Multifocal planes head-mounted displays

Jannick P. Rolland, Myron W. Krueger, and Alexei Goon

Stereoscopic head-mounted displays (HMD's) provide an effective capability to create dynamic virtual environments. For a user of such environments, virtual objects would be displayed ideally at the appropriate distances, and natural concordant accommodation and convergence would be provided. Under such image display conditions, the user perceives these objects as if they were objects in a real environment. Current HMD technology requires convergent eye movements. However, it is currently limited by fixed visual accommodation, which is inconsistent with real-world vision. A prototype multiplanar volumetric projection display based on a stack of laminated planes was built for medical visualization as discussed in a paper presented at a 1999 Advanced Research Projects Agency workshop (Sullivan, Advanced Research Projects Agency, Arlington, Va., 1999). We show how such technology can be engineered to create a set of virtual planes appropriately configured in visual space to suppress conflicts of convergence and accommodation in HMD's. Although some scanning mechanism could be employed to create a set of desirable planes from a two-dimensional conventional display, multiplanar technology accomplishes such function with no moving parts. Based on optical principles and human vision, we present a comprehensive investigation of the engineering specification of multiplanar technology for integration in HMD's. Using selected human visual acuity and stereoacuity criteria, we show that the display requires at most 27 equally spaced planes, which is within the capability of current research and development display devices, located within a maximal 26-mm-wide stack. We further show that the necessary in-plane resolution is of the order of 5 μm. © 2000 Optical Society of America

OCIS codes: 080.0080, 120.2040, 120.2820, 120.4640, 230.3990, 350.4600.

1. Introduction

Perhaps surprisingly, the vast majority of deployed virtual reality systems present the same images to both eyes. Such nonstereoscopic binocular systems provide neither a change in accommodation nor convergence. Elite systems that can afford two separate graphics generator, and thus provide a distinct image for each eye, are known as stereoscopic head-mounted displays (HMD's), or stereoscopic binocular HMD's. A comprehensive discussion of the trade-offs in designing binocular HMD's is given in Kocian (1988).[1] The conventional approach is to employ two-dimensional (2-D) miniature displays or the equivalent near or at the focal plane of an eyepiece to provide stereoscopic vision. Examples of miniature 2-D displays are 0.5–1-in. (1.27–2.54-cm) CRT displays or 1-in. (2.54-cm) transmissive or reflective flat-panel displays. In some cases, one can construct an equivalent miniature 2-D display using the end of a fiber-optic bundle that transports the image from a high-resolution light valve to the HMD.[2] The miniature 2-D displays are typically imaged through magnifying optics to create a virtual image for each eye at some fixed distance in space. That distance can vary from system to system, but it is critical to observe that it is fixed in space for any given system. When focused at that distance, a HMD user perceives the monocular images with optimal image sharpness; and by visually fusing the images in his brain, he can perceive objects in three dimensions. Under such image presentation, a conflict of accommodation and convergence necessarily arises.

The conflict can be summarized as follows. Stereoscopic HMD's require convergent eye movements to yield single vision of an object of interest located in depth. Similar to real-world vision, three-dimensional (3-D) objects located at other depths than the object of interest are seen double and thus are referred to as diplopic objects.[3,4] Contrary to real-world vision, however, these diplopic objects are seen sharp. Indeed, visual accommodation in such displays is intrinsically restricted to

---

J. P. Rolland and A. Goon are with the School of Optics, Center for Research and Education in Optics and Lasers, University of Central Florida, Orlando, Florida 32816-2700. M. W. Krueger is with Artificial Reality Corporation, Vernon, Connecticut 06066.

Received 27 July 1999; revised manuscript received 25 February 2000.

