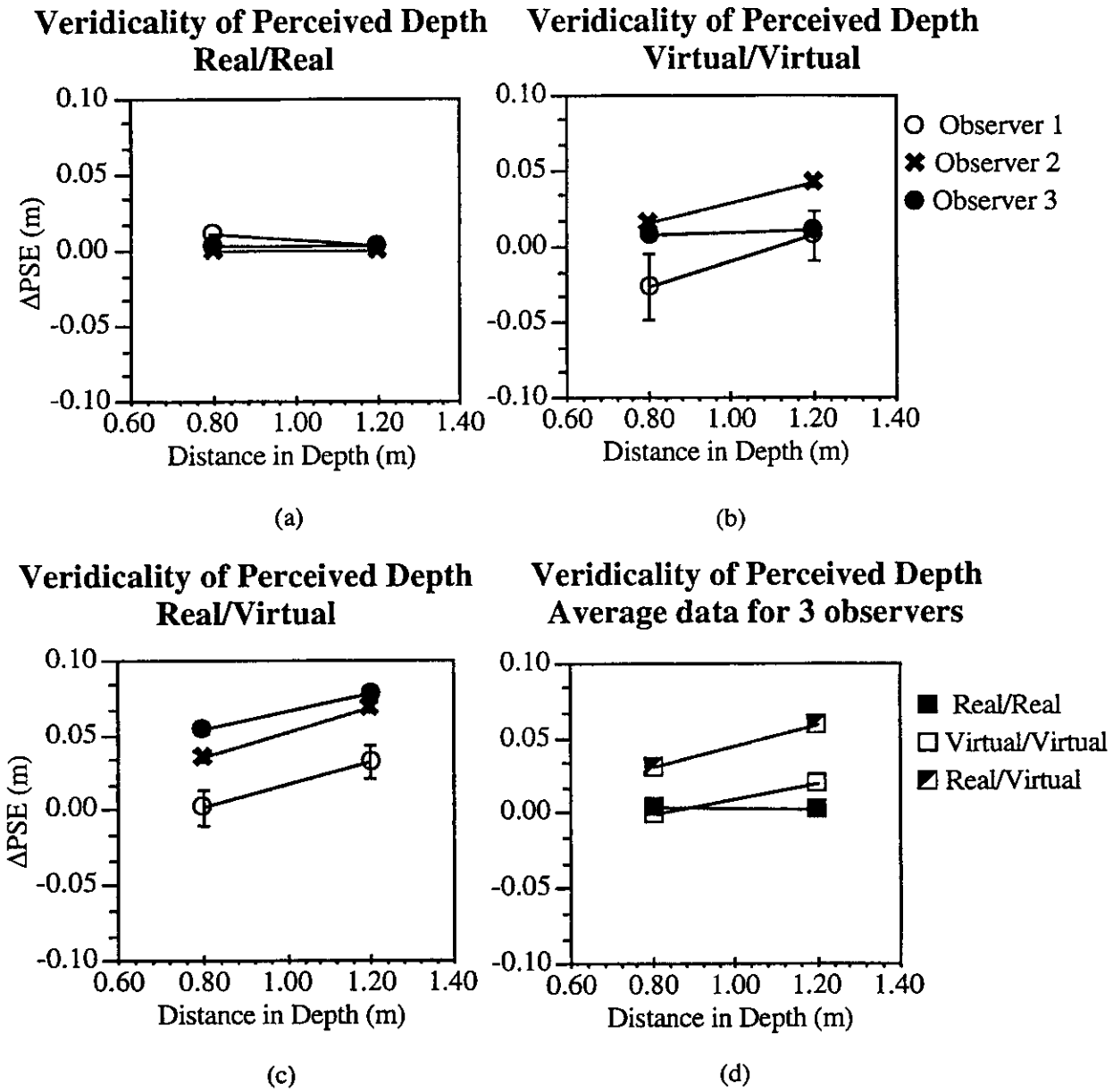

Figure 8. *The veridicality of perceived depth, graphed for the three conditions (a) real/real, (b) virtual/virtual, and (c) real/virtual, are plotted as a function of two depths 0.8 and 1.2 m for three subjects. Part (d) summarizes the results for the three conditions after averaging the data over the three subjects. The upward shifts of the data in (c) and (d) mean that the virtual objects are seen further away than the real objects.*

ure 8c and d means that there is a shift in perceived depth for virtual objects as compared to real objects. A positive ΔPSE, in this case, means that the virtual objects are perceived farther away than real objects.

