



Published in final edited form as:

*Inf Disp* (1975). 2009 January ; 25(1): 12–16.

## Perception of 3-D Layout in Stereo Displays

**Martin S. Banks,**

Professor in the Vision Science Program, School of Optometry, University of California at Berkeley, Berkeley, California; telephone 510/642-7679

**Robert T. Held,** and

Works in the Joint Graduate Group in Bio-engineering, University of California, San Francisco, California, and the University of California at Berkeley. Berkeley, California

**Ahna R. Girshick**

Works at the Department of Psychology, Center for Neural Science, New York University, New York, New York

Martin S. Banks: martybanks@berkeley.edu

### Abstract

As stereoscopic displays become more commonplace, it is more important than ever for those displays to create a faithful impression of the 3-D structure of the object or scene being portrayed. This article reviews current research on the ability of a viewer to perceive the 3-D layout specified by a stereo display.

---

Stereoscopic displays have become very important for many applications, including vision research, operation of remote devices, medical imaging, surgical training, scientific visualization, virtual prototyping, and more. It is important in these applications for the graphic image to create a faithful impression of the 3-D structure of the object or scene being portrayed. Here we review current research on the ability of a viewer to perceive the 3-D layout specified by a stereo display. To do so, we will first consider conventional displays (*i.e.*, pictures such as photographs) and then consider stereo displays.

Conventional pictures (photographs, cinema, computer-graphics images, *etc.*) are very useful because in the convenient format of a 2-D surface they allow viewers to perceive 3-D scene information. At one level, it seems obvious why pictures provide such useful information: A conventional picture viewed from its center of projection (CoP) generates the same retinal image as the original scene, so a well-positioned viewer understandably perceives the depicted scene as similar to the original scene. Such pictures, however, would not be very useful if the viewer's eye always had to be positioned at the CoP to create an acceptable impression. Imagine, for example, that there is only one seat in the cinema that produced a percept that was acceptably close to the depicted scene. Fortunately, when pictures are viewed from other locations, the perceived scene does not seem significantly different, even though the retinal image now specifies a different scene<sup>1</sup>; thus, people can sit in various locations in a theater and gain an acceptable impression of a motion picture.

We have been experimentally investigating the ability to compensate for incorrect viewing position when viewing conventional pictures. In one set of experiments,<sup>2</sup> we had subjects judge the aspect ratio of an ovoid-shaped object in a depicted scene rich with geometric cues. The CoP of the stimulus was directly in front of and 45 cm from the computer display. Subjects viewed the stimulus binocularly from a variety of positions ranging from the appropriate one (the CoP) to positions too far to the left or too far to the right. We accomplished this by rotating the display rather than by moving the subject. [Imagine an overhead view of the apparatus. The observer's head position was fixed and the CRT display

was rotated about a vertical axis. Different amounts of rotation corresponded to different “viewing angles” on the abscissa in Fig. 1(b) When the viewing angle was zero, the observer was positioned at the CoP of the stimulus; otherwise, the viewer was not at the CoP.] The rotation caused large changes in the shape of the projected ovoid in the retinal image. We found that subjects nonetheless perceived the shape of the ovoid on the display screen essentially correctly – provided that they viewed the display binocularly – even when they were more than  $30^\circ$  from the CoP. Thus, human viewers can compensate for incorrect viewing positions and thereby achieve essentially complete perceptual invariance with conventional pictures.

In another set of experiments, we investigated the perception of a 3-D shape depicted in a conventional picture. The stimulus was a vertical hinge in an open-book configuration; an example is shown in Fig. 1(a). The hinge was presented in perspective projection on a conventional display screen. Subjects viewed the stimulus from a variety of positions ranging from the appropriate one (the CoP) to positions too far to the left. We accomplished this by rotating the display rather than by moving the subject. Of course, the retinal images for a given hinge stimulus on the computer display differed depending on viewing position. By using a psychophysical procedure, we found the hinge angle that on average was perceived as  $90^\circ$ .

