Geometric Considerations for Stereoscopic Virtual Environments

by

Larry F. Hodges and
Elizabeth Thorpe Davis

GIT-GVU-93-02
January 1993

Graphics, Visualization & Usability Center

Georgia Institute of Technology
Atlanta GA 30332-0280

GEOMETRIC CONSIDERATIONS FOR STEREOSCOPIC VIRTUAL ENVIRONMENTS

Larry F. Hodges
College of Computing
(404) 894-8787 hodges@cc.gatech.edu

Elizabeth Thorpe Davis
School of Psychology
(404) 853-0194 ed15@prism.gatech.edu

Graphics, Visualization & Usability Center
Georgia Institute of Technology
Atlanta, GA 30332-0280

ABSTRACT

We examine the relationship between the different geometries implicit in a stereoscopic virtual environment. In particular, we examine in detail the relationship of retinal disparity, fixation point, binocular visual direction, and screen parallax. We introduce the concept of a volumetric spatial unit called a stereoscopic voxel. Due to the shape of stereoscopic voxels, apparent depth of points in space may be affected by their horizontal placement.

1.0 INTRODUCTION

The display component of the most common implementations of virtual environments provides the user with a visual image that incorporates stereopsis as a visual cue. Examples include head-mounted displays (Teitel, 1990), time-multiplexed CRT-based displays (Deering, 1992), and time-multiplexed projection systems (Cruz-Neira, et al., 1992). It is not always recognized, however, that the characteristics of a stereoscopic image can be very different from that of a monoscopic perspective image. The visual impression given by stereoscopic images are very sensitive to the geometry of the visual system of a user, the geometry of the display environment, and the modeling geometry assumed in the computation of the scene. In this paper we present a tutorial whose purpose is to review these basic geometries and to analyze their relationship to each other in a stereoscopic virtual environment.

In section 2.0 we review the geometry of binocular vision and retinal disparity. In section 3.0 we discuss modeling geometry and its relationship to retinal disparity. Section 4.0 describes the effects of the discrete nature of display geometry and the distortions causes by optical and tracking artifacts. In section 5.0 we revisit our approximate model of the visual system geometry and discuss its limitations.

2.0 VISUAL SYSTEM GEOMETRY

Stereopsis results from the two slightly different views of the external world that our laterally-displaced eyes receive (e.g., Schor, 1987; Tyler, 1983). This difference is quantified in terms of retinal disparity. If both eyes are fixated on a point, f1, in space, then an image of f1 is focused at corresponding points in the center of the fovea[1] of each eye. Another point, f2, at a different spatial location, would be imaged at points in each eye that may not be the same distance from the fovea. This difference in distance is the retinal disparity. Normally we measure this distance as the sum of the angles $\delta_1 + \delta_2$ as shown in Figure 1, where angles measured from the center of fovea toward the outside of each

[1]The fovea is the part of the human retina that possesses the best spatial resolution or visual acuity.

eye are negative. In the example shown in Figure 1, δ_1 has a positive value, δ_2 has a negative value, and their sum $\delta_1 + \delta_2$ is positive.

The retinal disparity that results from these two different views can provide information about the distance or depth of an object as well as about the shape of an object. Stereoacuity is the smallest depth that can be detected based on retinal disparity. In some humans, under optimal stimulus conditions, stereoacuities of 5" or less can be obtained (Westheimer and McKee, 1980; McKee, 1983), however stereoacuities of more than 5" are more common (Davis, King and Anoskey, 1992; Shor and Wood, 1983).

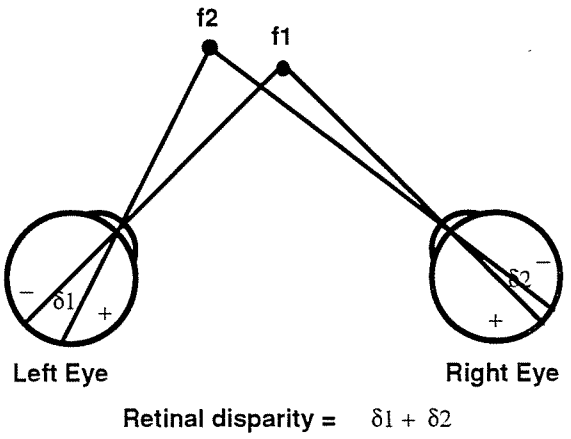

Figure 1. Retinal disparity.

