Published in final edited form as:
ACM Trans Graph. 2008 January 1; 2008: 23–32. doi:10.1145/1394281.1394285.

Misperceptions in Stereoscopic Displays: A Vision Science Perspective

Robert T. Held[1,2] and **Martin S. Banks**[1]

[1] University of California, Berkeley

[2] University of California, San Francisco

Abstract

3d shape and scene layout are often misperceived when viewing stereoscopic displays. For example, viewing from the wrong distance alters an object's perceived size and shape. It is crucial to understand the causes of such misperceptions so one can determine the best approaches for minimizing them. The standard model of misperception is geometric. The retinal images are calculated by projecting from the stereo images to the viewer's eyes. Rays are back-projected from corresponding retinal-image points into space and the ray intersections are determined. The intersections yield the coordinates of the predicted percept. We develop the mathematics of this model. In many cases its predictions are close to what viewers perceive. There are three important cases, however, in which the model fails: 1) when the viewer's head is rotated about a vertical axis relative to the stereo display (yaw rotation); 2) when the head is rotated about a forward axis (roll rotation); 3) when there is a mismatch between the camera convergence and the way in which the stereo images are displayed. In these cases, most rays from corresponding retinal-image points do not intersect, so the standard model cannot provide an estimate for the 3d percept. Nonetheless, viewers in these situations have coherent 3d percepts, so the visual system must use another method to estimate 3d structure. We show that the non-intersecting rays generate vertical disparities in the retinal images that do not arise otherwise. Findings in vision science show that such disparities are crucial signals in the visual system's interpretation of stereo images. We show that a model that incorporates vertical disparities predicts the percepts associated with improper viewing of stereoscopic displays. Improving the model of misperceptions will aid the design and presentation of 3d displays.

Keywords

Depth perception; Virtual Reality; 3D displays; Visualization

1 Introduction

Stereoscopic displays have become commonplace as they have made their way into cinema [Lipton 1982], medical imaging [Chan et al. 2005], scientific visualization [Fröhlich et al. 1999], and other applications. As the use of such displays has spread, the benefits and problems associated with them have become clearer. A well-documented problem is that perceived 3d shape and scene layout is often distorted. For instance, viewing a stereoscopic display from the wrong distance typically alters the perceived size and shape of an object [Masaoka et al. 2006; Woods et al. 1993]. In some applications, such as cinema, the distortions are not necessarily a serious problem [Lipton 1982], but in other applications, such as medical imaging or virtual reality [Deering 1992], they can have grave consequences. Here we examine the causes of perceptual distortions in stereography, show that in some cases the standard model of such distortions is incorrect, and describe how the model should be modified.

There are three steps in the production of a stereoscopic percept. (1) Images are acquired by stereo photography or generated by computer graphics. (2) The images are presented stereoscopically to a viewer. (3) The images are interpreted by the viewer's visual system. Geometric misperceptions are caused by inappropriate acquisition-viewing relationships (steps 1 & 2) such that the retinal images are not the same as those produced by the original scene. Perceptual misperceptions are produced by the viewer's visual system (step 3): the retinal images may each be geometrically correct, but visual cues such as vergence and accommodation cause them to be misinterpreted nonetheless. The graphics, stereocinema, and virtual-reality literatures [Diner 1991; Jones et al. 2001; Kusaka 1992; Kutka 1994; Leiser et al. 1995; Lipton 1982; Masaoka et al 2006; Son et al. 2002; Wartell et al. 2002; Woods et al. 1993; Yamanoue et al. 2006] have only used the geometric approach.

We begin by developing the mathematics of the geometric approach and summarizing the predicted distortions. We then describe the limitations of the approach, especially in dealing with the vertical disparities produced by some viewing situations; these situations include rotation of the viewer's head relative to the display and using converging cameras to acquire images that are then viewed on a single-surface display. Finally, we describe an approach derived from vision science that provides a better characterization.

2 Related Work: Geometric Approach

To describe the geometric approach, we use some derivations from Woods et al. [1993].

Step 1: Acquisition (*Object space to 2d camera sensors*)

We first determine the 2d coordinates of a point in 3d space (*P*) once projected onto the sensors of a pair of cameras. In 3d coordinates, *X* is the inter-camera axis, *Y* is the vertical axis perpendicular to the camera axis and running through the midpoint between the cameras, and *Z* is orthogonal to *X* and *Y*. The 3d coordinates of *P* are $\boldsymbol{P_o}$. The cameras are specified by focal length *f*, sensor width W_c, and inter-camera separation *t*. Each camera has two axes: the *lens axis*, which bisects and is normal to the lens and is perpendicular to the image sensor plane, and the *optical axis*, which contains the lens center and sensor center. Camera alignment is specified by V_c, the angle between the cameras' optical axes, and by *h*, the displacement of the lens axis in the sensor plane. With the camera lenses parallel to one another, the optical axes can be parallel ($h = 0$, $V_c = 0$) or converging ($h \neq 0$, $V_c \neq 0$) (Figure 1A and 1B). *P*'s coordinates in the left and right cameras are (x_{cl}, y_{cl}) and (x_{cr}, y_{cr}), where *x* and *y* are horizontal and vertical coordinates in the sensors:

$$\begin{aligned} x_{cl} &= f\tan\left[\tan^{-1}\left(\frac{t/2+P_o(x)}{P_o(z)}\right)-\frac{V_c}{2}\right]-h \\ x_{cr} &= f\tan\left[\frac{V_c}{2}-\tan^{-1}\left(\frac{t/2-P_o(x)}{P_o(z)}\right)\right]+h \\ y_{cl} &= \frac{P_o(y)f}{P_o(z)\cos(\frac{V_C}{2})+\left(P_o(x)+\frac{t}{2}\right)\sin(\frac{V_C}{2})} \\ y_{cr} &= \frac{P_o(y)f}{P_o(z)\cos(\frac{V_C}{2})-\left(P_o(x)-\frac{t}{2}\right)\sin(\frac{V_C}{2})} \end{aligned} \qquad (1)$$

Step 2: Presentation (*2d camera sensors to 2d projections*)

To present the stereo camera images, the sensor coordinates (x_{cl}, y_{cl}) and (x_{cr}, y_{cr}) must be transformed to 2d picture coordinates (X_{sl}, Y_{sl}) and (X_{sr}, Y_{sr}). In most applications, the pictures are presented on one display surface such as an LCD or projection screen. In vision science, they are often presented on two displays, one for each eye, in a device called a haploscope [Backus et al. 1999]. Single-surface displays are much more common, so we concentrate on them here. The pictures are characterized by their width W_p, and d, which is their horizontal displacement relative to one another. The ratio W_p/W_c is the magnification from the camera images to the picture. The 2d coordinates of corresponding points in the picture are:

$$X_{sl}=x_{cl}\left(\frac{W_p}{W_c}\right)-\frac{d}{2} \quad Y_{sl}=y_{cl}\left(\frac{W_p}{W_c}\right)$$
$$X_{sr}=x_{cr}\left(\frac{W_p}{W_c}\right)+\frac{d}{2} \quad Y_{sr}=y_{cr}\left(\frac{W_p}{W_c}\right)$$

Step 3: Viewing (*2d projections to percept*)

