CHAPTER 8

Space Perception and the Display of Data in Space

We live in a three-dimensional world (actually, four dimensions if time is included). In the short history of visualization research, most graphical display methods have required that data be plotted on sheets of paper, but computers have evolved to the point that this is no longer necessary. Now we can create the illusion of 3D space behind the monitor screen, changing over time if we desire. The big question is why we should do this. There are clear advantages to conventional 2D techniques, such as the bar chart and the scatter plot. Designers already know how to draw diagrams and represent data effectively in two dimensions, and the results can easily be included in books and reports. Of course, one compelling reason for an interest in 3D space perception is the explosive advance in 3D computer graphics. Because it is so inexpensive to display data in an interactive 3D virtual space, people are doing it—often for the wrong reasons. It is inevitable that there is now an abundance of ill conceived 3D design, just as the advent of desktop publishing brought poor use of typography and the advent of cheap color brought ineffective and often garish use of color. Through an understanding of space perception, we hope to reduce the amount of poor 3D design and clarify those instances in which 3D representation is really useful.

The first half of this chapter presents an overview of the different factors involved in the perception of 3D space. The second half gives a task-based analysis of the ways in which different kinds of spatial information are used in performing seven different tasks, ranging from tracing paths in 3D networks to judging the morphology of surfaces to appreciating an aesthetic impression of spaciousness. The way we use spatial information differs greatly, depending on the task at hand. Docking one object with another and trying to trace a path in a tangled web of virtual wires require different ways of seeing.

Depth Cue Theory

The visual world provides many different sources of information about 3D space. These sources are usually called *depth cues,* and a large body of research is related to the way the visual system processes depth-cue information to provide an accurate perception of space. Following is a list of the more important depth cues. They are divided into categories according to whether they can be reproduced in a static picture (*monocular static*) or a moving picture (*monocular dynamic*) or require two eyes (*binocular*).

Monocular Static (Pictorial):

- Linear perspective
- Texture gradient
- Size gradient
- Occlusion
- Depth of focus
- Cast shadows
- Shape-from-shading
- Depth-from-eye accommodation (this is nonpictorial)

Monocular Dynamic (Moving Picture):

- Structure-from-motion (kinetic depth, motion parallax)

Binocular:

- Eye convergence
- Stereoscopic depth

Shape-from-shading information has already been discussed in Chapter 7. The other cues are discussed in this chapter. More attention is devoted to stereoscopic depth perception than to the other depth cues, not because it is the most important, but because it is relatively complex and because it is difficult to use stereoscopic depth effectively.

Perspective Cues

Figure 8.1 shows how perspective geometry can be described for a particular viewpoint and a picture plane. The position of each feature on the picture plane is determined by extending a ray from the viewpoint to that feature in the environment. If the resulting picture is subsequently scaled up or down, the correct viewpoint is specified by similar triangles, as shown. If the eye is placed at the specified point with respect to the picture, the result is a correct perspective view of the scene. A number of the depth cues are direct results of the geometry of perspective. These are illustrated in Figures 8.2 and 8.3.

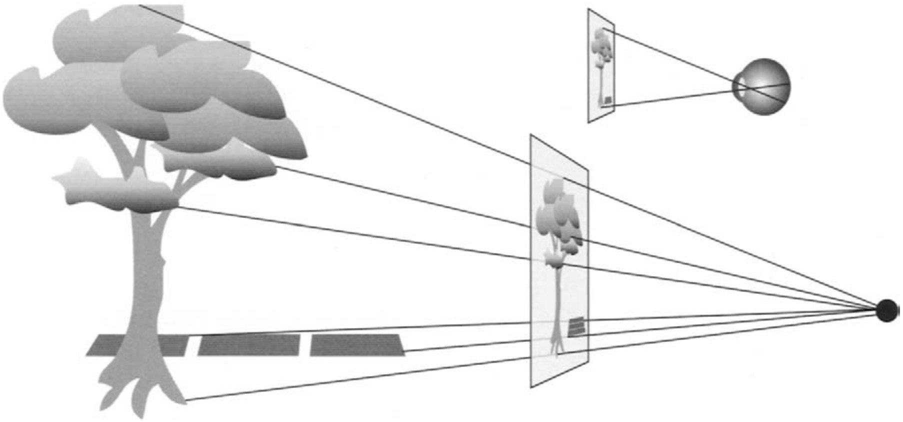


Figure 8.1 The geometry of linear perspective is obtained by sending a ray from each point in the environment through a picture plane to a single fixed point. Each point on the picture plane is colored according to the light that emanates from the corresponding region of the environment. The result is that objects vary in size on the picture plane in inverse proportion to their distance from the fixed point. If an image is created according to this principle, the correct viewpoint is determined by similar triangles, as shown in the upper right.