2.1 Binocular visual direction.

Visual direction is the perceived spatial location of an object relative to the observer. Usually, it is measured in terms of azimuth (left and right of the point of fixation) and of elevation (above and below the point of fixation). Often, the binocular visual direction of an object is not the same as the monocular visual direction of either eye. (You can verify this yourself by looking at a very close object first with one eye, then with the other eye, then with both eyes. Notice how the spatial location of the object changes.) Hering proposed that binocular visual direction will lie midway between the directions of the monocular images; others have reported that the binocular visual direction will lie somewhere between the left and right monocular visual directions, but not necessarily midway (e.g., Tyler, 1983).

2.2 Convergence Angles and Retinal Disparities.

For symmetric convergence of the two eyes on a fixated point in space, f_1, the angle of convergence is defined as

$$\alpha = 2 \arctan(i/2D_1)$$

where α is the angle of convergence, D_1 is the distance from the midpoint of the interocular axis to the fixated point, f_1, and i is the interocular distance (Arditi, 1986; Graham, 1965).[2] (See Figure 2.) Notice that the angle of convergence, α, is inversely related to the distance, D_1, of the fixated point from the observer; this inverse relation is nonlinear.

For another point in space, f_2, located at a distance D_2, the angle of convergence is β. (See Figure 2.) Note that since

$\alpha + a + c + b + d = 180$ and

$\beta + c + d = 180$, then

$\alpha - \beta = -a + -b = (\delta 1 + \delta 2)$.

The difference in vergence angles ($\alpha - \beta$) is equivalent to the retinal disparity between the two points in space, measured in units of visual angle. This retinal disparity is monotonically related to the depth between the two points in space (i.e., at a constant distance, D_1, a larger depth corresponds to a

[2]For asymmetric convergence of the two eyes, the formula for the angle of convergence is basically the same as that shown for symmetric convergence. The difference is that D now represents the perpendicular distance from the interocular axis to the frontoparallel plane that intersects the asymmetrically converged point of fixation. This interpretation of the convergence angle formula for asymmetric convergence is not exact, but it is a good approximation.

larger retinal disparity); this monotonic relationship is a nonlinear one.

If an object is closer than the fixation point, the retinal disparity will be a negative value. This is known as a *crossed* disparity because the two eyes must cross to fixate the closer object. Conversely, if an object is farther than the fixation point, the retinal disparity will be a positive value. This is known as *uncrossed* disparity because the two eyes must uncross to fixate the farther object. An object located at the fixation point or whose image falls on corresponding points in the two retinae has a *zero* disparity.

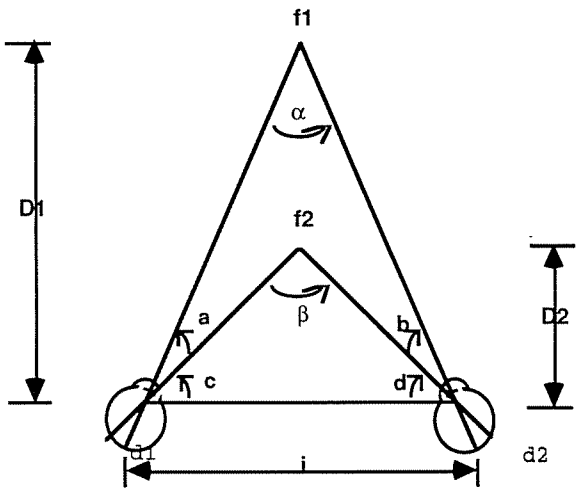

Figure 2. Convergence Angles

2.3 Horopters

Corresponding points on the two retinae are defined as being the same vertical and horizontal distance from the center of the fovea in each eye (e.g., Tyler, 1983; Arditi, 1986; Davis & Hodges, 1993). When the two eyes binocularly fixate on a given point in space, there is a locus of points in space that fall on corresponding points in the two retinae. This locus of points is the horopter, a term originally used by Aguilonius in 1613. The horopter can be defined either theoretically or empirically.

The *Vieth-Mueller Circle* is a theoretical horopter defined only in terms of geometrical considerations. This horopter is a circle in the horizontal plane that intersects each eye at the first nodal point of the eye's lens system (Gulick & Lawson, 1974; Ogle, 1968) (see Figure 3). This circle defines a locus of points with zero disparity. However, in devising this theoretical horopter it is assumed that the eyes are perfect spheres with perfectly spherical optics and that the eyes rotate about axes that pass only through their first optical nodal points (e.g., Arditi, 1986; Bishop, 1985, Gulick & Lawson, 1974). None of these assumptions is strictly true. Thus, when one compares the Vieth-Mueller Circle to any empirically determined horopter there is a discrepancy between the theoretical and empirical horopters. With few exceptions (Deering, 1992) the eye geometry used for image calculations for stereoscopic virtual environments assume a visual model consistent with the Veith-Mueller circle. We will address the results of this assumption in section 5.0.