The binocular viewer is positioned to view the pictures on the display surface. We use two new sets of 3d coordinates to describe this: one with its origin on the display surface and one with its origin at the viewer. For the first set, X and Y are the horizontal and vertical axes centered on the display surface and Z is orthogonal to them. In these coordinates, the eyes' positions are $\boldsymbol{E_l}$ and $\boldsymbol{E_r}$. The positions of the points in picture are:

$$\boldsymbol{P_l}=(X_{sl},Y_{sl},0) \quad \boldsymbol{P}_r=(X_{sr},Y_{sr},0)$$

To determine the viewer's estimate of the location of P given $\boldsymbol{E_l}$, $\boldsymbol{E_r}$, $\boldsymbol{P_l}$, and $\boldsymbol{P_r}$, we project rays from the eye centers through the corresponding points in the picture. The estimated location of P is assigned to the point of intersection $\boldsymbol{P_i}$ (Figure 1C). We want the location of $\boldsymbol{P_i}$ specified in viewer coordinates, so we transform $\boldsymbol{E_l}$, $\boldsymbol{E_r}$, $\boldsymbol{P_l}$, and $\boldsymbol{P_r}$ into $\boldsymbol{E'_l}$, $\boldsymbol{E'_r}$, $\boldsymbol{P'_l}$, and $\boldsymbol{P'_r}$ in a viewer-centered system. There, the origin is midway between the eyes, which is $\boldsymbol{E_c}$ in picture-centered coordinates; X is the inter-ocular axis, Y is the vertical axis, and Z is orthogonal to them. $\boldsymbol{E_c}$ is subtracted from $\boldsymbol{E_l}$, $\boldsymbol{E_r}$, $\boldsymbol{P_l}$, and $\boldsymbol{P_r}$. The transformations are listed below:

$$\boldsymbol{E}_{\boldsymbol{l}}'=R(\rho,\sigma)(\boldsymbol{E_l}-\boldsymbol{E_c}) \quad \boldsymbol{E}_{\boldsymbol{r}}'=R(\rho,\sigma)(\boldsymbol{E_r}-\boldsymbol{E_c})$$
$$\boldsymbol{P}_{\boldsymbol{l}}'=R(\rho,\sigma)(\boldsymbol{P_l}-\boldsymbol{E_c}) \quad \boldsymbol{P}_{\boldsymbol{r}}'=R(\rho,\sigma)(\boldsymbol{P_r}-\boldsymbol{E_c})$$
$$\text{where}\, R(\rho,\sigma)=\begin{bmatrix} \cos(\rho)\cos(\sigma) & \cos(\rho)\sin(\sigma) & -\sin(\rho) \\ -\sin(\sigma) & \cos(\sigma) & 0 \\ \sin(\rho)\cos(\sigma) & \sin(\rho)\sin(\sigma) & \cos(\rho) \end{bmatrix}$$
$$\text{and}\, \rho=-\tan^{-1}\left(\frac{E_r(z)-E_l(z)}{E_r(x)-E_l(x)}\right) \,\text{and}\, \sigma=\sin^{-1}\left(\frac{E_r(y)-E_l(y)}{\left|\left|\boldsymbol{E_r}-\boldsymbol{E}_l\right|\right|}\right)$$

The intersection of rays originating at $\boldsymbol{E'_l}$ and $\boldsymbol{E'_r}$ and passing through $\boldsymbol{P'_l}$, and $\boldsymbol{P'_r}$ can then be found from:

$$\boldsymbol{E}_{\boldsymbol{l}}'+(\boldsymbol{P}_{\boldsymbol{l}}'-\boldsymbol{E}_{\boldsymbol{l}}')m=\boldsymbol{E}_r'+(\boldsymbol{P}_r'-\boldsymbol{E}_r')n$$

$(\boldsymbol{P'_l}-\boldsymbol{E'_l})m$ and $(\boldsymbol{P'_r}-\boldsymbol{E'_r})n$ represent the exiting rays; m and n are used to define points along those rays. When the two sides of the equation are set equal to each other, one can find the rays' intersection. The solutions for m and n are:

$$m=\frac{|(\boldsymbol{E}_{\boldsymbol{r}}^{\prime}-\boldsymbol{E}_{\boldsymbol{l}}^{\prime})\times(\boldsymbol{P}_{\boldsymbol{r}}^{\prime}-\boldsymbol{E}_{\boldsymbol{r}}^{\prime})|}{|(\boldsymbol{P}_{\boldsymbol{l}}^{\prime}-\boldsymbol{E}_{\boldsymbol{l}}^{\prime})\times(\boldsymbol{P}_{\boldsymbol{r}}^{\prime}-\boldsymbol{E}_{r}^{\prime})|}\quad n=\frac{|(\boldsymbol{E}_{\boldsymbol{l}}^{\prime}-\boldsymbol{E}_{\boldsymbol{r}}^{\prime})\times(\boldsymbol{P}_{\boldsymbol{l}}^{\prime}-\boldsymbol{E}_{\boldsymbol{l}}^{\prime})|}{|(\boldsymbol{P}_{\boldsymbol{r}}^{\prime}-\boldsymbol{E}_{\boldsymbol{r}}^{\prime})\times(\boldsymbol{P}_{\boldsymbol{l}}^{\prime}-\boldsymbol{E}_{\boldsymbol{l}}^{\prime})|}$$

From this, we obtain:

$$\boldsymbol{P}_{\boldsymbol{i}}^{\prime}=\boldsymbol{E}_{\boldsymbol{l}}^{\prime}+(\boldsymbol{P}_{\boldsymbol{l}}^{\prime}-\boldsymbol{E}_{\boldsymbol{l}}^{\prime})m\quad\text{or}\quad\boldsymbol{P}_{\boldsymbol{i}}^{\prime}=\boldsymbol{E}_{\boldsymbol{r}}^{\prime}+(\boldsymbol{P}_{\boldsymbol{r}}^{\prime}-\boldsymbol{E}_{r}^{\prime})n$$

These terms are identical if the intersection exists. We discuss non-intersecting rays in Section 4. We now have the estimated location of the point P in viewer coordinates. Misperceptions can be quantified by differences between $\boldsymbol{P'_i}$ and $\boldsymbol{P_o}$.

Before examining the consequences of incorrect acquisition and viewing parameters, it is useful to consider what it means to have those parameters correct. The picture presented to each eye has a center of projection (COP) whose position depends on image magnification (W_p/W_c) and the orientation of the camera's optical axis relative to the sensor plane (h and V_c). The separation between the COPs depends on inter-camera separation t, magnification W_p/W_c, and picture offset d. Two constraints must be satisfied for the viewing situation to match the viewing of the original scene. 1) Both eyes must be positioned at the appropriate COPs [Leiser et al. 1995; Wartell 2002]. When the eyes are so positioned, the retinal images are the same while viewing the stereo picture as they would be while viewing the original scene. 2) The eyes' vergence (the angle between the eyes' optical axes) required to fixate a point in the virtual scene must be the same as the vergence required to fixate the corresponding point in the original scene [Leiser et al. 1995]. We will refer to viewing situations in which these constraints are satisfied as the *proper viewing condition*.

We are most interested in what happens when the viewing condition is not proper: specifically, when one or both eyes are not at the appropriate COPs and/or when the eye vergence is inappropriate. Incorrect positioning and vergence are common with single viewers and necessarily occur with multiple viewers.

3 Predicted Distortions: Geometric Approach

We implemented the geometric approach in software and investigated the consequences of modifying acquisition and viewing parameters. The investigation revealed viewing situations in which the geometric approach fails to produce a solution; we discuss those in Section 4.

Figure 1 in the color plate shows the results of one investigation. We presented a 30cm cube at a distance of 55cm from the cameras. In the proper viewing condition, the following parameters were used:

Acquisition Parameters:

Orientation of camera optical axes: Parallel

Inter-camera separation (t): 6.2cm

Camera focal length (f): 6.5mm

Viewing Parameters:

Magnification (W_p/W_c): 84.6

Picture separation (d): 6.2cm

Viewing distance: 55cm

Inter-ocular distance (I): 6.2cm

Viewer position: Midpoint of inter-ocular axis on central surface normal of display.

Viewer orientation: Face parallel to display surface

We modified parameters independently to observe their effects on the estimated 3d percept. Each panel of Color Plate Figure 1 presents the results for a set of parameters; E is the proper viewing condition.

