QUENTIN E. DOLECEK

COMPUTER–GENERATED STEREOSCOPIC DISPLAYS

Although stereoscopy has been with us for over 150 years, the preparation of computer-generated perspective views and formation of stereoscopic images are not obvious. This article discusses how people see three-dimensionally and stereoscopically, how electrostereoscopic displays work, and how they mesh with the visual system. The use of visual cues in stereoscopic images and the use of a distortion-free algorithm for generating stereoscopic images are also presented.

INTRODUCTION

Through the ages, men like Euclid and Leonardo da Vinci understood that we see different images of the world with each eye. Kepler and others wondered why we do not see a double image of the visual world. It was Wheatstone[1] in 1838 who explained to the world that the problem was actually the solution. He demonstrated with his stereoscope and drawings that the mind fuses two planar retinal images into one image with stereoscopic depth. This fusion by the mind of two images into one that contains depth information is the basis for all stereo imaging today.

RETINAL DISPARITY

Try the experiment shown in Figure 1. If you hold a thumb in front of your face and look at the thumb, the eyes will converge on the thumb; that is, the optical axes of both eyes will cross on the thumb. Sets of muscles move the eyes to accomplish this by placing the images of the thumb on each fovea or central portion of each retina. If you continue to converge your eyes on the thumb, paying attention to the background, the background will appear to be double. Conversely, when looking at the background, your eyes will converge on it, and the thumb will now appear to be double.

If the images on the left and right retinae were superimposed as if they were slides, you would see two almost overlapping images: left and right perspective viewpoints that have disparity. Disparity is the distance, in a horizontal direction, between the corresponding left and right image points of the superimposed retinal images. The corresponding points of the retinal images of an object on which the eyes are converged will have zero disparity.

Retinal disparity occurs because each eye sees the world from a different perspective, as on average the eyes of adults are 64 mm apart. Disparity is processed by the eye–brain into a single image of the visual world. The mind's ability to combine two different, although similar, images into one image is called fusion, and the resultant sense of depth is called stereopsis.

ACCOMMODATION AND CONVERGENCE

When looking at an object in the visual world, the eyes focus (*accommodate* by changing the shape of the lenses) and *converge* (rotate toward or away from each other). That is, the eyes focus on the object so that it may be seen clearly and also converge on the object so that it may be seen singly. The latter process is called fusion.

When looking at a stereoscopic display, the eyes accommodate on the plane of the screen but converge because of screen parallax. The axes of the eyes will be converged as if portions of a stereoscopic image are at different distances, but they will remain focused on the
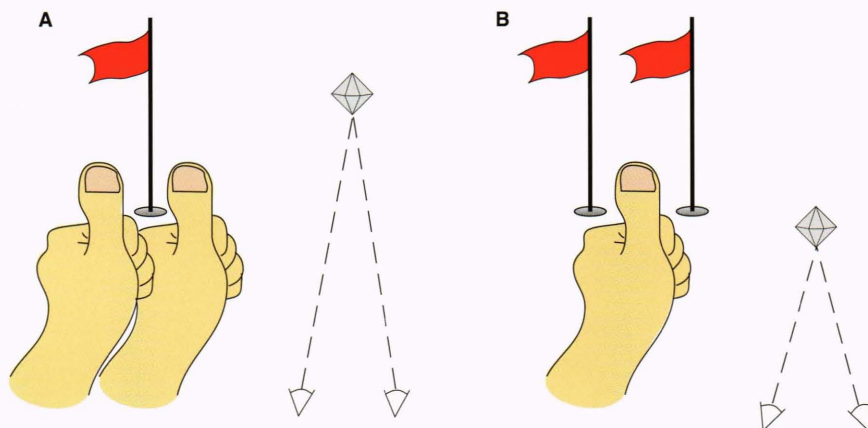

Figure 1. Each eye sees the world from a different perspective and rotates to fuse images of an object at different distances. **A.** If eyes are converged on the background, the thumb is seen double. **B.** If eyes are converged on the thumb, the background is seen double.

plane of the screen. This is the only significant manner in which an electrostereoscopic display differs from the way we see objects in the visual world.