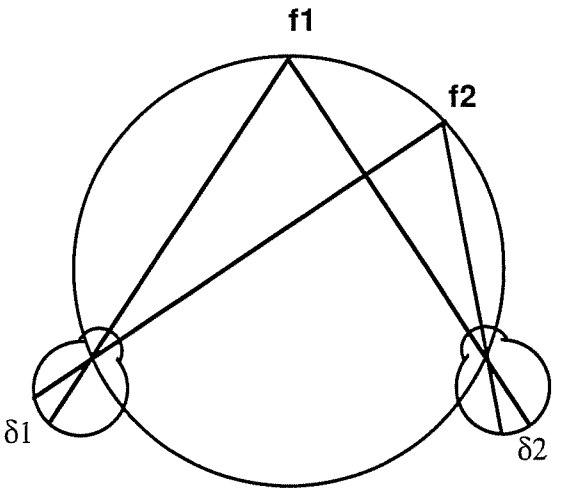

Figure 3. Vieth-Mueller Circle

Horopters usually describe a locus of points that should result in zero disparity. *Stereopsis*, however, occurs when there is a non-zero disparity that gives rise to the percept of depth. That is, an object or visual stimulus can appear closer or farther than the horopter for crossed and uncrossed disparity, respectively.

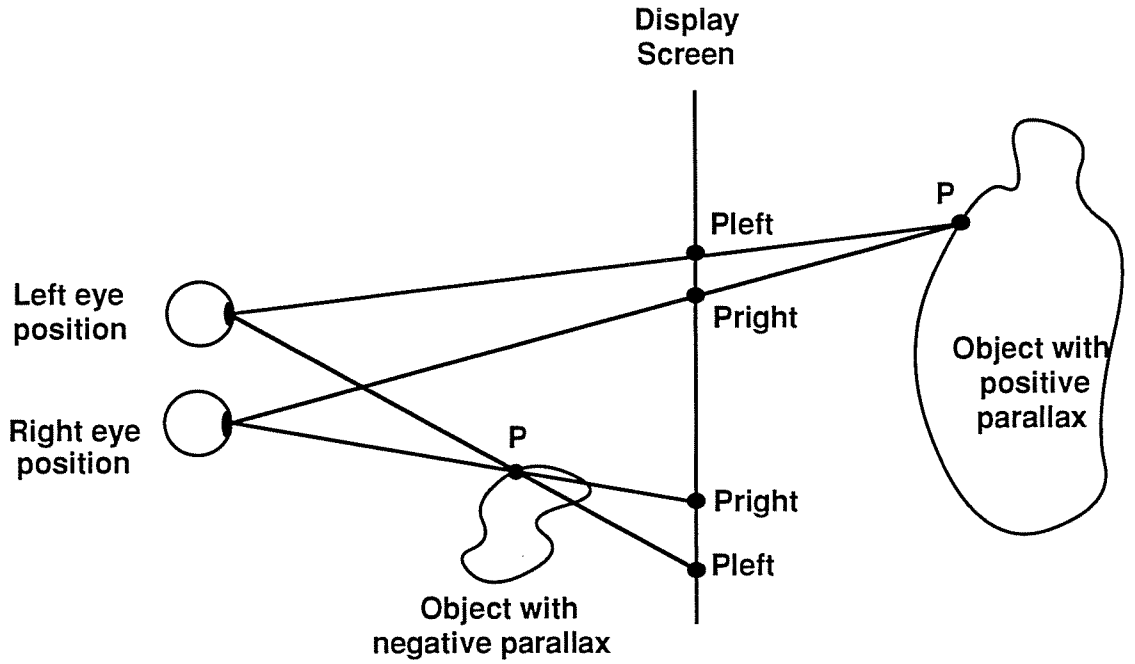

Figure 4. Screen Parallax

3.0 Modeling Geometry

For any type of stereoscopic display, we are modeling what would be seen by each eye if the image were projected onto a screen or window being viewed by an observer (Hodges, 1992). The simplest case, which is equivalent to the geometry of most head-mounted displays, occurs when the observer's eyes lie in a plane parallel to a single planar projection screen on which both the left- and right-eye projected views of an image are displayed (Figure 4). If we choose a point, P, on the object and follow its projection vectors to each eye position, then the *screen parallax* is the distance between the projected location of P on the screen, P_{left}, seen by the left eye and the projected location, P_{right}, seen by the right eye. Crossed or negative parallax (i.e., the left-eye view of P is to the right of the right-eye view of P) results in an image that appears spatially in front of the display screen. Uncrossed or positive parallax results in an image that appears behind the screen. Zero or no parallax results in a spatial position at the plane of the screen.