B and H show the consequences of moving the viewer respectively farther from (110cm) and closer to (27.5cm) the picture. When the viewer is too distant, the predicted perceived distance is greater and the predicted shape is stretched in depth. When the viewer is too close, the predicted perceived distance is less and the predicted shape is compressed in depth. These results are consistent with the analysis of Woods et al. [1993]. In D and F, we translated the viewer left and right of the proper viewing position. The translation was parallel to the display surface and the viewer's head remained parallel to the surface. The predicted shape is skewed toward the viewer. A and I show the effects of inter-camera separation. The proper separation was equal to the inter-ocular distance of 6.2cm. In A, the cameras are 3.1cm apart, and the predicted stimulus is larger and farther away. In I, the cameras are 12.4cm apart, and the predicted stimulus is smaller and closer to the viewer than the original stimulus. C and G show the effects of picture displacement. In the proper viewing condition, the centers of the pictures were 6.2cm apart. Changing the picture displacement increases or decreases all of the disparities in the retinal images. When the pictures are separated by 7.5cm, the disparities are increased, and the result is a predicted stimulus that is farther away and stretched in depth. When the separation is 3.1cm, the predicted stimulus is closer and the shape is compressed in depth. Woods et al. [1993] did not investigate picture displacement, but this result could be derived from their analysis.

The results in Color Plate Figure 1 are consistent with our empirical observations in these viewing situations and generally with the analysis of Woods et al [1993]. There are, however, acquisition-viewing conditions for which the geometric approach does not yield a solution; we now turn to them.

4 Failures of the Geometric Approach

In many viewing situations, rays from the eyes through corresponding points in the stereo pictures do not intersect, so the geometric approach cannot yield a solution for the predicted perceived stimulus. Interestingly, viewers in those situations perceive a coherent 3d scene, so the visual system finds a solution nevertheless. The presence of non-intersecting (skew) rays has been mostly unnoticed in the literature and the perceptual consequences have never been investigated. Our main contribution is an analysis of the causes of skew rays and a description of the manner in which the visual system finds a 3d solution when such rays exist.

4.1 Epipolar Planes

The causes of skew rays can be well understood in epipolar geometry. A point in real space and the two eye centers define a plane: the *epipolar plane* [Shapiro and Stockman 2001]. It can be shown that two corresponding points in a stereo picture produce intersecting rays as long as they lie in the same epipolar plane (and are non-parallel). Consider a viewer of a stereoscopic picture with the eyes ($\boldsymbol{E'_l}$ and $\boldsymbol{E'_r}$) positioned at the COPs (Figure 2B). A ray from the left eye to point $\boldsymbol{P'_l}$ and a ray from the right eye to $\boldsymbol{P'_r}$ are identical to the rays that would have passed from the eyes to the original point P. Thus, they lie in the same epipolar plane as P and will intersect at P in virtual space. Now consider viewing the stereo picture with the eyes not at the COPs. The line segment between the COPs is the *inter-COP axis*. If the viewer is translated from the proper position, the inter-ocular axis will be parallel to, but not coincident with, the inter-COP axis. Rays from the two eyes to the corresponding points in the picture still lie in a common epipolar plane, so they will intersect in space. Therefore, a geometric solution exists for $\boldsymbol{P'_I}$. This is why the geometric approach could produce solutions to the viewing situations in Section 3. If the viewer's head is rotated about the inter-ocular axis (defined as "pitch"), the two axes remain coincident, so a solution still exists. But if the viewer's head is rotated about a vertical axis (yaw rotation) or a forward axis (roll rotation), the inter-ocular and inter-COP axes will be neither coincident nor parallel. In those cases, there are corresponding points in the picture that produce rays in different epipolar planes (Figure 2C). The rays are therefore not guaranteed to intersect, so there may be no solution for $\boldsymbol{P'_i}$. The Appendix provides a mathematical derivation of these results.

Mismatches between camera setup and display surface can also produce skew rays. For instance, the imaging sensors of converging cameras (lens axes converging) lie in different planes, but the resulting stereo pictures are usually displayed on one plane. The mismatch causes the left and right stereo pictures to exhibit "keystone" distortion (Color Plate Figure 2), which creates non-zero on-screen vertical disparities between points that have non-zero Y coordinates in the picture. The vertical disparities produce rays that lie in different epipolar planes, so they do not provide a solution. A modification of the geometric approach provides a solution [Woods et al. 1993], but as we will show, the solution is very unlikely to match viewers' percepts.

4.2 Previous Solutions to Skew Rays

Most previous investigations of misperceptions in stereography have not discussed skew rays [Diner 1991; Jones et al. 2001; Kusaka 1992; Kutka 1994; Leiser et al. 1995; Masaoka et al. 2006, Strunk and Iwamoto 1990; Yamanoue et al. 2006], but a few have noted their existence in some situations [Agrawala et al 1997; Wartell et al 2002; Woods et al 1993]. Only one of those considered the possible perceptual consequences: Woods et al. [1993] modified the geometric approach to accommodate skew rays created by improper acquisition and viewing settings. In particular, they observed that using converging cameras and a single display surface causes vertical disparities in the picture surface; they did not consider viewer rotations as we have in Section 4.1. In modifying the geometric approach, Woods and colleagues first determined which pairs of corresponding points had unequal on-screen Y coordinates (Y_{sl} and Y_{sr}; Equations 1). They then reset the Y coordinates for each pair to the average Y value. This placed the on-screen points and the eye centers in a common epipolar plane, so ray intersections could be found. There are two important shortcomings with this approach. 1) It does not apply to yaw and roll rotations even though such rotations create skew rays (Figure 2). 2) Even in the converging camera situation for which the approach does apply, vision-science findings strongly suggest that the 3d estimate will not match human percepts. We describe these findings next.

4.3 Vertical Disparity

A point in a real scene projects in the same epipolar plane for both eyes, but as we have said, epipolar geometry does not necessarily hold in the viewing of stereo pictures. Consequently, corresponding points in a stereo picture may project to different elevations in the two eyes, thereby creating non-zero vertical disparities in an epipolar coordinate system. Such non-zero vertical disparities are known to influence 3d percepts. An example is the *induced effect*. A lens is placed before one eye that magnifies the image vertically and creates non-zero vertical disparities. When this is done, a frontoparallel surface appears slanted even though the horizontal disparities created by the surface are unaffected by the magnifier [Ogle 1932]. There are many other perceptual consequences of altering vertical disparity, so it is well accepted in vision science that 3d percepts are a product of horizontal and vertical disparities [Backus et al. 1999; Banks et al., 2001; Rogers and Bradshaw 1993; Rogers and Bradshaw 1995]. Indeed, the visual system uses a number of depth cues to estimate the 3d structure of the environment. Many are monocular cues such as perspective and shading, which are beyond the scope of our discussion. But two estimation methods are based on stereopsis and should therefore be considered here.

One stereoscopic estimation method is based on measuring horizontal disparities and eye position [Backus et al., 1999]:

$$S \approx -\tan^{-1}\left(\frac{1}{\mu}\ln(HSR) - \tan(\gamma)\right) \quad (2)$$

where *S* is the slant of a surface patch, *HSR* is the horizontal size ratio (a measure of horizontal disparity; defined in Figure 3), μ is the eyes' vergence (defined in the figure), and γ is the eyes' version (defined in the figure). The geometric approach discussed here is identical to this means of estimating surface orientation.