0003-6935/00/193209-07$15.00/0

one plane for optimal image sharpness. So all objects, regardless of their location in depth, are seen in focus if the user focuses on the virtual monocular images. If the user's accommodation varies with eye convergence, then all objects, including those gazed at, are seen blurred. Therefore, even the highest-resolution displays would not yield 3-D vision as experienced in the real world. So, although we shall continue the quest for sharper displays and ergonomic designs, it is necessary to also ask whether HMD's can be engineered to shun the conflict of accommodation and convergence.[5,6]

Although conflicts exist, they can be minimized for specific applications when the location of the monocular virtual images are set according to the tasks to be performed.[7] In HMD's for air pilots, for example, collimated images are typically preferred because of the need to visualize information essentially in the far field. In this case, the 2-D displays are best located at the focal plane of the eyepieces. In contrast, virtual reality systems for medical or engineering applications are likely to require visualization at arm's length, thus in the near field to allow manipulation of objects. In this case, the 2-D displays are best located within the focal distance of the eyepieces to create virtual images at a finite, roughly arm's length, distance from the user.

The challenge comes with applications whose tasks require visualization of the near and far field simultaneously, or any set of depth planes in between. Under such requirements, the conflicts cannot be minimized. There are many applications that have these requirements. For example, operating a car requires navigating in the far field, looking for targets that may be located anywhere in space, and simultaneously attending to instrumentation in the near field. In this case, one could argue that the instrumentation could be projected to the far field as it is for a pilot. Implementation of these tasks in a virtual environment can be perceptually challenging. Similarly, as we walk around a scene we are constantly moving our eyes around, looking out for obstacles that may be at different distances. Concordant accommodation and convergence is necessary for such tasks. In other applications such as athletics, out of focus may in fact be a common mode of operation. Athletes may never have time to converge and accommodate on particular objects. Thus athletes operate under mainly blurred vision. A virtual world where all objects would always be in focus may affect training under realistic conditions. There are other tasks such as the close examination of 3-D objects that may involve, to the contrary, many small focus adjustments. Those would not be possible with conventional HMD's.

Conflicts of accommodation and convergence in HMD's thus result from the inability to provide realistic accommodation cues.[6–9] Although accommodation is a weak cue in itself for the perception of depth compared with other cues (e.g., occlusion, head motion parallax, stereopsis), when we look around in the real world, it is not all in focus at once.[10–12] In an earlier paper we highlighted potential problems that HMD's could place on the accommodation and convergence system.[7] Studies of binocular stress have followed.[9] Thus the requirement for accommodation as well as for convergence is a necessary ingredient for the synthetic representation of realistic scenes.

In this paper we explain how multiplanar technology may provide an attractive solution to the problem. Limitations are also addressed, and the approach is situated in the context of previous research. Finally, a comprehensive engineering study of existing multiplanar technology for integration in HMD's is given.

2. Proposed Solution and Previous Research

A solution to matching convergence and accommodation is to make the HMD produce multiple image planes. This can be accomplished with multiplanar technology that consists of a stack of planar arrays, the pixels of which would be presented according to the distance of the simulated object. Although it might seem that scanning multiple displays would present a problem, this is not so. As part of the calculation of a rendered computer-graphics image, a Z buffer is created that has a distance value for each pixel currently displayed. Instead of ignoring this construct during the display process, we can use it to determine which of the display planes each pixel is written to as it is scanned. The fundamental of graphics can be found in Foley *et al.* (1990).[13] In all other planes, the visually aligned pixels would remain transparent. To ensure correct transparency, eye tracking must be performed as well.[14–16] In the case of no eye-tracking capability, simpler environments can be considered or respective occlusion of objects can be computed based on where in the scene the observer will most likely gaze. The application and tasks to be performed must guide the approach and technology requirements.

The approach of multiplanar technology is attractive in that it does not require any moving elements. The feasibility of the approach in itself comes from the observation that, in HMD optics, a small movement of the actual screens toward the viewer's eye or away from it leads to a large displacement in depth of the optical images. This results from working at magnifications greater than one, given that the longitudinal magnification in optical systems varies as the square of the transverse magnification. Therefore it is possible to envision a miniature thick-display system that can adjust the depth of the displayed objects on a plane-by-plane basis. The principle of operation of multiplanar technology for focusing in depth is shown in Fig. 1.