The amount of screen parallax in this case may be computed with respect to the geometric model of the scene as

$$p = i(D-d)/D$$

where p is the amount of screen parallax for a point, f1, when projected onto a plane a distance d from the plane containing two eyepoints. The modelled distance between eyepoints is i and the distance from f1 to the nearest point on the plane containing the two eyepoints is D (Figure 5).

Screen parallax does not correspond directly to retinal disparity. Retinal disparity is measured on the two retinae and its value is relative to the current binocular fixation point and convergence angle. Screen parallax is measured on the surface of a display device and its value is dependent on the computational model of the observer (i.e., distance of each eye position from the projection plane and the orientation of the plane relative to the plane containing the two eye positions) and the display geometry (how values are mapped to a display screen).

Points with constant screen parallax describe a straight line that is parallel to a horizontal plane through the two eye points. This assumes that the eyepoints do not change positions when the eyes converge and accommodate on a different point. In reality the location of the first nodal point of the eye's lens system will vary slightly with convergence and possibly also with accommodation. We will address this issue in section 5.0.

For now we assume a model in which perfectly spherical eyes rotate about the eyepoint so that the eyepoint is stationary with respect to the head position. For these assumptions the resulting screen parallax that is computed for each point in the scene will produce proper retinal parallax for any fixation point and resultant convergence angles. This is illustrated in Figure 6. Points f2 and f3 both have the same screen parallax as point f1. These equal screen parallaxes induce different convergence angles α, β and γ for points f1, f2, and f3, respectively. If the observer is fixated on f1, then the resultant retinal disparities for f2 and f3 are $(\alpha-\beta)$ and $(\alpha-\gamma)$, respectively.

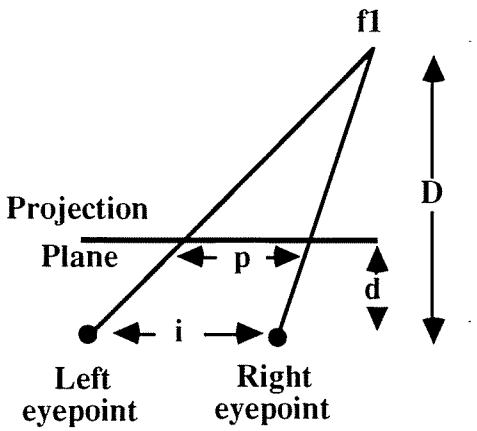

Figure 5. Screen parallax, p, is equal to i(D-d)/D

Screen parallax is an absolute measure of a point's distance from the plane of an observer's eyepoints. To calculate retinal disparity from screen parallax, however, we must look at values that are proportional to differences in screen parallax. For symmetric convergence, this calculation is straight-forward for points that lie on a line along a given binocular visual direction. If an observer's eyes are fixated on a point, f1, in a stereoscopic virtual environment, then there is a screen parallax p_α associated with that point and an angle of convergence, α. Another point, f2, with convergence angle, β, would also have an associated screen parallax, p_β. Remember that retinal disparity is the difference in vergence angles $(\alpha-\beta)$ between two points in space, measured in units of visual angle. For points that lie on a line along the binocular visual direction, the retinal disparity may be closely approximated directly from the screen parallax by the formula

$$2\,[\tan^{-1}(p_\alpha / 2(D-d)) - \tan^{-1}(p_\beta / 2(D-d))].$$

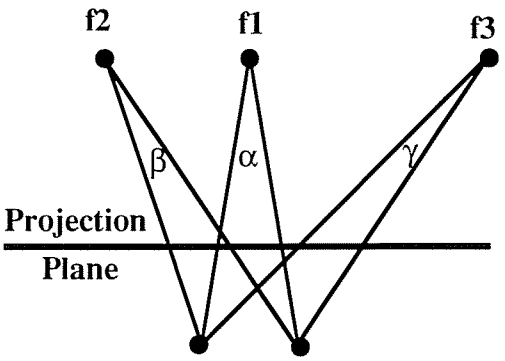

Figure 6. Screen parallax and convergence angles.