Another stereoscopic method is based on measuring horizontal and vertical disparity and does not require an estimate of eye position [Backus et al., 1999]:

$$S \approx -\tan^{-1}\left(\frac{1}{\tilde{\mu}}\ln\left(\frac{HSR}{VSR}\right)\right) \quad (3)$$

where *VSR* is the vertical size ratio (a measure of vertical disparity; defined in Figure 3B), and $\tilde{\mu}$ is a measure of vergence derived from the gradient of *VSR*.[1] The visual system uses both of these stereoscopic methods to estimate surface orientation from binocular disparity [Backus et al. 1999; Garding et al. 1995; Rogers and Bradshaw 1995]. When the two methods provide different estimates, the system's final estimate is a weighted average of the two with the weights determined by the relative reliabilities of the two methods [Backus et al. 1999; Rogers and Bradshaw, 1995].

4.4 Perception with Skew Rays Present

As we showed, three situations produce skew rays and the standard model cannot predict the resulting percepts. What then is the visual system doing in these situations? We consider this by examining the three situations in which skew rays occur.

Condition 1: Observer Rotation in *X–Z* Plane (Yaw)—Color Plate Figure 3 is a stereo picture of a cube. To see the perceptual consequences of a yaw rotation, rotate the picture about a vertical axis. The 3d percept changes in a few ways: the front and back surfaces appear to rotate relative to the viewer such that they remain roughly parallel to the picture surface; the front surface appears to rotate slightly less than the back surface, so the surfaces become non-parallel; the distance between the front and back surfaces appears to decrease. To understand the perceptual consequences, we need to consider the horizontal and vertical disparities created by the cube following a yaw rotation. Figure 4 plots those disparities as vectors; panels B and E show the disparities associated with the cube's front and back surfaces, respectively. There are regions in the stimulus in which the vertical disparities reverse sign from the front to back surface; the upper left corner is an example.

[1] These equations apply for tilt 0. Extensions have been derived for all tilts [Banks et al. 2001].

Such a sign reversal can never occur in natural viewing with aligned eyes.[2] For this reason, we cannot appeal to a natural situation to determine what the visual system perceives when a stereo picture undergoes a yaw rotation. The answer, however, is suggested by the vision science literature. Duke and Howard [2005] created stereograms of two transparent planes, one in front of the other. They applied one pattern of vertical disparity to one plane and the opposite pattern to the other plane; this creates reversals in the sign of vertical disparity (as we observed with yaw rotations). Viewers of these unnatural stimuli perceived different surface shapes for the front and back surface and those shapes are well predicted by a weighted combination of Equations 2 and 3, applied separately to the two surfaces. We found that the percept associated with yaw rotation while viewing stereo pictures is well predicted by a similar weighted combination of surface orientation estimates derived from the two means of estimation.

Condition 2: Observer Rotation in *X–Y* Plane (Roll)—To see the consequences of a roll rotation, rotate the upper picture (Color Plate Figure 3) about the forward axis. The 3d percept changes little with small rotations and then collapses with larger rotations as the visual system becomes unable to fuse the disparate images. As shown in Section 4.1, roll rotations cause non-intersecting rays, so once again the geometric approach cannot derive an estimate for the perceived 3d structure. To understand the perceptual effects, we again consider the horizontal and vertical disparities created by this viewing situation. Panels A and D in Figure 4 plots the disparities associated with the cube's front and back surfaces. The disparity pattern can be understood by considering how the eyes' positions change with head roll. If the roll is counterclockwise, the right eye moves up and becomes closer to the upper right corner and farther from the bottom right corner. The opposite is true for the left eye. As a result, the upper right corner creates a larger retinal image in the right than in the left eye. The opposite is true for the bottom right corner. In both corners, the vertical disparities in epipolar coordinates have changed from zero with no roll to non-zero after roll. The horizontal disparities have been altered as well. The changes in vertical and horizontal disparity are proportional to the distance of the point from the rotation axis. Because of this, the perceived shape of the front or back surface should become curved, one corner bending toward the viewer and the opposite bending away. The amount of curvature depends on the magnitude of roll and whether the points on the picture have crossed or uncrossed horizontal disparity. The predicted deviation is only significant with large rolls, so one expects little if any perceptual change for small rolls. We have been unable to observe the curvature effect because the ability to fuse the stimulus breaks down at the larger rolls where the effect is predicted.

Condition 3: Converging Cameras—Color Plate Figure 3 demonstrates the perceptual consequences of using converging cameras but a single display surface for viewing. The top panel is a stereo picture of a cube when the cameras were parallel and the bottom panel is a picture of the same cube when the cameras were converging. The 3d percepts for the two cases differ in two ways: the cube's front and back surfaces appear planar in the parallel-camera case and convex in the converging-camera case; the front and back surfaces appear closer to the viewer in the converging case. To understand the perceptual consequences of using converging cameras, we need to again consider the horizontal and vertical disparities. Using converging cameras causes keystoning (Color Plate Figure 2), thereby changing the pattern of vertical disparities. In particular, the horizontal gradient of vertical disparity is altered such that it specifies a nearer surface than is actually present; the alteration is

[2] The vertical disparity associated with a point in space is nonzero if the point is to the left or right of straight ahead (i.e., not in the head's mid-sagittal plane) and above or below the plane of fixation (i.e., not in the visual plane). For any combination of such azimuth and elevation the vertical disparities of points at all distances have the same sign.

different for near and far surfaces, but it always increases the vertical-disparity gradient. Converging cameras also alter horizontal disparities. Specifically, the horizontal gradient of horizontal disparity specifies a more convex surface than is actually present; again the change differs for near and far surfaces, but always increases the gradient. From Duke and Howard [2005], we know that the visual system is likely to estimate 3d structure by estimating the orientation and curvature of surfaces separately with a weighted combination of Equations 2 and 3. We found that this model predicts the percept associated with converging cameras quite well.

5 Conclusions and Future Work

It is important to understand the misperceptions that occur when viewing stereoscopic pictures; without such an understanding, it will be very difficult to create displays and viewing situations for which 3d percepts are faithful to the intended result. With that in mind, we evaluated the standard model for predicting 3d percepts from stereoscopic displays, particularly when the acquisition and viewing parameters are improper as necessarily occurs with multiple viewers. The standard model makes reasonable predictions in many situations, but fails to make predictions in some important ones that are known to produce misperceptions. Those situations involve rotation of the viewer's head relative to the display and the use of converging cameras in acquisition with single displays for viewing. The skew rays that occur in those situations give rise to vertical disparities in the retinal images that were not present before the viewer rotation or before converging cameras were used. We described findings in the vision-science literature that point to how the visual system determines 3d structure in these situations. In particular, the system uses vertical disparity as an additional signal for determining the structure. Preliminary observations are consistent with the predictions derived from this model.

Our analysis, guided by findings in vision science, can aid the design and evaluation of stereoscopic displays and viewing parameters in applications ranging from cinema to medical imaging. We hope the work will spur further collaboration between the vision-science and computer-graphics communities.

References

Agrawala, M.; Beers, AC.; McDowall, I.; Fröhlich, B.; Bolas, M.; Hanrahan, P. The two-user responsive workbench: Support for collaboration through individual views of a shared space. Proceedings of ACM SIGGRAPH 97; New York: ACM Press/Addison-Wesley; 1997. Computer Graphics Proceedings, Annual Conference Series; ACM; p. 327-332.

Backus BT, Banks MS, van Ee R, Crowell JA. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. Vision Research. 1999; 39(6):1143–1170. [PubMed: 10343832]

Banks MS, Hooge IT, Backus BT. Perceiving slant about a horizontal axis from stereopsis. Journal of Vision. 2001; 1(2):55–79. [PubMed: 12678602]

Chan HP, Goodsitt MM, Helvie MA, Hadjiiski LM, Lydick JT, Roubidoux MA, Bailey JE, Nees A, Blane CE, Sahiner B. ROC study of the effect of stereoscopic imaging on assessment of breast lesions. Medical Physics. 2005; 32(4):1001–1009. [PubMed: 15895583]

Deering MF. High Resolution Virtual Reality. SIGGRAPH Computer Graphics. 1992; 26(2):195–202.