PARALLAX

Parallax and disparity are similar entities. Parallax is measured at the display screen, and disparity is measured at the retinae; parallax produces retinal disparity, and disparity, in turn, produces stereopsis.

Four basic types of parallax are illustrated in Figure 2. In the first type, zero parallax, the homologous image points of the two images exactly correspond to or lie on top of each other, and the eyes are converged at the plane of the screen (i.e., the optical axes of the eyes cross at the plane of the screen). When image points have zero parallax, they are said to have "zero parallax setting."

In positive parallax the axes of the left and right eyes are parallel. (This occurs in the visual world when looking at objects at a great distance.) For a stereoscopic display, when the distance between the eyes *t* equals the parallax, the axes of the eyes will be parallel, just as they are when looking at a distant object in the visual world. For a small-screen display, parallax values approximately equal to *t* will produce discomfort.

In divergent parallax the images are separated by some distance greater than *t*. In this case, the axes of the eyes are diverging. Any uncrossed or positive value of parallax between zero and *t* will produce images appearing to be within the space of a cathode ray tube display or behind the screen. This divergence does not occur when looking at objects in the visual world, and the unusual muscular effort needed to fuse such images may cause discomfort. There is no valid reason for divergence in computer-generated stereoscopic images.

In negative parallax the axes are crossed and the parallax points are said to be crossed or negative. Objects with negative parallax appear to be closer than the plane of the screen or between the observer and the screen. We say that objects with negative parallax are within viewer space.

The accommodation/convergence relationship discussed in the previous section is a habitual, learned response. Invariably, in looking at objects in the visual world, accommodation and convergence correspond. In the example given earlier, when the eyes were converged on the thumb, they were also focused on the thumb. When the eyes were converged on the background, they were also focused on the background. However, an exception to this relationship results when looking at an electrostereoscopic image. The axes will be converged as if portions of a stereoscopic image are at different distances, but focused on the plane of the screen.

The muscles controlling convergence and those controlling focusing depart from their habitual relationship. This may cause some people to experience an unpleasant sensation when looking at stereoscopic images, especially images with large values of parallax. The lowest values of parallax compatible with a good depth will minimize the breakdown of accommodation/convergence, thus reducing viewer discomfort. A parallax value of about $1.5°$ produces clear and comfortable stereoscopic images. This translates to about 12 mm of separation for images to be observed from typical workstation-viewing distances.

In only one case, the zero parallax setting, will no breakdown of the accommodation/convergence relationship occur. If the composition allows, it is best to place the principal object(s) at or near the plane of the screen, or to split the difference in parallax by fixing the zero parallax setting on the middle of an object so that half the parallax values will be positive and half negative.

CONGRUENT NATURE OF THE FIELDS

To produce high-quality stereoscopic images, the left and right image fields must be identical except for the values of horizontal parallax. The color, geometry, and brightness of the left and right image fields must be the same or viewer "eyestrain" will result. If a system produces image fields that are not congruent in these respects, high-quality stereoscopic images cannot be achieved.

COMPOSITION

A stereoscopic image that has a one-to-one correspondence with the visual world may be uncomfortable to look at and of questionable value to the scientist, engineer, technician, or artist because of accommodation/convergence breakdown, conflicting screen surround cues, and the restrictions imposed for orthostereoscopic displays. Far more important than one-to-one correspondence with the visual world are the composition and depth in a scene.

Depth Cues

Monocular depth cues, illustrated in Figure 3, are the basis for the perception of depth in flat visual displays and are just as important as stereopsis for creating images
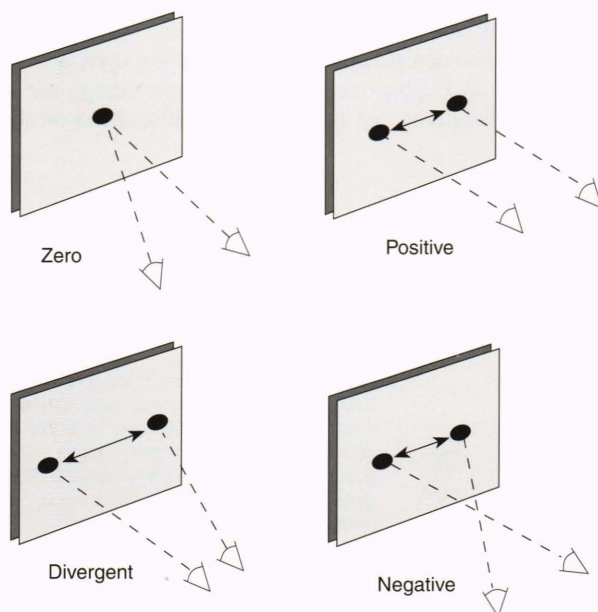