Figure 9d, which summarizes Figure 9a, b, and c, shows a significant increase in discrimination thresholds for the two conditions virtual/virtual and virtual/real as compared to the real/real condition.

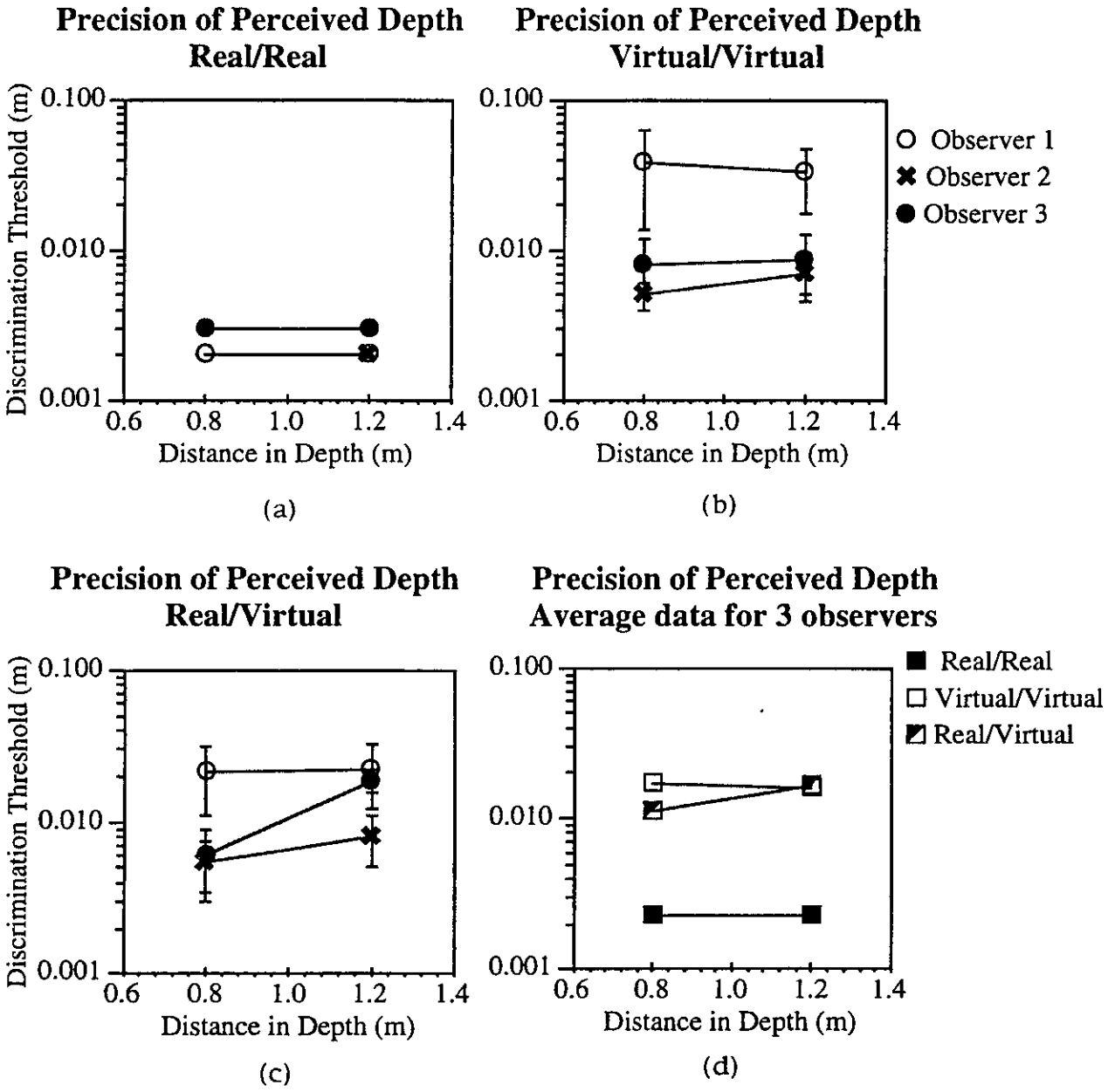

Figure 9. *The precision of perceived depth, graphed for the three conditions (a) real/real, (b) virtual/virtual, and (c) real/virtual, are plotted as a function of the two depths 0.8 and 1.2 m for three subjects. Part (d) summarizes the results for the three conditions after averaging the data over the three subjects. Those figures show that depth discrimination thresholds are being elevated from 2 mm in condition (a) to roughly 15 mm in conditions (b) and (c). Moreover, no systematic elevation of thresholds with depth was observed.*

6 Discussion

Some of the problems of studying perception in virtual environments were avoided in this study by using a static bench prototype system. Those include the potential slippage of the HMD as the observers move their heads around, the weight of the HMD (Lackner & Dizio, 1989), and the time lag with which the HMD tracks head movements.