Diner, DB. A new definition of orthostereopsis for 3d television. Proceedings of the 1991 IEEE International Conference on Systems, Man, and Cybernetics, Decision Aiding for Complex Systems; 1991. p. 1053-1058.

Duke PA, Howard IP. Vertical-disparity gradients are processed independently in different depth planes. Vision Research. 2005; 45(15):2025–2035. [PubMed: 15820519]

Fröhlich, B.; Barrass, S.; Zehner, B.; Plate, J.; Göbel, M. Exploring geo-scientific data in virtual environments. Proceedings of the Conference on Visualization '99: celebrating ten years; Los Alamitos, CA: IEEE Computer Society; 1999. p. 169-173.

Gårding J, Porrill J, Mayhew JEW, Frisby JP. Stereopsis, vertical disparity and relief transformations. Vision Research. 1995; 35(5):703–722. [PubMed: 7900308]

Jones, G.; Lee, D.; Holliman, N.; Ezra, D. In: Woods, AJ.; Bolas, MT.; Merritt, JO.; Benton, SA., editors. Controlling perceived depth in stereoscopic images; Proceedings of the SPIE, Volume 4297: Stereoscopic Displays and Virtual Reality Systems VIII; 2001. p. 42-53.

Kusaka, H. Apparent depth and size of stereoscopically-viewed images. In: Rogowitz, BE., editor. Proceedings of SPIE, Volume 1666: Human Vision, Visual Processing, and Digital Display III; 1992. p. 476-482.

Kutka R. Reconstruction of correct 3d perception on screens viewed at different distances. IEEE Transactions on Communications. 1994; 42(1):29–33.

Leiser D, Bereby Y, Melkman A. Minimizing distortions: seating requirements for stereo projection rooms. Ergonomics. 1995; 38(6):1231–1238.

Lipton, L. Foundations of the Stereoscopic Cinema: A Study in Depth. Van Nostrand Reinhold; 1982.

Masaoka K, Hanazato A, Emoto M, Yamanoue H, Nojiri Y, Okano F. Spatial distortion prediction system for stereoscopic images. Journal of Electronic Imaging. 2006; 15(1):13002-1–13002-12.

Ogle KN. Induced size effect. I. A new phenomenon in binocular space-perception associated with the relative sizes of the images of the two eyes. Archives of Ophthalmology. 1938; 20:604–623.

Rogers BJ, Bradshaw MF. Vertical disparities, differential perspective, and binocular stereopsis. Nature. 1993; 361(6409):253–255. [PubMed: 8423851]

Rogers BJ, Bradshaw MF. Disparity scaling and the perception of frontoparallel surfaces. Perception. 1995; 24(2):155–179. [PubMed: 7617423]

Shapiro, LG.; Stockman, GC. Computer Vision. Prentice Hall; 2001.

Son JY, Gruts Y, Chun J, Choi YJ, Bahn JE, Bobrinev VI. Distortion analysis in stereoscopic images. Optical Engineering. 2002; 41(3):680–685.

Strunk, LM.; Iwamoto, T. A linearly-mapping stereoscopic visual interface for teleoperation. Proceedings of the IEEE International Workshop of Intelligent Robots and Systems; 1990. p. 429-436.

Wartell Z, Hodges LF, Ribarsky W. A geometric comparison of algorithms for fusion control in stereoscopic HTDs. IEEE Transactions on Visualization and Computer Graphics. 2002; 8(2):129–143.

Woods, AJ.; Docherty, T.; Koch, R. In: Merritt, JO.; Fisher, SS., editors. Image distortions in stereoscopic video systems; Proceedings of the SPIE Volume 1915, Stereoscopic Displays and Applications IV; 1993. p. 36-47.

Yamanoue H, Okui M, Okano F. Geometrical analysis of puppet theater and cardboard effects in stereoscopic HDTV images. IEEE Transactions on Circuits and Systems for Video Technology. 2006; 16(6):744–752.

Appendix: Underlying Geometry of Skew Rays

In Section 4.1, we used epipolar geometry to illustrate the geometric approach's inability to provide a solution under certain viewing situations. Here we provide more detailed derivations that produce the same results. For a given 3d point in a stereo image, recall that the terms $(\boldsymbol{P'_l} - \boldsymbol{E'_l})m$ and $(\boldsymbol{P'_r} - \boldsymbol{E'_r})n$ represent the rays passing from the centers of the eyes ($\boldsymbol{E'_l}$ and $\boldsymbol{E'_r}$) to the corresponding points on the screen ($\boldsymbol{P'_l}$ and $\boldsymbol{P'_r}$) in viewer-centered coordinates. For two rays to intersect, they must be non-parallel and lie in a common plane. We determine whether two rays lie in a plane by first finding the plane defined by the two eye centers and a point in the left stereo picture, and then by finding the plane defined by the two eyes and the corresponding point in the right stereo picture. If the two planes are coincident, the rays must intersect, provided that they are non-parallel. We define the planes by their surface normals. The first normal is given by the cross product between two vectors that originate at the left eye and extend to the point in the left stereo picture ($\boldsymbol{v_{l1}}$) and to the right eye ($\boldsymbol{v_{l2}}$) (Figure A1). The second plane is given by the cross product between the vectors originating at the left eye and extending to the point in the right stereo picture ($\boldsymbol{v_{r1}}$) and to the right eye ($\boldsymbol{v_{r2}}$).

$$\begin{aligned}\boldsymbol{v_{l1}}=\boldsymbol{P_l'}-\boldsymbol{E_l'} \quad \boldsymbol{v_{r1}}=\boldsymbol{P_r'}-\boldsymbol{E_l'}\\ \boldsymbol{v_{l2}}=\boldsymbol{E_r'}-\boldsymbol{E_l'} \quad \boldsymbol{v_{r2}}=\boldsymbol{E_r'}-\boldsymbol{E_l'}\end{aligned}$$

To include the viewer's position and orientation relative to the picture, we replace $\boldsymbol{P'_l}$ and $\boldsymbol{P'_r}$ with:

$$\boldsymbol{P_l'}=R(\rho,\sigma)(\boldsymbol{P_l}-\boldsymbol{E_c}) \quad \boldsymbol{P_r'}=R(\rho,\sigma)(\boldsymbol{P_r}-\boldsymbol{E_c})$$

We can now investigate how viewer translation ($\boldsymbol{E_c}$) and rotation ($R(\rho, \sigma)$) affect the above-defined planes. $\boldsymbol{v_{l2}}$ and $\boldsymbol{v_{r2}}$, the same vectors originating at the left eye and extending to the right eye, can be expressed as (I, 0, 0), where I is inter-ocular distance. Combining the equations, multiplying out the elements of $R(\rho, \sigma)$, and taking into account that $P_l(z)$ and $P_r(z)$ are both zero produces the following equations for the cross products:

$$\begin{aligned}\boldsymbol{v_{l1}}\times\boldsymbol{v_{l2}}=\boldsymbol{j}(I(\sin(\rho)\cos(\sigma)(P_l(x)-E_c(x))+\sin(\rho)\sin(\sigma)(P_l(y)-E_c(y))+\cos(\rho)(-E_c(z))))+\boldsymbol{k}(I(\sin(\sigma)(P_l(x)-E_c(x))+\cos(\\ \boldsymbol{v_{r1}}\times\boldsymbol{v_{r2}}=\boldsymbol{j}(I(\sin(\rho)\cos(\sigma)(P_r(x)-E_c(x))+\sin(\rho)\sin(\sigma)(P_r(y)-E_c(y))+\cos(\rho)(-E_c(z))))+\boldsymbol{k}(I(\sin(\sigma)(P_r(x)-E_c(x))+\cos\end{aligned} \quad ($$