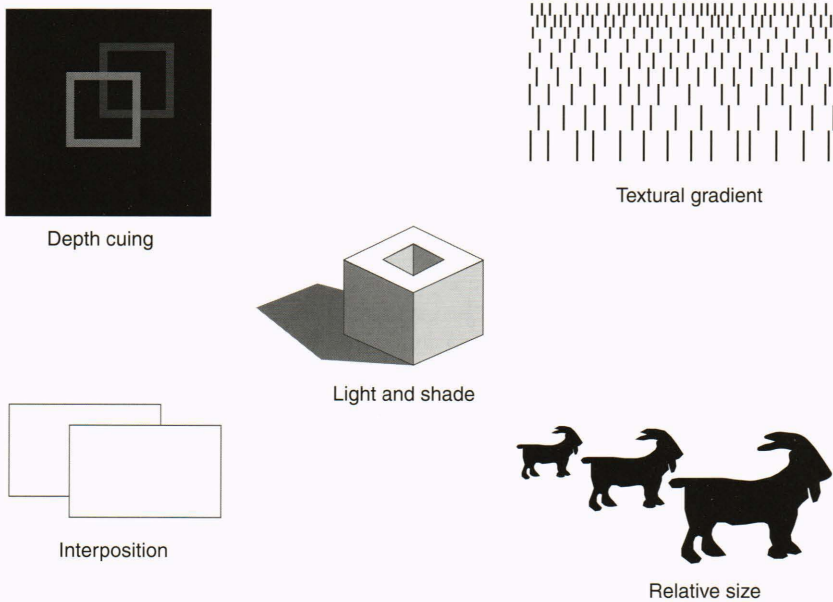**Figure 2.** The four basic types of parallax.

Depth cuing

Textural gradient

Light and shade

Figure 3. Examples of monocular depth cues.

Interposition

Relative size

that are perceived as truly three-dimensional. These cues include perspective, light and shade, relative size, interposition, textural gradient, aerial perspective, and motion parallax. Images rich in monocular depth cues will be even easier to visualize when the stereoscopic cue is added.

Perspective is the relationship between foreground and background objects and is the most important extrastereoscopic depth cue for our purposes, since it scales the stereoscopic depth. If perspective is exaggerated or if perspective cues such as lines recede to a vanishing point, the depth of the image will be enhanced.

Light and shade provide a basic depth cue. Objects can be made to appear solid or round by shading them. Cast shadows can make an object appear to be resting on a surface. Vivid objects appear to be nearer than dim ones, and bright-colored objects look closer than dark-colored ones.

Relative size involves the size of the image of an object projected by the lens of the eye onto the retina. We know that objects appear larger when they are closer and smaller when farther away. Memory helps us to judge the distance of familiar objects. A person seen at some great distance is interpreted to be far away rather than small.

Interposition is seemingly so obvious that we often take it for granted. You perceive that the journal you are now reading is closer to you or in front of whatever is behind it (e.g., your desk) because you cannot see through the journal. It is interposed between you and objects that are farther away.

Aerial perspective is the diminution in visibility of distant objects. For example, distant vistas will often have a bluish haze because the intervening atmosphere scatters red light.

Textural gradient provides a depth cue because the texture is more apparent when the object is closer to the observer. A grassy lawn or the tweed of a jacket is an example of this cue.

Motion parallax is a depth cue that can be provided in the visual world or in moving-image displays and can be achieved, for instance, through the rotation of an object. A familiar example of this cue is seen from a moving car: Everyone has noticed that telephone poles move past more rapidly than the distant hills.