The averaged measured shifts in perceived depth of

virtual objects with respect to real objects were in the order of 50 mm for our system. It is also the case that virtual objects were seen farther away than real objects when both objects were presented at the same depth in visual space. Similar findings were reported by Roscoe and others for familiar objects such as airport runways. Pilots are thought to overshoot the runway because the decrease in apparent size makes them think they are farther from touchdown than they in fact are. This is based on familiar size cues which they learn when learning to fly. They, in essence, have learned how big the runway is, what its angular size is, and how it changes during a correct approach. One fundamental difference between this study and others is that two generically shaped objects, here a cube and a cylinder, were chosen as stimuli, rather than familiar objects. Another important difference is that there was no object motion, neither egomotion or induced egomotion of the observer. Moreover, most of Roscoe's studies were based on size estimation rather than depth estimation. Because some efforts were taken in our experimental studies to remove familiar size cue, precisely the cue that may have caused Roscoe's depth error, the greater apparent distance to the virtual objects in this paper and in Roscoe's observations of hard landings due to distance errors, may have different causes. In our experiment, no feedback was available to the subject.

Before postulating on the causes of possible artifacts, one more optical parameter of our system not previously mentioned, optical distortion, needs to be addressed. At the time of doing this first set of studies, optical predistortion of the display was not yet implemented. Although the optical distortion should not have affected the calibration of the system as explained earlier, it did affect the placement in depth of the virtual objects because the objects were laterally separated by 110 mm and were displayed away from the center of distortion in at least one of the two displays. An illustration of the amount of distortion caused by the optics is shown Figure 10. Because the exact layout and parameters of the optical system are known, the effect of residual distortion on assigned depth to objects can be estimated. An exact calculation of the impact of optical distortion on the results reported in Figures 8 and 9 was carried out

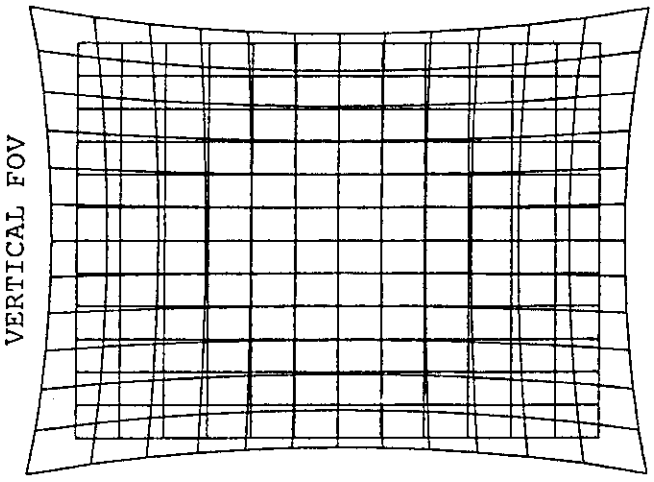

Figure 10. *Illustration of the pincushion optical distortion of the optical system.*

using an optical ray-tracing program. Calculations show that pincushion distortion for the individual monocular images has the effect of bringing the 3D virtual objects created from the two disparate virtual monocular images closer to the eyes. The calculated values were about 11 and 7 mm for the three subjects at the viewing distances of 0.8 and 1.2 m, respectively. This means that our measurements of shifts in depth of virtual objects with respect to real objects may be too conservative. New data are now being acquired using predistorded images (Robinett & Rolland, 1992; Rolland & Hopkins, 1993) to confirm that the perceived shifts are indeed larger than the ones reported here.