We now have surface-normal representations for the two planes that originate at $\boldsymbol{E'_l}$. Before testing for equality, we normalize the vectors by dividing by their magnitudes $|\boldsymbol{v_{l1}} \times \boldsymbol{v_{l2}}|$ and $|\boldsymbol{v_{r1}} \times \boldsymbol{v_{r2}}|$. Then, to determine if the planes are coincident, we check to see if their $\boldsymbol{j}$ and $\boldsymbol{k}$ terms are equal to one another (there are no $\boldsymbol{i}$ terms in the equations above). The $\boldsymbol{j}$ terms are:

$$\begin{aligned}&\sin(\rho)\cos(\sigma)(P_l(x)\\&\quad-E_c(x))\\&\quad+\sin(\rho)\sin(\sigma)(P_l(y)\\&\quad-E_c(y))\\&\quad+\cos(\rho)(\\&\quad-E_c(z))\\&\quad=\frac{|\boldsymbol{v_{1r}}\times\boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}}\times\boldsymbol{v_{2l}}|}(\sin(\rho)\cos(\sigma)(P_r(x)\\&\quad-E_c(x))\\&\quad+\sin(\rho)\sin(\sigma)(P_r(y)\\&\quad-E_c(y))\\&\quad+\cos(\rho)(-E_c(z)))\end{aligned} \quad \text{(A2)}$$

and the $\boldsymbol{k}$ terms are:

$$\sin(\sigma)(P_l(x)-E_c(x))+\cos(\sigma)(P_l(y)-E_c(y))=\frac{|\boldsymbol{v_{1r}}\times\boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}}\times\boldsymbol{v_{2l}}|}(\sin(\sigma)(P_r(x)-E_c(x))+\cos(\sigma)(P_r(y)-E_c(y))) \quad \text{(A3)}$$

These two equalities can be used as tests for intersecting rays. If both equalities are valid, the tested rays will intersect, provided that they are not parallel; if the rays are parallel, we set the intersection at infinity, which is a sensible result. If they are not valid, the rays will not intersect and there will be no solution for the geometric approach. We now explore four acquisition-viewing conditions to determine which ones produce skew rays.

Observer Translation

In this condition, the viewer is translated relative to the picture. The viewer's face is parallel to the picture surface and not rotated, so the *X*, *Y*, and *Z* axes in the picture- and viewer-centered coordinate systems are parallel. Because the axes are parallel, ρ and σ in the rotation matrix *R*(ρ, σ) are 0. If these values are plugged in to Equations A1, A2, and A3, the ***j*** and ***k*** terms of the surface normals become:

$$-E_c(z) = -\frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|} E_c(z)$$
$$\&P_l(y) - E_c(y) = \frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|}(P_r(y) - P_c(y)),$$
$$where\ \frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|} = \frac{|\boldsymbol{j}(I(-E_c(z)) + \boldsymbol{k}(I(P_l(y) - E_c(y)))|}{|\boldsymbol{j}(I(-E_c(z)) + \boldsymbol{k}(I(P_r(y) - E_c(y)))|}$$

Examining the equations, it is clear that the two equalities are only valid if the *Y* coordinates of corresponding points on the picture are equal to one another. In other words, there must be no *vertical* disparities in the picture. On-screen vertical disparity is always zero when the images are captured with cameras whose lens axes are parallel, even if the optical axes are converged (i.e., the sensors are offset relative to the lens axes). Therefore, when the cameras' lens axes are parallel and the viewer's eye coordinates are parallel to the picture-centered coordinates, there are intersections for all ray pairs and the geometric approach provides a solution.

Observer Rotation in *X–Z* Plane

Viewer rotation in the *X–Z* plane—"yaw" rotation—occurs when the viewer is positioned to the left or right of the proper position and turns the head toward the center of the stereo picture. ρ is now nonzero, but σ is still zero. Equations A1 and A2 then become:

$$\sin(\rho)(P_l(x) - E_c(x)) + \cos(\rho)(-E_c(z)) = \frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|}(\sin(\rho)(P_r(x) - E_c(x)) + \cos(\rho)(-E_c(z)))$$
$$\&P_l(y) - E_c(y) = \frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|}(P_r(y) - E_c(y)),$$
$$where\ \frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|} = \frac{|\boldsymbol{j}(I(\sin(\rho)(P_l(x) - E_c(x)) + \cos(\rho)(-E_c(z)))) + \boldsymbol{k}(I(P_l(y) - E_c(y))|}{|\boldsymbol{j}(I(\sin(\rho)(P_r(x) - E_c(x)) + \cos(\rho)(-E_c(z)))) + \boldsymbol{k}(I(P_r(y) - E_c(y))|}$$

In this case, $P_l(y)$ and $P_r(y)$ must be equal to $E_c(y)$ for the dual equalities to be valid. The only corresponding points that will produce intersecting rays are points in the *X–Z* plane; all rays above and below that plane will be non-intersecting (skew rays), except when $\boldsymbol{P_l}$ and $\boldsymbol{P_r}$ are identical (zero disparity). Because rays for many points in the stereo picture do not intersect, the geometric approach cannot provide a prediction for what viewers should perceive. This is disappointing because yaw rotations are commonplace in the viewing of stereo media.

Observer Rotation in *X–Y* Plane

With viewer rotation in the *X–Y* plane—"roll" rotation—σ is nonzero and ρ is zero, which results in

$$
\begin{gathered}
-E_c(z) = -\frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|} E_c(z) \\
\& \sin(\sigma)(P_l(x) - E_c(x)) + \cos(\sigma)(P_l(y) - E_c(y)) = \frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|} \sin(\sigma)(P_r(x) - E_c(x)) + \cos(\sigma)(P_r(y) - E_c(y))), \\
where\ \frac{|\boldsymbol{v_{1r}} \times \boldsymbol{v_{2r}}|}{|\boldsymbol{v_{1l}} \times \boldsymbol{v_{2l}}|} = \frac{|\boldsymbol{j}(-IE_c(z)) + \boldsymbol{k}(I(\sin(\sigma)(P_l(x) - E_c(x)) + \cos(\sigma)(P_l(y) - E_c(y))))|}{|\boldsymbol{j}(-IE_c(z)) + \boldsymbol{k}(I(\sin(\sigma)(P_r(x) - E_c(x)) + \cos(\sigma)(P_r(y) - E_c(y))))|}
\end{gathered}
$$

Equations A2 and A3 are only valid if $\boldsymbol{P_l}$ is identical to $\boldsymbol{P_r}$. Any non-zero disparity produces skew rays if the head is rolled relative to the picture. Thus, this is another situation in which the geometric approach cannot provide a prediction of viewers’ percepts.

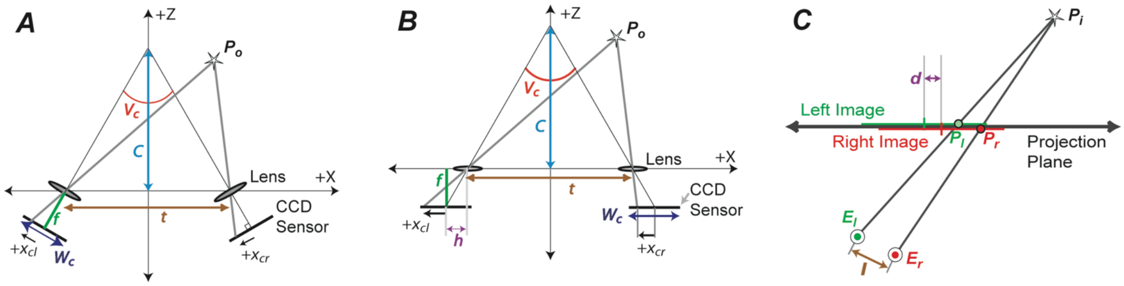

Figure 1.