The multifocal approach to the problem is not new. Wann *et al.* addressed the conflicts of accommodation and convergence in HMD's and suggested varying the focal depth either by using an oscillating lens or by adjusting the image depth based on the user's gaze point.[17] Marran and Schor have also outlined various focal solutions for virtual reality systems.[18] Solutions included designing pinhole optics to create an

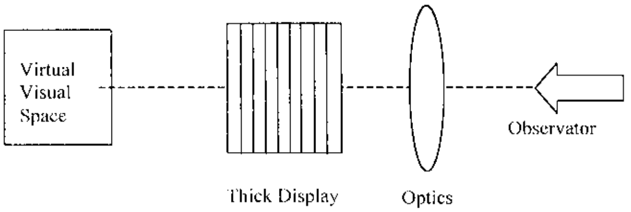


Fig. 1. Principle of multifocal planes HMD.

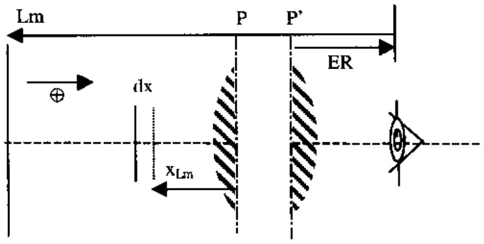


Fig. 2. Basic layout of the imaging optics in a HMD.

infinite depth of focus for the visual system, adding a different monocular lens in front of each eye to focus each eye at a different depth, or using bifocal lenses or the like to focus at various depths in various parts of the field of view. Although these proposed solutions may find specific applications, none of them intends to provide a general solution to suppressing conflicts of convergence and accommodation in HMD's.

From a conceptual point of view, research that is closely related to the multiplanar approach is the real-depth imaging proposed by Dolgoff.[19] It consists in providing floating images for computer, television, and projection applications without recourse to either stereoscopy or autostereoscopy technology. The technique consists in dividing a scene into a foreground and a background image with each placed on a different plane. The basis of Dolgoff's technique was to use accommodation cues in addition to 2-D cues such as perspective and background-object occlusion to provide a sense of three dimensions from one single image presented to the eyes. For scientific visualization, we must establish the rationale for the number of planes, the interplane spacing, and the planes' resolution.

To expand on the concept of multiple planes, a volumetric projection display based on multiplanar technology was developed for surgical simulation and training, telepresence surgery, and medical volume visualization.[20,21] The approach provides an alternative to true 3-D displays.[22–24] In a volumetric projection display, a series of 2-D images is projected at 30 Hz into a multiplanar optical element (MOE). The MOE acts as a variable depth projection screen that is synchronized to the projector frame rate. The first MOE covered a 6 in. × 6 in. × 3 in. (15.24 cm × 15.24 cm × 7.62 cm) volume with a 480 × 480 × 12 plane resolution, totaling 2.76 million voxels, each independently and simultaneously addressable.[21] A 6.5-bit gray scale was achieved. Some multiplanar antialiasing algorithms provide improved subjective 3-D continuity. The current generation display that is being developed covers a 15.5 in. × 13.5 in. × 9 in. (39.8 cm × 34.3 cm × 23 cm) volume with a 512 × 512 × 50 plane resolution, totaling 13.1 million voxels.[25] A 24-bit gray scale is being implemented. The system is designed to operate at 30-Hz update rates. Sullivan[25] recently proposed using a high frame-rate image projector to illuminate the MOE. Spatial light modulators, which spatially modulate an incident collimated beam, can also be used to provide the display resolution and speed (≥30 Hz) if not the contrast ratio. The multiplanar technology must be miniaturized and engineered for integration in HMD's.