Another source of depth judgment error is due to the change in IPD of the observers as they converge their eyes on nearby objects as mentioned in Section 4.2. We estimate the error in depth based on our computational model to be in the order of 19 mm for a change in IPD of 1.5 mm. Again our experimental results on veridicality of perceived depth may be conservative because the psychophysically measured depth errors could have been in fact larger if IPD change had been accounted for. A precise value for IPD change, however, requires a model of eye movements with eye rotations during convergence. We are currently investigating such models to be included in future research.

Another more subtle, yet relevant parameter that could have biased our results is the one of illumination. Illumination in real environments is often punctual, i.e., it comes from a point source of light. On the other hand, illumination in virtual environments is more likely to be directional, i.e., it is not spatially localized. Moreover, the spectral range of illumination can be different in subtle ways. For example, the illumination in the virtual environment was slightly bluish, while illumination of real objects was slightly more reddish. This is of importance because it has been shown that the binocular observation of a red light and a blue light, located at the same physical depth, often will cause the red one to appear closer than the blue one, a phenomenon known as chromostereopsis (Vos, 1960). Experimental data reported by Vos show, however, that the maximum amplitude of the effect over four observers was only 4 mm at 0.8 m viewing distance. Moreover, half of the observers reported that the red object (a slit in Vos's case) was perceived behind the blue slit. Due to the very slight difference in the spectral range of illumination sources in our real and virtual environments, and considering the measurements reported by Vos in saturated environments, it would appear that chromostereopsis has a negligible role in our findings.

An elevation in variance in the measure of veridicality of perceived depth of virtual objects was recorded. It can be interpreted as a decrease in strength of the percept of depth for virtual objects in the context of that specific experiment. A variant to this effect is the need for anchor points on the psychometric functions. The use of anchor points was mentioned in Section 5. In their absence, subjects had a tendency to drift toward larger differences between perceived and physical depth, and to be more inconsistent across repeated measures. This is indeed a point of interest because it suggests that the percept of depth of the virtual objects was somewhat unstable (Foley, 1991; Ellis, 1994). That is, it lacked strength in the sense that the objects did not seem anchored to one depth but rather to be susceptible to drift more easily in space. Further perceptual investigations of this finding are necessary.

Depth discrimination thresholds were found to be higher in virtual than in real environments. The elevation in thresholds, however, is less than the elevation that one would predict based on the addressability of the display. This can be explained by the fact that antialiasing is performed on the images. The lower bound for depth displacement that can be computationally generated is given by one over the number of addressable lines per millimeter divided by the number of available gray levels. However, this assumes both an ideal observer as well as an interpretation of a change in luminance as a change in location by the human visual system. A more veridical prediction based on the human visual system must involve a visual task such as a vernier acuity task. In such a task, changes in location would be manipulated by changing luminance of the top bar and asking the observer to judge if the latter is to the right or to the left of the bottom bar. Such a measure would give us the perceived resolution of the system that includes the observer.

Size and depth perception are intimately related. As a virtual or real object is moved closer it tends to look bigger, or similarly, as it becomes bigger it tends to look closer. This is generally true for familiar objects of known size but less true for generically shaped objects (e.g., a cube or a cylinder) or for unfamiliar objects. Although the subject could not compare the relative sizes of the two objects to assess depth, because that possibility was eliminated by choosing two objects of different shapes, the subjects may still have used the size change in trying to assess depth.

To determine if size change was an important factor in making depth judgments, pilot studies were run under two conditions: (1) the angular subtense of the virtual object that is being translated in depth from trial to trial was kept constant, and (2) the size of the moving object was randomly selected from trial to trial within a 20% range of the mean value.

In the first case, the displayed object was of constant size when using one eye. When using stereo vision, on the other hand, the object appeared to actually shrink in size as it moved forward. This did create a countercue to depth because real objects seemed bigger when looked at closer. Under this condition, one of the subjects noticed the underlying structure of the experimental setup and the results were unchanged. The second subject was at

first fooled and reported the objects as always farther away than the real object. In an attempt to record the new range of perceived depth of the virtual object using a staircase method, the subject was able to recalibrate himself and final results remained unchanged as compared to the normal viewing condition.

In the second case, no significant changes in the results were observed for the few studies ran. More data need to be collected to fully capture what is the impact of size perception on the perceived depth of generically shaped objects.

Because we suspect that the collimation of the images plays a significant role in the measured errors to depth veridicality, the collimation remains to be experimentally manipulated. The crucial test of whether this or some other, perhaps artifactual, cause is responsible is to compare human performance of collimated images with images that are not collimated but rather are placed at the same optically specified distance for accommodation and stereopsis.