The most important variable to manipulate for strong off-screen effects is not parallax, but rather the depth cues. It is tempting to use huge parallax values to obtain strong viewer-space effects. A better approach is to stress depth cues, particularly perspective. This can be accomplished by using a wide-angle view (at least 40° horizontal) and by scaling the object to fill the screen.

Screen Surround

Looking at a monitor is like looking through a window. The plane of the display screen is the stereo window, and the vertical and horizontal edges of the screen are called the surround. A stereo image can be thought of as existing in two regions: one within the cathode ray tube of a workstation (screen space) and one between it and the viewer (viewer space). When composing a stereoscopic image, a basic decision must made: where to position the scene or the object with respect to the screen surface or stereo window.

Optimally, the surround should not cut off any portion of an object in viewer space (negative parallax). If the surround cuts off or touches an object in viewer space, as illustrated in Figure 4, many people will experience a conflict of depth cues. Because this anomaly never occurs in the visual world, we lack the vocabulary to describe it, so people typically call the effect "blurry," or "out of focus." For many, this conflict of cues will result in the image being pulled back into screen space, even though it has negative parallax.

For example, one cue, the cue of interposition, tells the viewer the object must be behind the window since it is cut off by the surround. The other cue, the stereoscopic
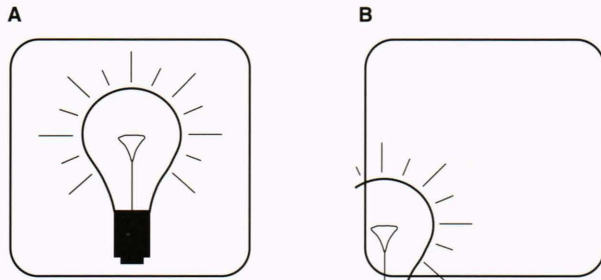
A **B**

Figure 4. Interposition and parallax cues for negative parallax displays. **A.** No conflict of cues; the bulb is centered within and does not touch the screen surround. **B.** Conflict of cues; with the bulb cut off by the surround, interposition tells the eye–brain that the bulb is behind the surround, but stereopsis says it is in front of the surround.

cue provided by parallax, tells the viewer that the object is in front of the window. Experience tells us that when looking through a window, objects cannot be between the window and us. One can accept the stereo illusion only if the object is not occluded by a portion of the window.

Objects touching the horizontal portions of the surround (top and bottom) are less troublesome than objects touching the vertical portions. Objects can be hung from or cut off by the horizontal portions and still look all right.

However, the vertical portion of the surround creates problems due to the paradoxical stereoscopic window effect. When looking outdoors through a real window— for example, out the left edge of the window—the right eye sees more of the outdoor image at the edge than the left eye sees. The same kind of experience occurs for the right edge of the window. The difficulty arises in stereoscopic imaging because the right eye, when looking at the left edge of the screen for objects with negative parallax, will see less image, and the left eye will see more. This phenomenon is at odds with what we see when we look through a real window. The departure from our usual experience may cause the exacerbated difficulty associated with the perception of images at the vertical surrounds.

Viewing stereoscopic images on a large screen can reduce or eliminate the conflict of cues associated with the surround. The edges of the screen may be difficult to see because they lie on the periphery of vision. Thus, the conflict of cues occurring at the surround will not be noticed. This is one benefit of watching stereo movies projected in 70 mm using the Disney, United Artists, or IMAX systems instead of a workstation screen. In this regard, people at workstations have a more difficult task than the cinematographers who are working for projection on enormous screens.

CREATING STEREOSCOPIC DISPLAYS

Software

Implementing a stereoscopic image involves drawing two monocular viewpoints from different positions. In the past, people have created computer-generated stereo-

scopic pairs by rotating an object through a few degrees, but this practice is not advised. The use of rotation to generate stereoscopic pairs will result in vertical misalignment of corresponding left and right image points. Such a vertical misalignment causes the eye muscles to work in an extraordinary way. Most people will experience discomfort fusing the image. If the eyes try to fuse vertical parallax, which contains no depth information, the result can be painful. The approach described here uses horizontal perspective shifts without rotation to produce the proper stereoscopic images.