4.0 DISPLAY GEOMETRY

4.1 Stereoscopic Voxels

A stereoscopic display system provides a partitioning of three-dimensional space into volumetric spatial units which we shall refer to as *stereoscopic voxels*. Each stereoscopic voxel is defined by the intersection of the lines of sight from each eye through two distinct pixels. The size and shape of stereoscopic voxels are determined by the position of each eye as well as the pixel pitch and shape. If we assume an idealized

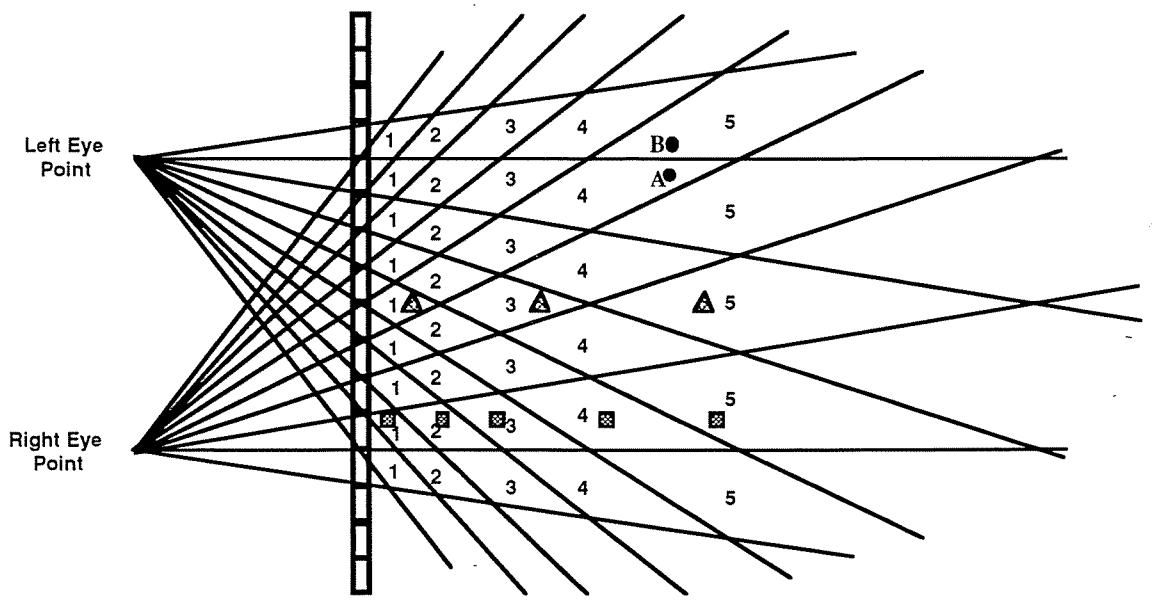

Figure 7. Cross-section of stereoscopic voxels.

rectangular pixel shape[3], then stereoscopic voxels can be modeled as six-sided polyhedrons with diamond-shaped horizontal cross section. A horizontal slice (created by one row of pixels) is shown in Figure 7. The distance, measured in number of pixels between the eye points, specifies the total number of discrete depth positions at which a point may be placed. In Figure 7, the number in each diamond cell indicates the number of pixels of parallax that defines that row of stereoscopic voxels. All stereoscopic voxels in a particular row have the same volume. They differ in shape in that each is a shear of the center (between the eyepoints) stereoscopic voxel.

A point's apparent position may be defined as the center of the cross-sectional area in which it resides.[4] The diamond cross-sectional shape of stereoscopic voxels means that a point's horizontal placement can affect its apparent depth location. In Figure 7, point A would appear to be closer to an observer than point B. Similarly the distinct steps in depth at which a point may be placed is affected by its horizontal placement. For example, the position of the triangular points in Figure 7 puts them only in three different depth positions (1, 3, and 5). Yet, the square points, which exist in the same range of depth, are positioned in five depth positions (1, 2,3, 4 and 5).

Conversely, if we allow some jitter in a pixel's depth location, then we can provide an effective doubling of horizontal spatial resolution for point positioning in stereoscopic display as compared to monoscopic display. For example, if we allow points to be positioned in either rows 0 and 1 of Figure 8, then we have fifteen horizontal positions represented across the front of the eight-pixel screen in Figure 8.

[3] When displayed on a CRT, pixels actually have a Gaussian shape and adjacent pixels may blend together. LCD pixels have well-defined edges, but most head-mounted displays incorporate an optical low-pass filter to reduce the perception of this effect. HMDs also incorporate optics that distort size and location of pixels.