Finally, a natural extension of the study of spatial relationships in HMD is to investigate the importance of the amount of spatial overlap between the two types of information (real and virtual) being presented (Weisheimer, 1986; Patterson & Martin, 1992). To interpret experimental results with spatially overlapping objects, however, a solid understanding of the percept of depth and size of spatially nonoverlapping objects in combined real and virtual environments is first required. Work is currently underway to quantify and understand how spatial overlap affects veridicality of perceived depth and sizes.

7 Conclusion

This paper discusses some of the important current problems with see-through HMDs, the problems of information integration in combined real and virtual environments. The measurement of perceived depths and sizes of virtual objects relative to real objects in a combined real and virtual environment is proposed as a starting point to studying these problems. A large number of parameters had to be addressed, carefully considered, evaluated, and set before any attempt to quantify depth and size perception in HMDs could be performed. This part is described in detail in Sections 3 and 4. A set of preliminary experiments in depth perception for three conditions real/real, virtual/virtual, and real/virtual is reported. These experiments were conducted on an optical bench prototype see-through HMD. Data are presented for two viewing distances, 0.8 and 1.2 m, and for three observers. Results seems to indicate that even for generically shaped objects, virtual objects are perceived systematically farther away than real objects given our current computational model for depth perception and our experimental setup. Two factors, the change in subjects' interpupillary distance with eyes convergence and the optical distortion of the optics, have not been accounted for in the model at the time of those experiments. Theoretical predictions for those two factors indicate that the shift in perceived depth of virtual objects with respect to real objects would have been larger than reported. Further studies are currently under way to confirm those findings. The measurements reported here were made for collimated images. The discrepancy between accommodation and convergence, which may have caused part of the perceptual bias, is currently being investigated. If this discrepancy proves to be important to depth and size judgments in combined real and virtual environments, results from these experiments (1) call for a modified computational model that would yield the correct depth percept of virtual objects, and (2) may drive the technology of HMDs to new directions such as building in an autofocus capability so that accommodation and convergence operate in a more natural way.

Acknowledgments

I thank the HMD and Pixel-Plane teams at UNC for their support with the various components, hardware and software, necessary to carry out the studies. Especially, I would like to thank Vern Chi and John Thomas for their significant help with the experimental setup. I thank Anselmo Lastra for his stimulating discussions about Pixel Plane 5 and the graphics library and Erik Erickson and Mark Mine for providing help with part of the software used to carry out the psychophysical

studies. Special thanks go to Warren Robinett for his constant encouragement during this research and for his help with deriving a calibration method for the system. I thank Irene Snyder for analyzing the data, and Russ Taylor and Rich Holloway for stimulating discussions. Many thanks go to Christina Burbeck for her help with interpreting the experimental results and her inspiring discussions about this research. The authors also greatly benefitted from stimulating discussions with Stanley Roscoe. A special thanks goes to Marvin Scher for editorial advice. Finally, I thank my husband, Bruce Scher, for helping with the alignment of the system and for his constant support during this work. This research was supported by NIH PO1 CA47982-04, ARPA DABT 63-93-C-0048, NSF Cooperative Agreement ASC-8920219, and ARPA: "Science and Technology Center for Computer Graphics and Scientific Visualization," and ONR N00014-94-1-0503.

References

Adelstein, B. D., Johnston, E. R., & Ellis, S. R. (1992). A testbed for characterizing the response of virtual environment spatial sensors. *The 5th Annual ACM Symposium on User Interface and Technology* (pp. 15–22). Monterey, California: ACM.

Bajura, M., Fuchs, H., & Ohbuchi, R. (1992). Merging virtual objects with the real world. *Computer Graphics (Proceedings of SIGGRAPH '92), 26*(2), 203–210.

Barrett, H. H., & Swindell, W. (1981). *Radiological imaging: The theory of image formation, detection, and processing,* Vol. 1. New York: Academic Press.

Benel, R. A. (1979). *Visual accommodation, the Mandelbaum effect, and apparent size.* Technical Report BEL-79-1/AFOSR-79-5. Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory.

Benel, R. A., & Benel, D. C. R. (1979). *Accommodation in untextured stimuli fields.* Technical Report Eng. Psy. 79-1/AFOSR-79-1. Champaign, IL: University of Illinois at Urbana-Champaign, Department of Psychology.