To be consistent with most programming environments, a "right-handed" coordinate system is used, with the positive z-axis pointing out from the screen. Keep in mind that the positive nature of the z-axis may create confusion because image points along this portion of the axis are defined to have negative parallax. Assuming that the screen lies in the xy plane, and that the eye point lies on the positive z-axis at $(0, 0, d)$, an arbitrary point (x, y, z) will be plotted at

$$\frac{xd}{d-z}, \frac{yd}{d-z}$$

provided that the point is in front of rather than behind the eye. When the eye is moved a distance t to the right, the arbitrary point (x, y, z) will be plotted at

$$\frac{xd - zt}{d-z}, \frac{yd}{d-z}.$$

Left- and right-eye images are created for stereoscopic viewing by calculating projections from two different eye positions onto a single screen. Rather than offsetting the right eye by t as just shown, both eyes are offset by $t/2$ in opposite directions. The twin centers of projection straddle the negative z-axis. Point (x, y, z) will be plotted at

$$\frac{(xd + zt/2)}{d-z}, \frac{yd}{d-z}$$

for the left eye and at

$$\frac{xd - zt}{d-z}, \frac{yd}{d-z}$$

for the right eye, where t is the eye separation and d is the distance from the eyes to the viewing plane. These last two equations assume that the projection plane is the zero parallax setting xy plane, z_{zps}, and the eye viewpoints straddle the z-axis ($x_{mid} = 0$ and $y_{mid} = 0$). If one accepts that z_{zps}, x_{mid}, and y_{mid} may have any value, the arbitrary point (x, y, z) now plots to

$$\frac{[d(x - x_{mid}) - (z_{zps} - z)(t/2)]}{d + z_{zps} - z} + x_{mid},$$

$$\frac{d(y - y_{mid})}{d + z_{zps} - z} + y_{mid}$$

for the left-eye view, and

$$\frac{[d(x - x_{\mathrm{mid}}) + (z_{\mathrm{zps}} - z)(t/2)]}{d + z_{\mathrm{zps}} - z} + x_{\mathrm{mid}},$$

$$\frac{d(y - y_{\mathrm{mid}})}{d + z_{\mathrm{zps}} - z} + y_{\mathrm{mid}}$$

for the right-eye view. These equations can also be specified as a set of transformation matrices, that is, one for the left eye and one for the right.

Some graphics applications give the user more freedom to manipulate images than others. If the user can change the viewing configuration, the application must make the necessary adjustments to maintain high-quality stereoscopic images with comfortable parallax. For example,

1. If an x or y coordinate translation occurs, the two centers of projection (eyes) must remain centered about a point perpendicular to the center of the screen.

2. If a z coordinate translation occurs, the z coordinate of z_{zps} must be changed.

3. If a scaling operation changes the x, y, or z coordinate ranges, z_{zps} as well as the distance from the eyes to z_{zps} and the eye separation (t) may need to be changed.

Cursors have the same benefits they have in the planar displays, but for a stereoscopic display they can also locate the z-axis with great accuracy. A cursor appearing in the left and right images in the same location will have zero parallax, but one must also be able to move the cursor in z space. By moving one or both cursors in the left and right subfields in the horizontal direction (the x direction), the cursor will appear to be moving in two-dimensional space. By counting pixels in the x plane, the exact location of the cursor can be determined. The depth sense of binocular stereopsis is very sensitive. A skilled operator can usually place a cursor to an accuracy of ±1 pixel.

Hardware

Electronic stereoscopic hardware provides a means for passing the appropriate image to the appropriate eye and blocking the inappropriate image from the inappropriate eye to produce stereopsis. Two methods have been developed for achieving this: passive techniques using polarizing glasses, and active techniques in which the viewer uses eyewear incorporating a liquid-crystal shutter to separate images shown on one display screen or incorporating two displays (one for each eye) separated by a septum. The two-display technique is usually found in head-mounted stereoscopic devices where separate, but synchronized, display images are generated for each eye to produce stereopsis.