[4] Alternately, we could think of the volume of a stereoscopic pixel as a measure of the *uncertainty* in the position of a point or as a type of spatial aliasing.

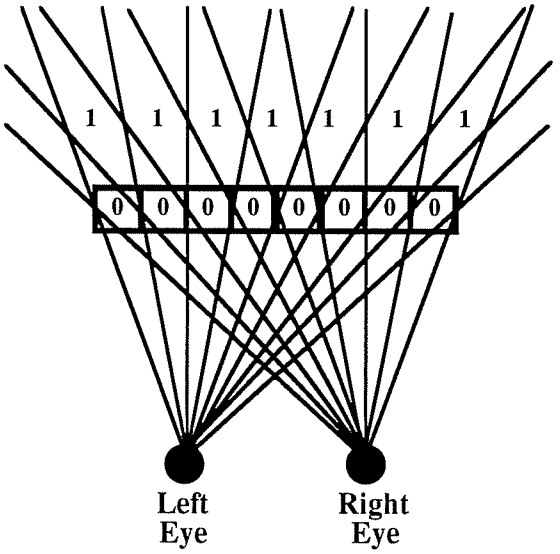

Figure 8. Jittering depth locations to achieve more horizontal resolution.

We can analytically characterize a stereoscopic voxel in row j by the values, d_{jnear}, d_{jfar}, d_{jmid} and d_{jwidth} (figure 9). The distance of nearest approach of a voxel j to the plane of the display screen is d_{jnear}, the distance from the plane to the most distance point of the voxel is d_{jfar}, the distance to the widest width of the voxel is d_{jmid}, and is the maximum width of the voxel is d_{jwidth}.

From similar triangles we can derive the following equations:

$d_{jnear} = d_{(j-2)far}$
$d_{jfar} = (j+1)pD/(\mathbf{i} - (j+1)p)$
$d_{jmid} = d_{(j-1)far}$
$d_{jwidth} = (j+1)p(d_{jfar} - d_{jmid})/ (d_{jfar})$

where p is the pixel pitch, the interocular distance is **i**, and the distance from the eyeplane to the projection plane is D.

Table 1 shows some representative values of these parameters assuming an interocular distance of 6.4 centimeters, a distance from the eyeplane to the plane of the screen of 40 centimeters, and a pixel width (through the optics) of 0.5 centimeters. These parameters are consistent with those of currently available head-mounted displays in the $5000-7,000 price range (Robinett & Rolland, 1992).

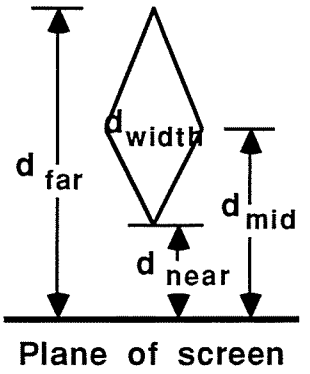

Figure 9. Characterization of stereoscopic voxel.

Table 1. Width and most distance point from projection surface of voxels in rows 0 through 11 in centimeters assuming interocular separation of 6.4 centimeters, projection surface at 40 centimeters, uncrossed parallax, and pixel width of 0.5 centimeters.

Row	dfar	dwidth
0	3.39	0.5
1	7.41	0.54
2	12.24	0.59
3	18.18	0.65
4	25.64	0.73
5	35.29	0.82
6	48.28	0.94
7	66.67	1.10
8	94.74	1.33
9	142.86	1.68
10	244.44	2.29
11	600.00	3.56

4.2 Geometric Distortion

Image distortion caused by display technology can be a problem for both time-

multiplexed and head-mounted displays. Two possible sources of distortion for time-multiplexed CRT displays have been analyzed by Deering (1992). The first source of distortion is that standard projection geometry assumes the display surface of a CRT is planar. Modern CRT screens, however, are shaped as a section of a sphere or a cylinder with a radius of curvature of about two meters. The second source of distortion is that the phosphor screen on a CRT is viewed through a thick glass faceplate with an index of refraction significantly higher than that of air. For a viewing angle of 60 ^{o} relative to a normal vector to the plane of the CRT, the curvature of the CRT screen can cause positional errors for objects at the surface of the display screen on the order of 1.8 centimeters. For the same viewing angle, the thickness of the glass faceplate can cause positional errors at the surface of display screen of up to 2.0 cm. These errors are both viewpoint dependent and nonlinear. Equations to compensate for these distortions are described by Deering (1992).