Figure 1(b) plots predictions and results. The hinge angle that was perceived as  $90^\circ$  is plotted as a function of viewing angle; different colors correspond to different base slants. If subjects were able to compensate for incorrect viewing position, a hinge that was depicted as  $90^\circ$  would be perceived as such: responses would follow the horizontal black line at  $90^\circ$ . If subjects were unable to compensate for incorrect viewing position and instead estimated the hinge angle from the geometric pattern in the retinal images, a  $90^\circ$  hinge would no longer be perceived as such; responses would then follow the dashed colored lines, one for each base slant. The results were generally in between the compensation and no-compensation predictions, so they show that human viewers of 2-D pictures are able to compensate partially for incorrect viewing position and thereby achieve some degree of perceptual invariance. This result is reasonably consistent with our previous work,<sup>2</sup> but shows that the amount of perceptual invariance depends on the depth variation in the stimulus.

## Perception of Stereo Pictures

Stereo pictures have all the properties of conventional pictures plus binocular disparity (*i.e.*, spatial differences in the two retinal images); disparity yields the compelling sensation of depth we enjoy when viewing 3-D content. The viewing parameters are often not correct in practical uses of stereo displays. For example, the great majority, if not all, of the people viewing a stereo movie will not have their left and right eyes at the appropriate CoPs. We next examined whether viewers can compensate for incorrect viewing position with stereo pictures as they do with conventional pictures.

The standard model in the stereo-cinema literature equates changes in the pattern of disparities at the retinas with the predicted 3-D percept<sup>3</sup>; *i.e.*, it assumes that viewers of stereo pictures do not compensate for incorrect viewing position. This is a significant assumption that should be seriously examined, particularly in light of the fact that viewers of conventional pictures do compensate for incorrect position. We will return to this assumption later. The standard model uses a ray-intersection algorithm. Each corresponding point within a pair of stereo pictures is projected onto the left and right retinas. From the retinal points, rays are projected out through the centers of the eyes into space. The intersection of those rays is the predicted 3-D location of the specified point in space. Applying the ray-intersection algorithm to each pair of corresponding points in the stereo

picture produces a 3-D percept of the entire virtual scene. For the geometrically predicted 3-D percept to match the original scene, several image acquisition, display, and viewing parameters must be appropriate for one another. The acquisition (camera) parameters include orientation (whether the cameras' optical axes are parallel or toed-in), inter-camera separation, and focal length. Display parameters include the magnification of the pictures and whether one or two display devices are used to present the pictures (in vision research, two displays are commonly used, one for each eye; in most everyday applications, one display is used and both pictures are presented on it). Whether one or two displays are used, the lateral separation between the two pictures must be appropriate to preserve the correct vergence angle for the viewer's eyes. The viewing parameters are the positions of the two eyes relative to the CoPs of the stereo pictures and the vergence angle induced by disparate points on the display.

We used a software implementation of the geometric approach to investigate the effects of viewer position and orientation on retinal images.

To allow comparison with our experimental results (Figs. 1 and 4), the stimulus in the simulation presented here was a vertical hinge. All of the parameters are correct in Fig. 2(c), so the predicted 3-D percept is identical to the original hinge photographed by the stereo cameras. Figures 2(a) and 2(e) show the predicted consequences of positioning the viewer respectively too close to or too far from the display. When the viewing distance is too short, the predicted perceived hinge angle is larger than  $90^\circ$ ; when the distance is too great, the predicted angle is smaller than  $90^\circ$ . Figures 2(b) and 2(d) show the consequences of translating the viewer to the left or right: the predicted perceived hinge rotates toward the viewer and the predicted angle becomes less than  $90^\circ$ . These predictions are derivable from previous analyses in the stereo cinema literature<sup>3</sup>

When the viewer translates, the intersecting-ray approach still works because all pairs of corresponding points in the retinal images produce rays that intersect in space. The fact that they intersect can be understood from epipolar geometry.<sup>4</sup> An epipolar plane is the plane containing a point in visible space and the centers of the two eyes. If the viewer is translated relative to the correct viewing position but does not rotate the head, it can be shown that the rays produced by point pairs in the stereo pictures lie in the same epipolar plane<sup>5</sup> [Figs. 3(a) and 3(b)]. Any two non-parallel rays that lie in a common plane are guaranteed to intersect, so the intersecting-ray approach yields a prediction for those viewing situations.