In the passive approach the viewer wears passive polarizing glasses to see projected polarized images on a screen. Either a large liquid-crystal panel is used to change the polarization characteristics of light emitted from a single projector, or two synchronized polarized projectors are used to display right- and left-eye images. The combination of the polarized images and the passive polarizing glasses forms a shutter sending the correct images to each eye. This technique works well for movie audiences and for applications such as flight simulation.

In the active liquid-crystal shutter approach, the left and right images are rapidly alternated on a workstation or personal computer screen, and a selection device is used to separate the images so that each eye sees only its perspective view. Since the images are separated in time, shutters are used to open and close in sequence with the images. Flickerless electrostereoscopic field-sequential displays operate at twice the rate of planar systems. As the better flickerless planar systems run at 60 fields per second for computer graphics, a flickerless stereoscopic image requires 120 fields per second. The selection device shutters (opens and closes) so that the left lens is open when the left image is written and the right lens is closed, and so on. Each eye sees its perspective view and not the other.

To double the number of fields per second displayed on a monitor without increasing the bandwidth required of the computer or monitor hardware, the video buffer is divided in two—top and bottom—each half containing a left and right view that are squeezed in the vertical direction by a factor of about 2. This effectively doubles the vertical scanning rate, producing 120 instead of the usual 60 fields per second. The number of lines per field is halved. When displayed on a 60-field-per-second monitor, each image will appear to be anamorphically compressed in the vertical direction. Left and right subfields, above and below, will be seen on the unmodified monitor. However, when displayed on a 120-field-per-second monitor, the subfields will be written in sequence.

Many modern computer graphics stereo-ready monitors will automatically switch between 60 and 120 fields per second and adjust the height of the image to restore the aspect ratio. These monitors require a stereo-ready graphics board to generate a vertical synchronization pulse between each field.

Cross Talk

In a perfect stereoscopic system, each eye sees only its assigned image. With cross talk the eye receives unwanted images intended for the other eye. There are two opportunities for cross talk in the electronic stereoscopic display of a workstation: departures from the ideal shutter in the eyewear, and phosphor afterglow on the monitor. Liquid-crystal shutters are good enough to prevent the perception of unwanted images due to incomplete shutter occlusion.

In an ideal stereoscopic display, the image of each field, composed of glowing phosphors, would vanish before the next field was written, but that does not happen. After the right image is written, it will persist while the left image is being written. Thus, an unwanted fading right image will persist into the left image (and vice versa). Perceived cross talk varies with the brightness of the image, color, parallax, and image contrast. Images with large values of parallax will have more perceived cross talk than images with low parallax. High-contrast images, like black lines on a white background, will show the most cross talk. Given the present state of the art for

monitor display tubes, the green phosphor has the longest afterglow and produces the most "ghosting," which can be reduced by reducing the green in the image.

SUMMARY

All computer-generated images produced from a three-dimensional database require the computation of an image from a single perspective or location in space. As stereoscopic images differ from three-dimensional images only in that the image must be generated from two locations, any three-dimensional database or software package can be augmented by stereoscopic displays. The composition and use of depth cues in a scene are the same for both three-dimensional "flat" images and stereoscopic images.

Pleasing computer-generated stereoscopic images can always be produced if sufficient care is taken.

REFERENCE

[1] Wheatstone, C., "On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision (Part the First)," *Philos. Trans. R. Soc. London*, 371-394 (1838).

THE AUTHOR

QUENTIN E. DOLECEK received a B.S. degree in electronics from the University of Maryland in 1963. After service in the Navy, he received an M.S. in acoustics and a D.Sc. in electronics from the Catholic University of America in 1970 and 1980, respectively. Dr. Dolecek began his professional career in 1958 at the Naval Ship Research and Development Center. He joined the Strategic Systems Department at APL in 1980, where he is a specialist in signal processing, systems analysis, and computer graphics. He participated in the very high speed integrated circuit (VHSIC) program, both in system design and VHSIC insertion activities. He was principal investigator for high-speed sonar processing algorithms, the memory-linked wavefont array processor, and the personal visualization system. Inventor of the last two, Dr. Dolecek led APL's development of a prototype and transfer of the technology to industry.