For head-mounted displays there are image distortions introduced by the optics of the system, by improper alignment of the user's eye position with respect to the mathematical model of eye position, and by mismatches between the display's field-of-view and the software's implicit mathematical model for field-of-view. Of particular interest are the distortions caused by the optical system in the head-mounted display. The purpose of the optical system is to provide an image to the user that is located at a comfortable distance for accommodation and magnified to provide a reasonable field-of view. The optics also cause nonlinear distortions in the image so that straight lines in the model appear curved in the visual image. Robinett and Rolland (1992) have developed a computational model for the geometry of head-mounted displays to correct for these distortions. Hodges & Watson (1993) have demonstrated a real-time implementation of this model.

4.3 Head-Tracking

Accurate head-tracking is crucial to the presentation of spatially consistent stereoscopic images. Screen parallax information is computed based on implied eye-point positions relative to head position and head orientation. If the observer's head is not exactly in the position reported by the head-tracker, then scaling and distortion effects in the image occur. If the observer's head is further away from the plane of the display than reported by the head-tracker, then the image is elongated in depth. If his head is closer, then the image is compressed in depth. If the observer's head is incorrectly tracked for side-to-side or up-down motion, the image is distorted by a shear transformation.

5.0 A Dose of Real Reality

From our analysis in section 3 we observed that screen parallax is dependent on the display geometry and the value chosen for interocular distance in the modeling geometry, but independent of the fixation point and of the visual direction. In that section we assumed that the eyes were perfect spheres with perfectly spherical optics and that each eye rotated about its first nodal point (the eyepoint for perspective projection). This model is implicit in all current implementations of stereoscopic virtual environments.

As indicated by the Hering-Hillenbrand deviation between an empirically determined horopter and the Vieth-Mueller Circle, however, these assumptions are only approximations of the true geometry of the eye. For example, the first nodal point does not actually correspond to the center of rotation of the eye (Deering, 1992, Ogle, 1968). As a result, if the eye rotates, then the first nodal point of each eye also shifts position depending on the current visual direction. The shifting creates a change in eyepoint separation. The greatest changes occur for symmetric convergence on points close to the observer. Figure 8 illustrates the range of change for an assumed separation of 6.4 centimeters between the centers of rotation of the eyes and a distance of 0.6

centimeters between the first nodal point and the center of rotation for each eye.

From section 2.2 we know that the angle of convergence, α, for a fixated point, f_1, is dependent not only on the distance to the point but also on eyepoint separation. If we correctly model the location of the eyepoint of a user and locate the center of rotation of each of his eyes, then we can compute the eye separation for any fixated point. Stereopsis is dependent on relative convergence angles specific to a particular fixated point. The result is that the screen parallax must be recomputed based on the current fixation point if we want to get retinal disparities in the three-dimensional virtual scene that are identical to retinal disparities created by a real image with the same geometry. For example, consider symmetric convergence on a fixated point thirty centimeters from the eyeplane, 6.4 centimeters interocular distance and 0.6 centimeters between the center of rotation of each eye and the first nodal point. We can compute the retinal disparity between that point and another point 100 centimeters along the same gaze direction as approximately 521 minutes of arc. If we now fixate on the point 100 centimeters away, the change in eye separation results in a retinal disparity of approximately 514 minutes of arc between the two points. This difference based on fixation point and gaze direction can not be modeled without an accurate description of the optical characteristics and physiology of the individual user's eye.

6.0 SUMMARY

Geometric considerations are critical in the visual display of stereoscopic virtual environments. The geometry of the two eyes, the geometry of the display system and the geometry implicit in the graphics model all must be carefully coordinated to create an accurate visual representation.

In particular, we have examined in detail the relationship of retinal disparity, fixation point, binocular visual direction, and screen parallax. We also have shown geometrically that a stereoscopic display partitions three-dimensional space into volumetric spatial units (stereoscopic voxels). Due to the shape of stereoscopic voxels, apparent depth of points in space may be affected by their horizontal placement.