Unfortunately, many common stereo viewing conditions violate epipolar geometry and therefore preclude a solution based on ray intersection. One such condition occurs when a viewer is positioned to the left or right of center and rotates the head to face the center (a yaw rotation). In this case, most of the rays produced by the corresponding points in the retinal images do not intersect [Fig. 3(c)]. The standard model relies on ray intersections, so with yaw rotations it cannot predict a percept. Interestingly, human viewers in this situation still have a coherent 3-D percept. The standard model, therefore, has to be modified. One modification of the model forces the non-intersecting rays into a common epipolar plane,<sup>3</sup> but there is no evidence that the human visual system uses such a method. The non-intersecting rays introduce vertical disparities at the retinas and research has shown that those disparities are used to estimate the 3-D layout of the scene.<sup>5,6</sup> A more complete model of the perception of 3-D pictures would incorporate the use of vertical disparities. With an appropriate modification, the model would be able to make predictions for 3-D percepts for a wider range of viewing situations, including combinations of viewer translation and rotation that are likely to be encountered in the viewing of stereo pictures.

As we noted, the standard model<sup>3,5</sup> assumes that the 3-D percept is dictated solely by the retinal images, which is equivalent to assuming that viewers do not compensate for incorrect viewing position. This is a significant claim with far-reaching implications for the creation and presentation of stereo content. Thus, we decided to test the assumption in an experiment similar to the one described in Fig. 1. The hinge stimulus, which is shown in Fig. 4(a), was similar to the one used in the conventional picture experiment [Fig. 1(a)] except that now its 3-D shape was specified by disparity along with the perspective cues present in the 2-D version of the experiment. As before, subjects viewed the stimulus from various positions ranging from the appropriate one to positions that were too far to the left. Figure 4(b) plots predicted and observed hinge angles that were perceived as 90°. If subjects were able to compensate for incorrect viewing position, any hinge that was depicted as 90° would be perceived as such: the results would lie on the horizontal black line at 90°. The no-compensation predictions were generated from the model in Fig. 2. If subjects did not compensate for incorrect position and instead estimated the hinge angle from the retinal disparities, a 90° hinge would no longer be perceived as 90°: the results would then follow the dashed colored curves. As one can see, the results were nearly identical to the no-compensation predictions.

As the results in Fig. 4(b) show, misperceptions occur when the viewer's eyes are not positioned correctly relative to a stereo picture. The percepts are well predicted from the ray-intersection model (Fig. 2). The results of this experiment coupled with the results for viewing of conventional pictures have profound implications: percepts from stereo pictures are significantly more affected by incorrect viewing position than are percepts from conventional pictures.

We hasten to point out that other visual cues are frequently incorrect in stereo displays – blur and accommodation are two prominent ones<sup>7,8</sup> – and they too can cause misperceptions. Those perceptual effects are, however, beyond the scope of this brief review.

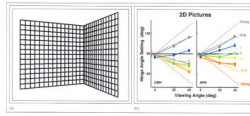
## Conclusion

In summary, our findings to date indicate that human viewers of stereo pictures are unable to compensate for incorrect viewing position. As a result, the 3-D percept seems to be determined only by the disparities in the retinal images. Further research is needed to determine whether other information, such as motion parallax, can aid compensation. At the moment, however, it appears that the perceptual invariance that makes audience viewing of conventional pictures acceptable does not occur to nearly the same degree with stereo pictures. Designers of stereo viewing systems should therefore carefully plan the acquisition, display, and viewing parameters so that the viewer can have a 3-D percept that is as faithful to the original scene as possible.