Burton, G. J., & Home, R. (1980). Binocular summation, fusion and tolerances. *Optica Acta, 27,* 809.

Campbell, C. J., McEachern, L. T., & Marg, E. (1955). *Flight by periscope.* Technical Report WADC TR 55-142. Wright-Patterson Air Force Base, OH: Wright Air Development Center.

Cogan, D. G. (1937). Accommodation and the autonomic nervous system. *Archives of Ophthalmology, 18,* 739–766.

Deering, M. (1992). High resolution virtual reality. *Computer Graphics, 26,* 2.

Edwards, E. K., Rolland, J. P., & Keller, K. P. (1993). Video see-through design for merging of real and virtual environments. *Proceedings of the IEEE Virtual Reality Annual International Symposium,* 223–233.

Ellis, S. R. (1993 and 1994). Personal communications.

Ellis, S. R., Tyler, M., Kim, W. S., & Stark, L. (1992). Three dimensional tracking with misalignment between display and control axes. *SAE Transactions: Journal of Aerospace, 100*(1), 985–989.

Enright, J. T. (1989). The eye, the brain, and the size of the moon: Toward a unified oculomotor hypothesis for the moon illusion. In M. Hershenson (Ed.), *The moon illusion.* Hillsdale, NJ: Erlbaum, 59–121.

Foley, J. D., & Van Dam, A. (1982). *Fundamentals of interactive computer graphics.* Reading, MA: Addison Wesley.

Foley, J. M. (1991). Stereoscopic distance perception. In S. R. Ellis (Ed.), *Pictorial communication in virtual and real environments.* Taylor and Francis. London and New York, 558–566.

Fuchs, H., Poulton, J., Eyles, J., Greer, T., Goldfeather, J., Ellsworth, D., Molnar, S., & Israel, L. (1989). Pixel-Planes 5: A heterogeneous multiprocessor graphics system using processor-enhanced memories. *Computer Graphics (Proceedings of SIGGRAPH '89), 23*(3), 79–88.

Gulick, W. L., & Lawson, R. B. (1976). *Human stereopsis.* London: Oxford University Press.

Hadani, I. (1991). Corneal lens goggles and visual space perception. *Applied Optics, 30*(28), 4136–4147.

Holloway, R. L. (1994). Dissertation in progress. University of North Carolina at Chapel Hill.

Hull, J. C., Gill, R. T., & Roscoe, S. N. (1982). Locus of the stimulus to visual accommodation: Where in the world or where in the eye? *Human Factors, 24,* 311–319.

Iavecchia, J. H., Iavecchia, H. P., & Roscoe, S. N. (1983). The moon illusion revisited. *Aviation, Space, and Environment Medicine, 54,* 39–64.

Iavecchia, J. H., Iavecchia, H. P., & Roscoe, S. N. (1988). Eye accommodation to head-up virtual images. *Human Factors, 30,* 689–702.

Janin, A. L., Mizell, D. W., & Caudell, T. P. (1993). Calibration of head-mounted displays for augmented reality applications. *Proceedings of the IEEE Virtual Reality Annual International Symposium,* 246–255.

Kotulak, J. C., & Morse, S. E. (1994). Relationship among

accommodation focus, and resolution with optical instruments. *Journal of the Optical Society of America A, 11,* 71–79.

Lackner, J. R., & Dizio, P. (1989). Altered sensory-motor control of the head as an etiological factor in space-motion sickness. *Perceptual and Motor Skills, 68,* 784–786.

Leibowitz, H. W., & Owens, D. A. (1975a). New evidence for the intermediate position of relaxed accommodation. *Documenta Ophthalmologica, 46,* 133–147.

Leibowitz, H. W., & Owens, D. A. (1975b). Anomalous myopias and the intermediate dark focus of accommodation. *Science, 189,* 646–648.

Levine, M. D. (1985). *Vision in man and machine.* New York: McGraw-Hill.

Lockhead, G. D., & Wolbarsht, M. L. (1989). The moon and other toys. In M. Hershenson (Ed.), *The moon illusion.* Hillsdale, NJ: Erlbaum.

Longhurst, R. S. (1973). *Geometrical and physical optics.* New York: Longman.

Mandelbaum, J. (1960). An accommodation phenomenon. *Archives of Ophthalmology, 63,* 923–926.

McKay, H. C. (1948). *Principles of stereoscopy.* Boston: American Photographic Publishing Company.