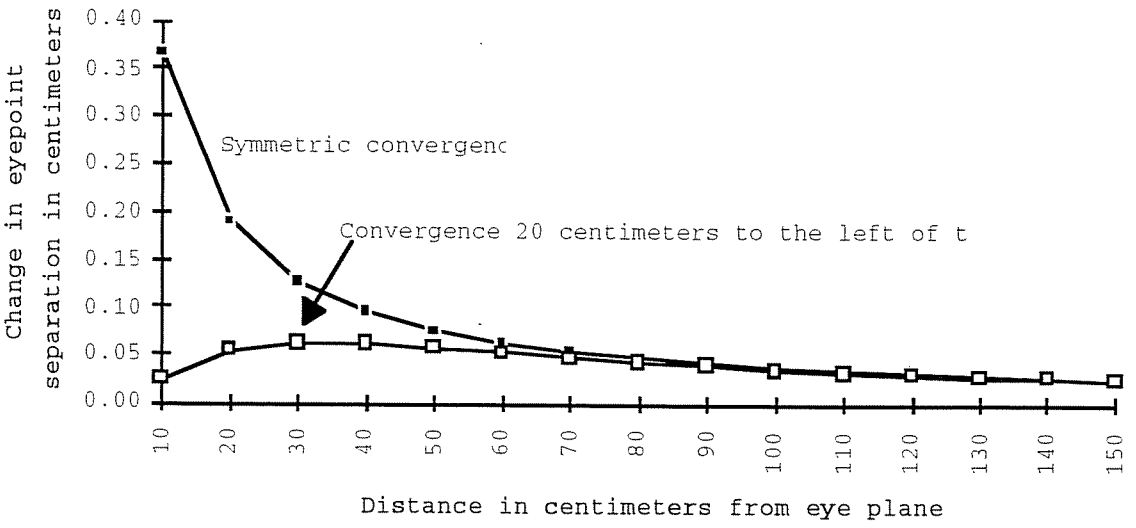

Figure 8. Change in eyepoint separation with change in point of fixation.
Centers of rotation of the eyes are assumed to be 6.4 centimeters apart.

7.0 REFERENCES

Arditi, A. (1986). Binocular vision. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.) *Handbook of Perception and Human Performance. Vol. 1 Sensory Processes and Perception.* New York: Wiley.

Bishop, P.O. (1985). Binocular vision. In *Adler's Physiology of the Eye.*

Bryson, S. (1992). Survey of Virtual Environment Technologies and Techniques. Siggraph '92 course notes for course #9: Implementation of Immersive Virtual Environments.

Cruz-Neira, Sandin, D.J., DeFanti, T.A., Kenyon, R.V., Hart, J.C. (1992). The CAVE: audio visual experience automatic virtual environment. *Communications ACM, 35*, 6, 64-72.

Davis, E.T. and Hodges, L.F. Human Stereopsis & Fusion and Stereoscopic Virtual Environments. To appear in Virtual Environments and Advanced Interface Design, W. Barfield and T. Furness, eds. Oxford University Press.

Davis, E.T.,King, R.A. and Anoskey, A. (1992). Oblique effect in stereopsis? *Human Vision, Visual Processing and Digital Display III*, Proc. SPIE 1666, 465-475.

Deering, M. (1992). High resolution virtual reality. *Computer Graphics* , 26, 2, 195-202.

Graham, C.H. (1965) Visual space perception. In C.H. Graham (Ed.), *Vision and Visual Perception.* New York: Wiley.

Gulick, W.L., and Lawson, R.B. (1974). *Human Stereopsis: A Psychophysical Analysis.* New York: Oxford.

Hodges, L.F. (1992). Time-multiplexed stereoscopic display. *IEEE Computer Graphics & Applications,* 12 , 2, 20-30.

McKee, S.P. (1983). The spatial requirements for fine stereoacuity. *Vision Research*, 23, 191-198.

Ogle, K.N. (1968) *Optics.* Charles C. Thomas, Publisher. Springfiled, Ill.

Robinett, Warren and Rolland, Jannick (1992). A computational model for stereoscopic optics of a head-mounted display. *Presence, 1*(1), 45-62.

Schor, C.M. (1987). Spatial factors limiting stereopsis and fusion. *Optic News, 13*(5),14-17.

Schor, C.M. and Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research* 23, 1649-1654.

Teitel, M.A. (1990). The eyephone: a head-mounted stereo display. *SPIE Proc. 1256*, 168-171.

Tyler, C.W. (1983). Sensory Processing of Binocular Disparity. In C.M. Schor and K.J. Ciuffreda (Eds.) *Vergence Eye Movements: Basic and Clinical Aspects.* Boston: Butterworths.

Watson, B.A. and Hodges, L.F. (1993) Real-time compensation for optical distortion in head-mounted displays. Technical report GIT-GVU-93-10, Georgia Tech Graphics, Visualization & Usability Center.

Westheimer, G. and McKee, S.P. (1980). Stereo design for testing local stereopsis. *Investigative Ophthalmology and Visual Science*, 19, 802-809.