3. Range of Multiplanes Focusing

We first establish the thickness dx of the multiplanes display that is needed to cover a given range of accommodation. We consider the imaging equation given by

$$\frac{n'}{x'} = \frac{n}{x} - \frac{n}{f}, \tag{1}$$

where n and n' are the indices of refraction in object and image space of the imaging optics, respectively; x and x' are the distances of the object (i.e., the miniature display) and the image (i.e., the virtual image plane) with respect to the principal planes P and P', respectively; and f is the focal length of the imaging optics. We denote as x_{Lm} the value of x that corresponds to an image located at the closest point of accommodation—i.e., *punctum proximum*—*Lm*. To focus, x varies from f to x_{Lm}, and x' varies accordingly from infinity to Lm-ER where ER is the eye relief measured from P' for simplicity and generality as shown in Fig. 2.

The focal plane in object space is the reference plane from which the range of focusing dx is measured. By manipulation of Eq. (1), dx is given by

$$dx = x_{Lm} - f = -\frac{f^2}{f + L_m - \mathrm{ER}}. \tag{2}$$

If P' is located close to the last optical surface of the optics, |ER| equal to 25 mm will allow the wearing of a wide variety of eyeglasses. The shorter the focal length and the larger the value of Lm, the less ranging is required. The range of focusing dx ranges from approximately 3.5 to 26 mm for focal lengths between 30 and 90 mm. Three-dimensional focal plane arrays of such sizes can currently be implemented in various materials.

4. Number of Planes for Complete Range Focusing

The minimum number of planes required to focus from a nearest plane to infinity is determined by the available range of accommodation and the depth of focus of the human visual system on each side of a plane of fixation. Specifically, by expressing the distal depth of focus dL_+ and the proximal depth of focus

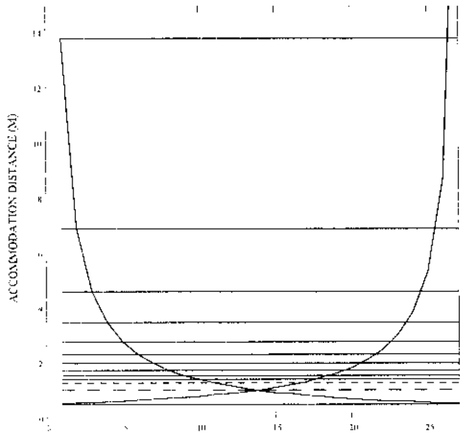


Fig. 3. Location of the planes of fixation for accommodation based on the computed depth of focus planes for the human visual system. A visual acuity of 1 arc min is assumed. Both schemes starting from *Lm* equal 0.5 m to infinity (i.e., solid curve decreasing from left to right) and from infinity to 0.5 m (i.e., solid curve increasing from left to right) were considered. In both cases, we find that 27 planes are required for a range of accommodation from infinity to 0.5 m. A few of these planes are represented as horizontal lines in the figure.

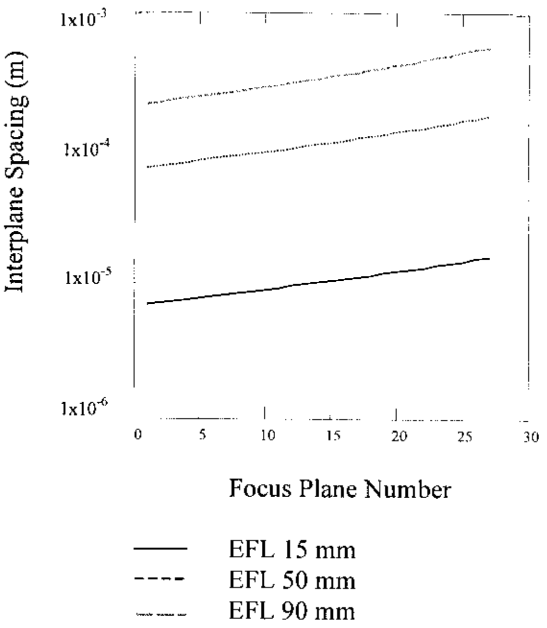


Fig. 4. Interplane spacing as a function of the focus plane number for three different values of the effective focal length (EFL).