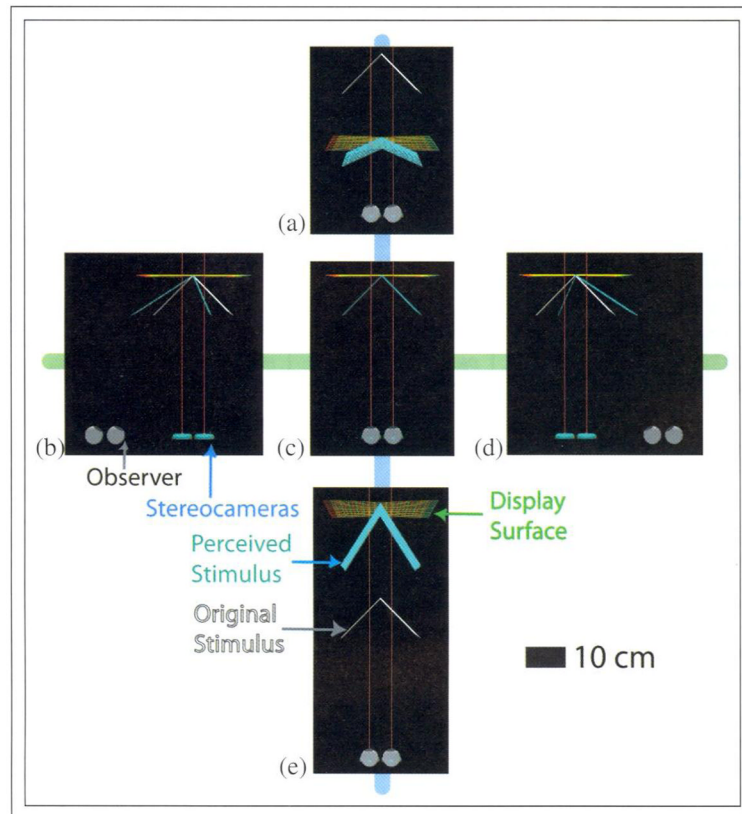
## References

1. Kubovy, M. *The Psychology of Perspective and Renaissance Art*. Cambridge University Press; Cambridge. U.K: 1986.
2. Vishwanath D, Girshick AR, Banks MS. Why pictures look right when viewed from the wrong place. *Nature Neuroscience*. 2005; 8(10):1401–1410.
3. Woods, AJ.; Docherty, T.; Koch, R. Image distortions in stereoscopic video systems. In: Merritt, JO.; Fisher, SS., editors. *Proc SPIE: Stereoscopic Displays and Applications IV*. 1993. p. 36-47.
4. Shapiro, LG.; Stockman, GC. *Computer Vision*. Prentice Hall; 2001.

5. Held, RT.; Banks, MS. Misperceptions in stereoscopic displays: a vision science perspective. Proc. 5th Symposium on Applied Perception in Graphics and Visualization (APGV 08); New York, NY: ACM; 2008.
6. Backus BT, Banks MS, van Ee ER, Crowell JA. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*. 1999; 39(6):1143–1170. [PubMed: 10343832]
7. Watt SJ, Akeley K, Ernst MO, Banks MS. Focus cues affect perceived depth. *J Vision*. 2005; 5(10): 7, 834–862.
8. Hoffman DM, Girshick AR, Akeley K, Banks MS. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J Vision*. 2008; 8(3):33, 1–30.

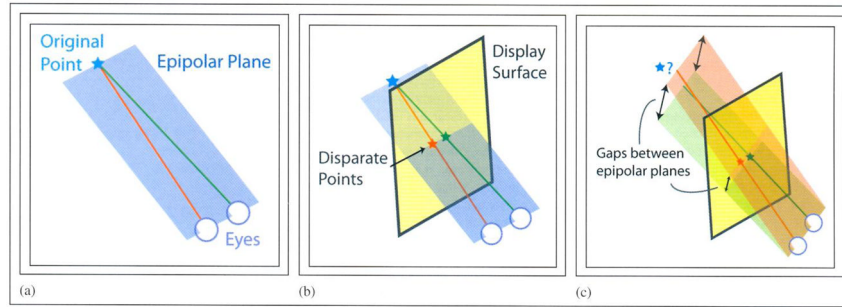


**Fig. 1.** Hinge stimulus, predictions, and results for the experiment with conventional pictures. (a) An example of the hinge stimulus. The stimulus was presented on a conventional flat-screen display and viewed binocularly. Viewing angle (the angle between a line from the hinge to the center of projection and a line from the hinge to the viewer) was varied by rotating the display about a vertical axis. (b) Predictions and results. The hinge angle in the depicted stimulus that was on average perceived as  $90^\circ$  is plotted as a function of viewing angle. The left and right panels show data from subjects DMH and HRF, respectively. The compensation prediction is represented by the horizontal black lines at  $90^\circ$ . The no-compensation predictions are represented by the dashed colored lines, each line representing a different base slant. The symbols represent the data, the colors corresponding to different base slants. Error bars represent 95% confidence intervals.