Norman, J., & Ehrlich, S. (1986). Visual accommodation and virtual image displays: Target detection and recognition. *Human Factors, 28,* 135–151.

Ogle, K. N. (1962). Spatial localization through binocular vision. In *The eye* (Vol. 4, pp. 271–324). New York: Academic Press.

Palmer, E., & Cronn, F. W. (1973). Touchdown performance with computer graphics night visual attachment. In *Proceedings of the AIAA Visual and Motion Simulation Conference,* 1–6. New York: American Institute of Aeronautics and Astronautics.

Patterson, R., & Martin, W. L. (1992). Human stereopsis. *Human Factors, 34,* 669–692.

Patterson, R., Moe, L., & Hewitt, T. (1992). Factors that affect depth perception in stereoscopic displays. *Human Factors, 34,* 655–667.

Randle, R. J., Roscoe, S. N., & Petitt, J. (1980). *Effects of accommodation and magnification on aimpoint estimation in a simulated landing task.* Tech Paper NASA-TP-1635. Washington DC: National Aeronautics and Space Administration.

Reading, R. W. (1983). *Binocular vision.* London: Butterworths.

Robinett, W., & Holloway, R. (1994). The visual display computation for virtual reality. Submitted for publication.

Robinett, W., & Rolland, J. P. (1992). A computational model for the stereoscopic optics of a head mounted display. *Presence, 1*(1), 45–62.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8,* 125–134.

Rolland, J. P., & Hopkins, T. (1993). *A method of computational correction for optical distortion in head-mounted displays.* Technical Report TR-93-045, Dept. of Computer Science, University of North Carolina at Chapel Hill.

Rolland, J. P., Holloway, R. L., & Fuchs, H. (1994). A comparison of optical and video see-through head-mounted displays. *Proc. Soc. Photo-Opt. Instrum. Eng.,* forthcoming.

Roscoe, S. N. (1948). The effect of eliminating binocular and peripheral monocular visual cues upon airplane pilot performance in landing. *Journal of Applied Psychology, 32,* 649–662.

Roscoe, S. N. (1979). When day is done and shadows fall, we miss the airport most of all. *Human Factors, 21*(6), 721–731.

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

Roscoe, S. N. (1985). Bigness is in the eye of the beholder. *Human Factors, 27*(6), 615–636.

Roscoe, S. N. (1987). Improving visual performance through volitional focus control. *Human Factors, 29*(3), 311–325.

Roscoe, S. N. (1991). The eyes prefer real images. In S. R. Ellis (Ed.), *Pictorial communication in virtual and real environments.* Taylor and Francis. London and New York, 577–585.

Roscoe, S. N. (1993). Visual orientation: facts and hypotheses. *International Journal of Aviation Psychology, 3,* 221–229.

Roscoe, S. N., Hasler, S. G., & Dougherty, D. J. (1966). Flight by periscope: Making tradeoffs and landings; the influence of image magnification, practice, and various conditions of flight. *Human Factors, 8,* 13–40.

Roscoe, S. N., Olzak, L. A., & Randle, R. J. (1976). Ground-referenced visual orientation with imaging displays: monocular versus binocular accommodation and judgments of relative size. *Proceedings of the AGARD Conference on Visual Presentation of Cockpit Information Including Special Devices for Particular Conditions of Flying* (pp. A5.1–A5.9). Neuilly-sur-Seine, France: NATO.

Sheehy, J. B., & Wilkinson, M. (1989). Depth perception after prolonged usage of night vision goggles. *Aviation, Space, and Environment Medicine, 60,* 573–579.

Simonelli, N. M. (1979). Apparent size and visual accommodation under day and night conditions. *Proceedings of the Hu-*

man Factors Society 23rd Annual Meeting. Santa Monica, CA: Human Factors Society, 374–378.

Simonelli, N. M. (1980). *The dark focus of visual accommodation: Its existence, its measurements, its effects.* Doctoral dissertation, University of Illinois at Urbana-Champaign.

Virtual Research. Flight Helmet/3193 Belick Street #2, Santa Clara, CA 95054.

Vos, J. J. (1960). Some new aspects of color stereoscopy. *Journal of the Optical Society of America, 50,* 785–790.

Weisheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. *Journal of Physiology (London), 370,* 619–629.

Woodson, W. E. (1992). *Human factors design handbook,* 2nd ed. New York: McGraw-Hill.