dL_- with respect to the fixation plane as a function of the size d of the pupil of the user's eye, and the distance L of accommodation, such that a defocused point on either side of the fixation plane would subtend an angle equivalent to the human visual acuity η, we obtain

$$|dL_\pm| = \left|\frac{\eta L^2}{d \pm \eta L}\right|. \tag{3}$$

Note that L is negative in Eq. (2), yielding $|dL_+| > |dL_-|$ as also observed in Fig. 3. Based on a value of visual acuity of 1 arc min, Eqs. (3) and (2) combined yield 27 planes from infinity to 0.5 m. A few planes are represented in Fig. 3. Based on this theoretical prediction, only 14 planes would be needed because every other plane can be selected as a plane of accommodation.

5. Interplane Spacing

Using Eq. (1), we mapped the values of L back in object space to dx values. The interplane spacing is given by the consecutive differences of the computed dx values. The interplane spacing as a function of the plane number is shown in Fig. 4 for the focal length values of 15, 50, and 90 mm. The interspacing is quasi-constant, and a constant spacing that was chosen to be that computed for plane number one can serve as a practical solution.

Although the smallest interplane spacing is approximately 10 μm for an extremely short focal length, typical values are an order of magnitude larger. Interplane spacing of the order of 100 μm is more readily feasible.

6. Resolution Requirement

We now ask how many depth units can be resolved within the range of accommodation imposed by the depth of focus of the human visual system around a fixation plane. The number of resolvable units sets the requirements for the in-plane resolution of the displays. A HMD user has stereoscopic information available from the disparate images provided to the two eyes. Binocular disparity can be defined as the angular disparity δ between any two object points in the field of view.[26] Simple geometry yields an expression for δ_+ behind the fixation plane (i.e., distal) and δ_- in front of the fixation plane (i.e., proximal) given by

$$\delta_\pm = \pm \frac{\Delta l \,\mathrm{IOD}}{(L \mp \Delta l)L}, \tag{4}$$

where IOD is the interocular distance and Δl is the resolvable depth at a given fixation distance L taken to be negative following our sign convention shown in Fig. 2. Given a value of δ, L, and IOD, the solution for Δl yields Δl_+ and Δl_- on the distal and proximal side, respectively, whose expressions are given by

$$|\Delta l_\pm| = \frac{|\delta| L^2}{\mathrm{IOD} \pm |\delta| L}. \tag{5}$$

Also note that according to sign conventions, δ is positive and negative on the distal and the proximal sides, respectively. Thresholds for stereoacuity δ vary widely between users in the extreme between approximately 2 and 130 arc sec.[26] Both extreme values are considered in addition to a more typical value of 30 arc sec. It has been shown that δ is constant over distance.[27] We next consider the dis-

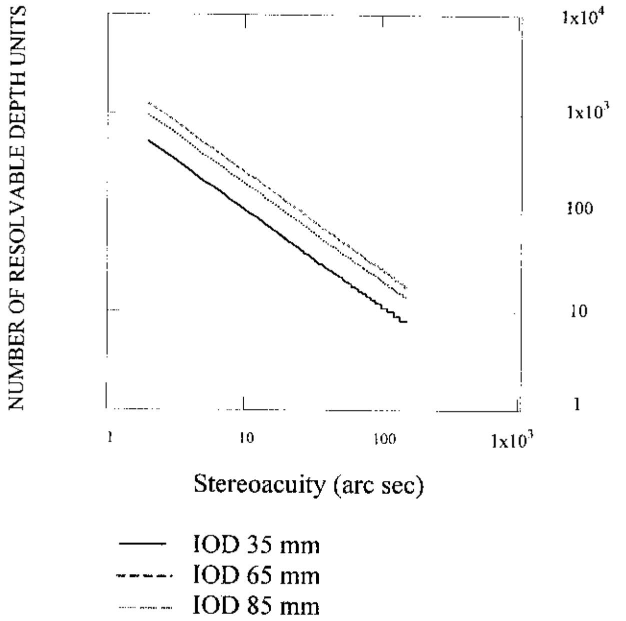


Fig. 5. Number of resolvable depth units as a function of stereoacuity for three values of the user interocular distance.