**Fig. 2.**

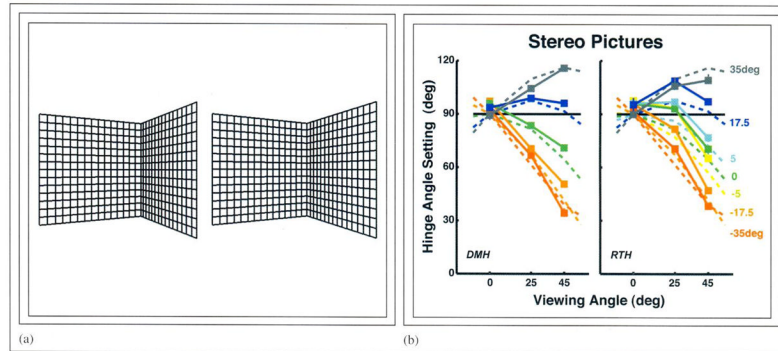
Predicted 3-D percepts for the hinge stimulus for different viewing situations. Each panel shows an overhead view of the observer (gray), stereo cameras (blue), display surface (yellow), original stimulus (gray), and the predicted perceived stimulus (blue). The parameters used in the simulation are the following. Acquisition (camera) parameters: Parallel orientation of optical axes, inter-camera separation of 6.2 cm, focal length of 6.5 mm. Display parameters: one display device, picture magnification (projected size divided by film size) of 69.2. Viewing parameters: viewing distance of 45 cm, inter-ocular distance of 6.2 cm, viewer positioned such that midpoint of inter-ocular axis is on central surface normal of display device, viewer oriented with face parallel to display surface, stimulus is a  $30 \times 30$ -cm vertical hinge with a hinge angle of  $90^\circ$ . (c) With all parameters correctly set, the original and predicted perceived stimuli are identical. (a) The viewer is too close to the display. The predicted perceived hinge angle is greater than  $90^\circ$ . (e) Viewer is too far from the display. The perceived angle is now less than  $90^\circ$ . (b) Viewer is translated to the left of the display. The predicted hinge rotates toward the viewer and the predicted angle is less than  $90^\circ$ . (d) Viewer is translated to the right of the display. The predicted hinge rotates toward the viewer and the predicted angle is less than  $90^\circ$ .



**Fig. 3.**

Epipolar geometry. (a) In the natural environment, an epipolar plane is defined by a point in visible space and the centers of the two eyes, (b) If the acquisition, display, and viewing parameters are correctly set, the epipolar planes produced by two corresponding points in the left- and right-eye pictures will coincide and the rays projected from the eyes through those points will intersect in space. (c) If the viewer's head is rotated about a vertical axis (yaw rotation), the corresponding points in the left- and right-eye pictures produce rays that generally do not intersect because they lie in different epipolar planes.





**Fig. 4.** Hinge stimulus, predictions, and results for the experiment with stereo pictures, (a) An example of the hinge stimulus. Cross-fuse the stimulus (direct the right eye to the left image and the left eye to the right image) to see it stereoscopically. Separate stimulation of the two eyes was accomplished by using liquid-crystal shutter glasses that were synchronized to the computer display. Viewing angle was varied by rotating the display about a vertical axis. (b) Predictions and results. The hinge angle in the depicted stimulus that was on average perceived as  $90^\circ$  is plotted as a function of viewing angle. The left and right panels show data from subjects DMH and RTH, respectively. The compensation prediction is represented by the horizontal black lines at  $90^\circ$ . The no-compensation predictions are represented by the dashed colored lines, each line representing a different base slant. Because the viewer was translated and rotated from the correct viewing position, epipolar geometry was not strictly followed. We made the predictions based on the disparities at the horizontal meridians of the eyes where rays from the eyes do intersect in space. The symbols represent the data, the colors correspond to different base slants. Error bars represent standard errors.