(A) Image formation with converging cameras. $\mathbf{P_o}$ is coordinates of point P, f is camera focal length, t is separation between the cameras, C is distance to which the camera optical axes are converged, V_c is angle between cameras' optical axes, W_c is width of camera sensors, x_{cl} and x_{cr} are x-coordinates of P's projection onto left and right camera sensors. (B) The cameras' optical axes can be made to converge by laterally offsetting the sensors relative to the lens axes. h is offset between sensor center and intersection of lens axis with the sensor. (C) Reconstruction of P from sensor images. Rays are projected from eye centers through corresponding points on picture. The ray intersection is estimated location of P. $\mathbf{E_l}$ and $\mathbf{E_r}$ are 3d coordinates of left and right eyes; $\mathbf{P_l}$ and $\mathbf{P_r}$ are locations of image points in the picture of P for left and right eyes; I is inter-ocular distance; d is distance between centers of pictures. The green and red horizontal lines represent the images presented to the left and right eyes, respectively.

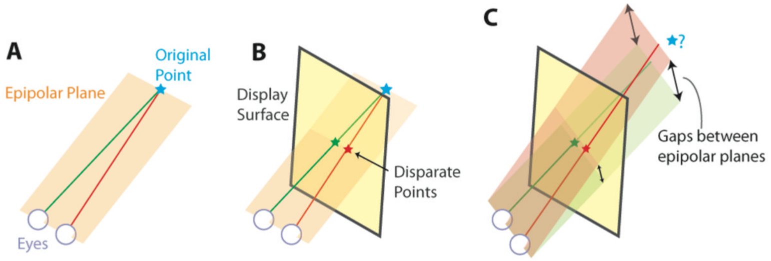

Figure 2.
Epipolar geometry. A) In natural viewing, a point in space and the two eye centers define an epipolar plane. B) If a viewer is correctly positioned relative to the picture, the rays emanating through the eyes and passing through a pair of corresponding points in the picture lie in the same epipolar plane and intersect in space. C) With oblique viewing (head rotated about a vertical axis such that inter-ocular axis is not parallel to picture surface), rays will generally lie in different epipolar planes and never intersect.

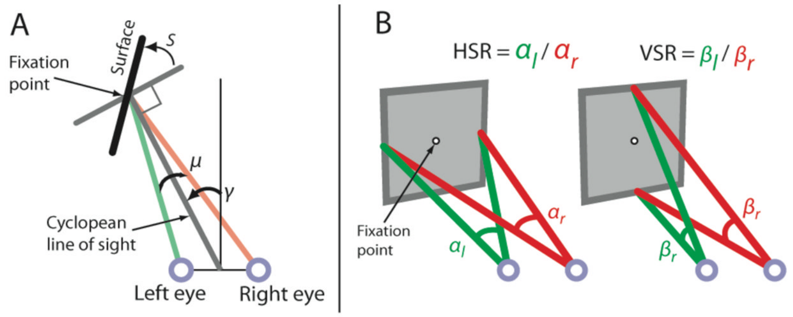

Figure 3.

A) Plan view of viewer fixating a planar surface. S is the slant of the patch. μ is eyes' horizontal vergence; γ is eyes' horizontal version. B) Definitions of HSR (horizontal size ratio) and VSR (vertical size ratio). HSR is the ratio of the horizontal angles a surface patch subtends at left and right eyes. VSR is the ratio of vertical angles. Adapted from Backus et al. [1999].

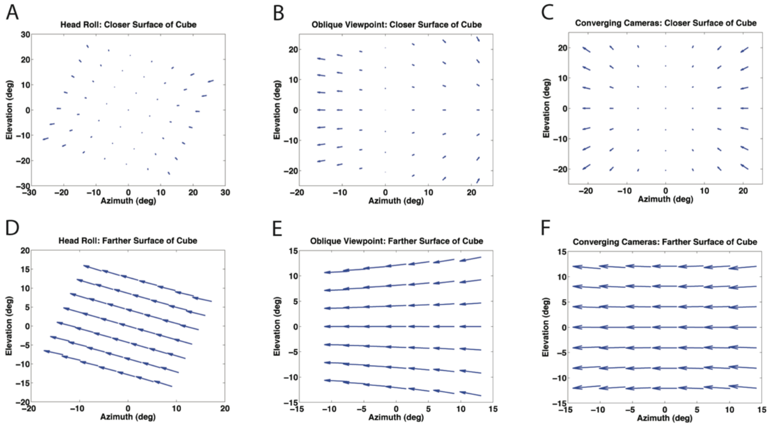

Figure 4.

Disparity as a function of azimuth and elevation. Fick coordinates (azimuth and elevation measured as longitudes and latitudes, respectively) were used. Vectors represent the direction and magnitude of disparities on the retinas produced by a stereoscopic image of a cube 0.3m on a side and placed 0.55m in front of the stereo cameras. Unless otherwise noted, the conditions listed in Section 3 were used to generate the figures. Arrow tails represent points on right eye's retina, and arrowheads represent corresponding points in left eye's retina. Panels A, B, and C contain points from the proximal face of the cube, where the eyes are fixating. D, E, and F represent the cube's distal face. In A and D, the observer is viewing the display at a 45deg angle. In B and E, the viewer's head has been rolled 20deg. In C and F, the cameras converge at 0.55m.

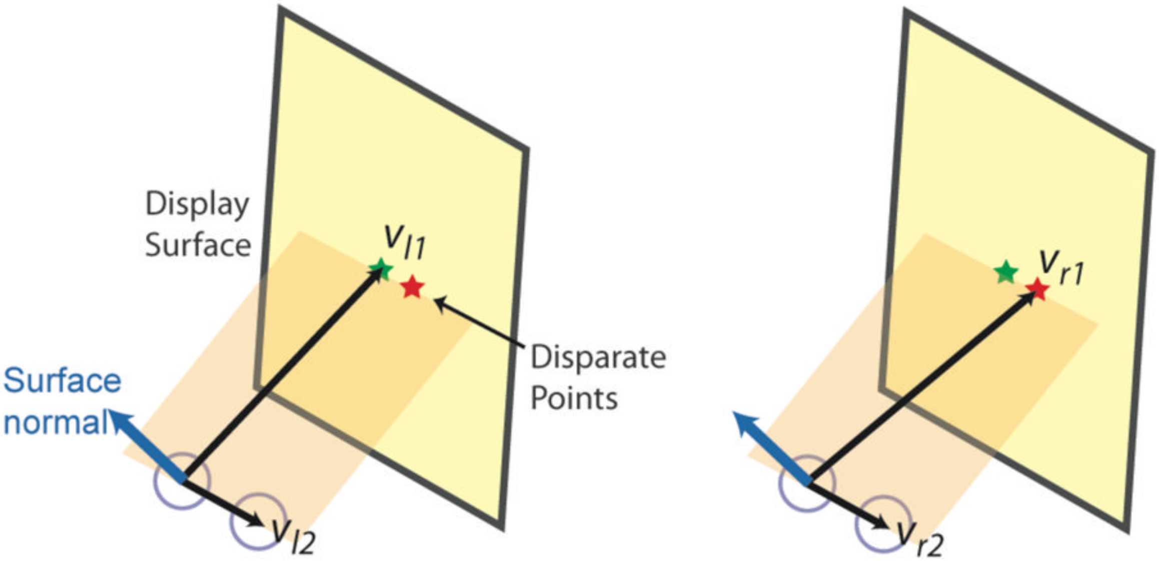

Figure A1.