tal value of Δl_+ in computing the resolvable depth units. L is recursively given by

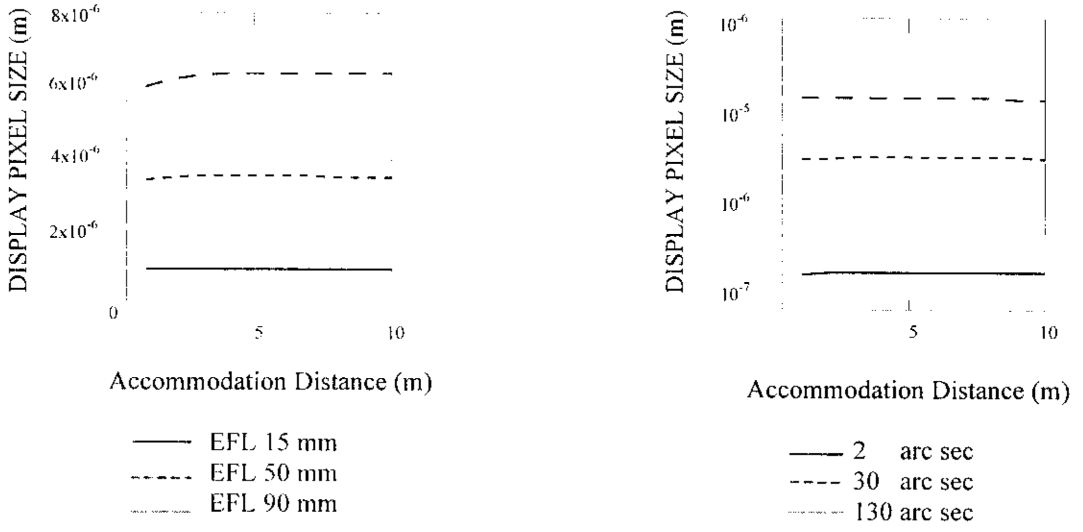
$$L_k = L_{k-1} + \Delta l_{k-1} = L_{k-1} + \frac{|\delta| L^2_{k-1}}{\mathrm{IOD} + |\delta| L_{k-1}}. \qquad (6)$$

For computational purpose, the IOD is set to an average value of 65 mm. Using Eqs. (5) and (6), we show the number of distances N resolved in depth as a function of δ for three values of observer IOD in Fig. 5. N is constant for all planes of accommodation previously computed, and these resolvable distances in depth are found to be equally spaced in the object space of the display device.

The number of resolvable depth units between the discrete set of planes of accommodation shown in Fig. 3 imposes some requirement on the resolution of the display device. Based on the smallest resolvable depth around any fixation plane, we can derive the required display resolution p by mapping a pixel on the display in image space using the lens magnification and the imaging equation given by Eq. (1) and by expressing the fact that the pixel size in image space is equal to its angular subtend multiplied by the distance of accommodation L, where the angular subtend is set by stereoacuity. After laying out such conditions, we can define p by

$$p = \frac{\mathrm{IOD}}{2(L + \Delta l)\left(1 + \dfrac{L}{f}\right)} \left(\frac{|\delta| L^2}{\mathrm{IOD} - |\delta| L}\right). \qquad (7)$$

The values of p as a function of the accommodation distance for three focal lengths at a stereoacuity of 30 arc sec, as well as a function of stereoacuity for a focal length of 50 mm, are shown in Fig. 6(a) and 6(b), respectively. Except at the highest resolution of 2 arc sec or for a short focal length (e.g., 15 mm) where the resolution requirement is quite stringent (i.e., 1 μm), a display resolution of approximately 5 μm is required. Such resolutions are achievable with current technology.