Planes defined using the centers of both eyes and either of the corresponding points in the pictures. The cross product of the illustrated vectors (from one eye to the other and from one eye to the image point) is a normal vector that defines the plane.

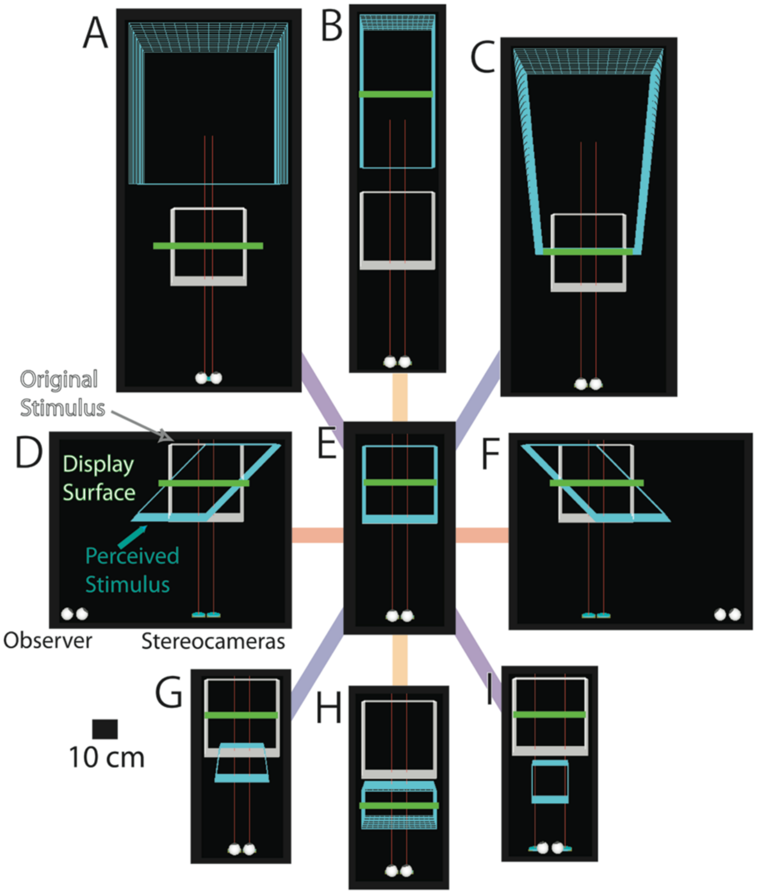

Figure 1.
Estimated 3d scenes for different acquisition and viewing situations. Each panel is a plan view of the viewer, stereo cameras, display surface, actual 3d stimulus, and estimated 3d stimulus. Red lines represent cameras' optical axes. E) Proper viewing situation. Parameters are listed in Section 3. The actual and estimated stimuli are the same. B) Viewer is too distant from picture. H) Viewer is too close. D) Viewer is too far to the left relative to the picture. F) Viewer is too far to the right. A) Cameras are too close together for viewer's inter-ocular distance. I) Cameras are too far apart. C) Distance between centers of the left and right stereo pictures is too great. G) Distance between the centers of pictures is too small.

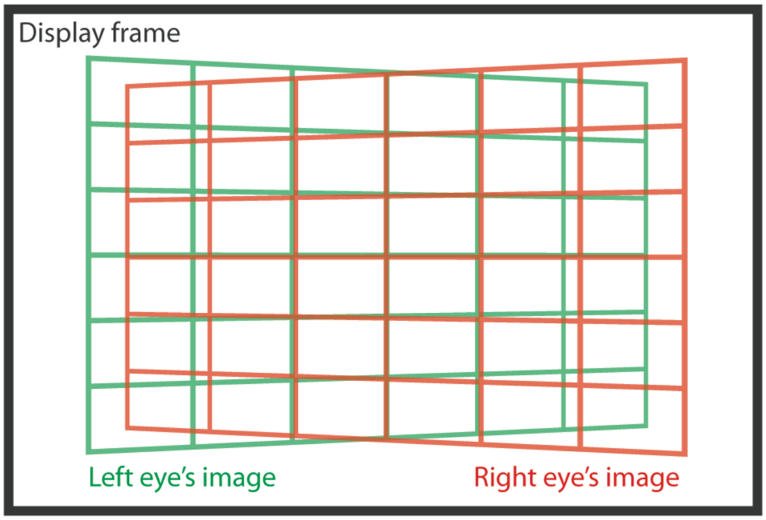

Figure 2.
The keystone effect. A rectangular grid was captured using converging cameras and displayed on a single flat display surface. Note the vertical disparities between the corresponding points in the corners.

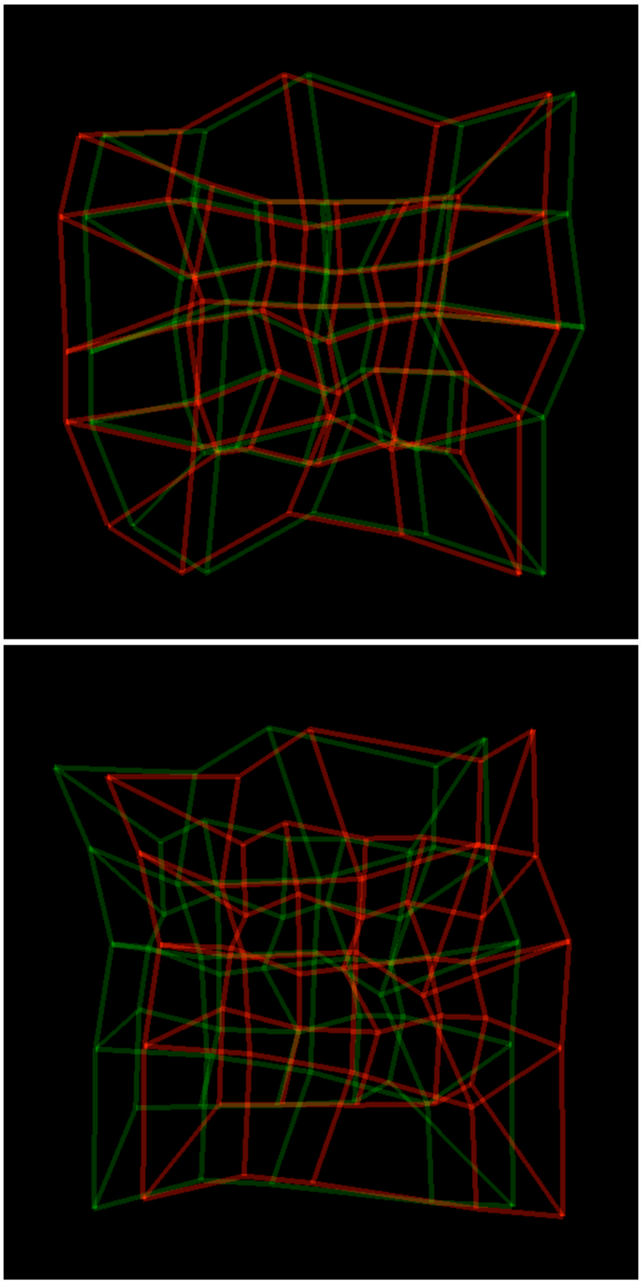

Figure 3.
Anaglyph stereograms captured with the acquisition settings listed in Section 3.1. Top: cameras with parallel optical axes. Bottom: cameras' optical axes were converged at 0.55m (center of cube). To view the stereograms, use red-green glasses with green filter over left eye. Try different viewing situations. 1) Move closer to and farther away from the page. 2) Move left and right while holding the head parallel to the page. 3) Position yourself directly in front of the page and rotate the head about a vertical axis (yaw) and then about a forward axis (roll). In each case, notice the changes in the cube's apparent shape. Points in the cube were randomly perturbed to lessen contributions of perspective cues to 3d percept.