(a) (b)

Fig. 6. Display resolution requirements. (a) Plot of the display resolution required as a function of the distance of accommodation L for three values of the effective focal length (EFL): 15, 50, and 90 mm. The stereoacuity threshold δ is 30 arc sec in this computation. (b) Plot of the display resolution required as a function of the distance of accommodation L for three values of the stereoacuity threshold δ: 2, 30, and 130 arc sec. In this case the focal length is set to 50 mm.

7. Conclusions

In conceiving and developing technology, we must ask both whether and to what extent perturbations in a fundamental mode of operation affect task performance and how a failure to provide a natural visual experience may cause difficulties for the visual system over an extended period of exposure. In this paper we addressed how the current conflicts of convergence and accommodation in HMD's can be suppressed. Given that current virtual reality exposure is measured in minutes rather than hours, there is no reason to be alarmed in the short run. However, if we contemplate a future in which virtual reality is a ubiquitous feature of video games, viewing conflicting visual cues may cause changes to the visual system over years of use. Therefore it behooves us to design displays that incorporate all visual cues as realistically as possible to ensure safety as well as to make the viewing experience most believable.

To resolve conflicts of convergence and accommodation in HMD's, we have proposed to make the miniature displays multifocal. We presented a comprehensive investigation of engineering requirements for adding multiplanes focusing capability to HMD's. For the parameters considered, the range of focusing to accommodate from infinity to 0.5 m goes from approximately 3.5 to 26 mm. The number of planes within this range is 27, which can be reduced to 14 for a standard visual acuity of 1 arc min and a 4-mm pupil diameter. Although 14 is the minimum theoretical value, it remains to be experimentally determined how many planes should actually be adopted for various applications. The framework we laid out will allow one to compute the number of planes required. Under the most stringent conditions imposed by this theoretical study, the interplane spacing is found to be constant and may be as small as 10 μm, but more typically approximately 100 μm. Finally, stereoacuity imposes that the transverse resolution of the display be of the order of 5 μm. Based on this investigation, we conclude that the addition of multiplanes focusing to HMD's may be challenging but nevertheless realizable with today's technology.

This research was supported by a First Award 1-R29-LM06322-01A1 from the National Institute of Health, Office of Naval Research grant N00014-97-1-0654, and Defense Advanced Research Projects Agency grant DAMD17-94-J-4113. We thank Allan Sullivan for providing us a description and update on multiplanar technology. We also thank Laser Power Corporation for stimulating discussions on providing accommodation cues in HMD's.

References

1. D. F. Kocian, "Design considerations for virtual panoramic display (VPD) helmet systems," Armstrong Aerospace Medical Research Laboratory, Visual Display Systems Branch (Wright-Patterson Air Force Base, Dayton, Ohio, 1988).
2. M. L. Thomas, R. Robinson, W. P. Siegmund, and S. E. Antos, "Fiber optic development for use on the fiber optic helmet-mounted display," Opt. Eng. **29,** 855–862 (1990).
3. A. Arditi, "Binocular vision," in *Handbook of Perception and Human Performance* (Wiley Interscience, New York, 1986), Chap. 23, p. 32.
4. J. Siderov and R. S. Harwerth, "Precision of stereoscopic depth perception from double images," Vision Res. **33,** 1553–1560 (1993).
5. C. Roumes, J. Plantier, and J. P. Menu, "L'avenir est-il au visual de casque binoculaire?" in *AGARD CP 517* (Advisory Group for Aerospace Research and Development, Paris, France, 1992), Vol. 11, pp. 11–19.
6. J. Wann, S. Rushton, and M. Mon-Williams, "Natural problems in the perception of virtual environments," Vision Res. **35,** 2731–2736 (1995).
7. J. P. Rolland, W. Gibson, and D. Ariely, "Towards quantifying depth and size perception in virtual environments," *Presence: Teleoperators Virtual Environ.* **4**(1), 24–29 (1995).
8. W. Robinett and J. P. Rolland, "A computational model for the stereoscopic optics of a head-mounted display," *Presence: Teleoperators Virtual Environ.* **1**(1), 45–62 (1992).
9. E. Peli, "Real vision and virtual reality," Opt. Photon. News, July 1995, pp. 28–34.
10. M. Mon-Williams, J. P. Wann, and S. Rushton, "Binocular vision in a virtual world: visual deficits following the wearing of head-mounted displays," Ophthalmic Physiol. Opt. **13,** 387–391 (1993).
11. W. C. Gogel and J. D. Tietz, "Relative cues and absolute distance perception," Percept. and Psychophys. **28,** 321–328 (1980).
12. H. A. Sedgwick, "Space perception," in *Handbook of Perception and Human Performance* (Wiley Interscience, New York, 1986), Chap. 4, p. 6.
13. J. Foley, A. vanDam, S. Feiner, and J. Hughes, *Computer Graphics: Principles and Practice*, 2nd ed. (Addison-Wesley, Reading, Mass., 1990).
14. J. P. Rolland, A. Yoshida, L. Davis, and J. H. Reif, "High-resolution inset head-mounted display," Appl. Opt. **37,** 4183–4193 (1998).
15. L. Vaissie and J. P. Rolland, "Eye-tracking integration in head-mounted displays," U.S. patent (Filed 23 December 1999).
16. J. P. Rolland, L. Vaissie, and G. Bochenek, "Analysis of eye-point locations and accuracy of rendered depth in binocular head-mounted displays" in *Stereoscopic Displays and Virtual Reality Systems VI*, J. O. Merritt, M. T. Bolas, and S. S. Fisher, eds., Proc. SPIE **3639A,** 57–64 (1999).
17. J. P. Wann, S. K. Rushton, and M. Mon-Williams, "Natural problems for stereoscopic depth perception in virtual environments," Vision Res. **35,** 2731–2736 (1995).
18. L. Marran and C. Schor, "Multiaccommodative stimuli in VR systems: problems and solutions," Hum. Factors **37,** 382–388 (1997).
19. E. Dolgoff, "Real-depth imaging," SID Digest **28,** 269–272 (1997).
20. A. F. Crabtree, "Method and apparatus for manipulating, projecting, and displaying light in a volumetric format," U.S. patent 5,572,375 (5 November 1996).
21. A. Sullivan, "Progress report on multiplanar volumetric displays," presented at the Defense Advanced Research Projects Agency Biomedical Computing Workshop, Innovative Modeling & Advanced Generation of Environments Conference, Scottsdale, Ariz., 14–18 July 1997.
22. J. D. Lewis, C. M. Verber, and R. B. MacGhee, "A true three-dimensional display," IEEE Trans. Electron Devices **18,** 724–732 (1971).
23. E. Downing, L. Hesselink, J. Ralston, and R. Macfarlane, "A

three-color, solid-state, three-dimensional display," Science **273,** 1185–1189 (1996).

24. I. I. Kim, E. Korevar, and H. Hakakha, "Three-dimensional volumetric display in rubidium vapor," in *Projection Displays II*, M. H. Wu, ed., Proc. SPIE **2650,** 274–284 (1996).
25. A. Sullivan, "Multiplanar volumetric displays for medical visualization," presented at Medicine Meets Virtual Reality '99, Advanced Research Projects Agency Workshop, San Francisco, 20 Jan. 1999.
26. R. W. Reading, *Binocular Vision: Foundations and Applications* (Butterworths, London, 1983), 110–128.
27. A. Lit, and J. P. Finn, "Variability of depth-discrimination thresholds as a function of observation distance," J. Opt. Soc. Am. A **66,** 740–742 (1976).