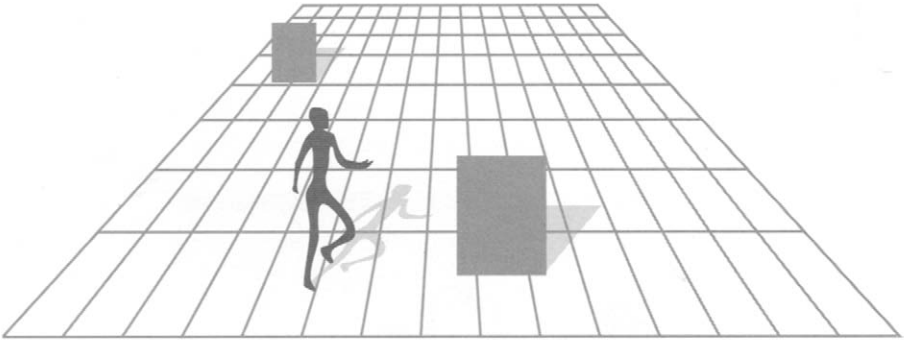


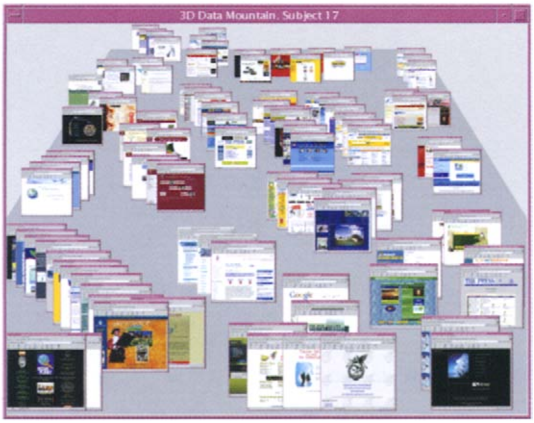
Figure 8.2 Perspective cues arising from perspective geometry include the convergence of lines and the fact that more distant objects become smaller on the picture plane.

Figure 8.3 A texture gradient is produced when a uniformly textured surface is projected onto the picture plane.

- Parallel lines converge to a single point.
- Objects at a distance appear smaller on the picture plane than do nearby objects. Objects of known size may have a very powerful role in determining the perceived size of adjacent unknown objects. Thus, an image of a person placed in a picture of otherwise abstract objects gives a scale to the entire scene.
- Uniformly textured surfaces result in texture gradients in which the texture elements become smaller with distance.

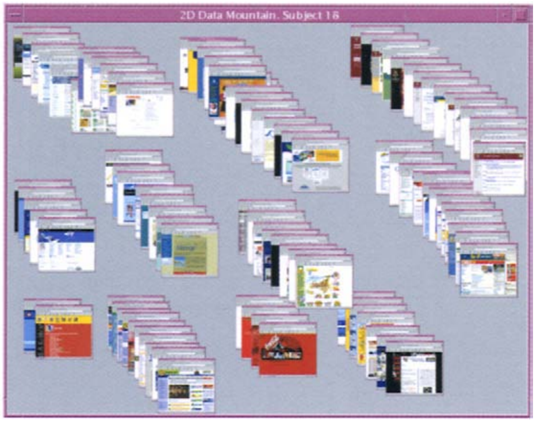
In the real world, we generally perceive the actual size of an object rather than the size at which it appears on a picture plane (or on the retina). This phenomenon is called *size constancy.* The degree to which size constancy is obtained is a useful measure of the relative effectiveness of depth cues. However, when we perceive pictures of objects, we enter a kind of dual perception mode. To some extent, we have a choice between accurately perceiving the size of the depicted object as though it were in a 3D space and accurately perceiving the size of the object at the picture plane (Hagen, 1974). The amount and effectiveness of the depth cues used will, to some extent, make it easy to see in one mode or the other. The picture-plane sizes of objects in a very sketchy schematic picture are easy to perceive. Conversely, the real 3D sizes of objects will be more readily perceived with a highly realistic moving picture, although large errors will be made in estimating picture-plane sizes.

In terms of the total amount of information available from an information display, there is little evidence that a perspective picture lets us see more than a nonperspective image. A study by Cockburn and McKenzie showed that perspective cues added no advantage to a version of the Data Mountain display of Robertson et al. (1998). The version shown in Figure 8.4(b) was just as effective as the one in Figure 8.4(a). However, both of these versions make extensive use of other depth cues (occlusion and height on the picture plane).



a

Figure 8.4 (a) Variations on Robertson et al.'s (1998) Data Mountain Display. *Courtesy of Andy Cockburn.*



b

Figure 8.4 *Continued* (b) Same as part (a) but without perspective.

Pictures Seen from the Wrong Viewpoint

It is obvious that most pictures are not viewed from their correct centers of perspective. In a movie theater, only one person can occupy this optimal viewpoint (determined by the focal length of the original camera and the scale of the final picture). When a picture is viewed from an incorrect viewpoint, the laws of geometry suggest that significant distortions should occur. Figure 8.5 illustrates this. If the mesh shown in Figure 8.5 is projected on a screen with a geometry based on viewpoint (a), but it is actually viewed from position (b), it should be perceived to stretch along the line of sight as shown (if the visual system were a simple geometry processor). However, although people report seeing some distortion initially when looking at moving pictures from the wrong viewpoint, they become unaware of the distortion after a few minutes. Kubovy (1986) calls this the *robustness of linear perspective*. Apparently, the human visual system overrides some aspects of perspective in constructing the 3D world that we perceive.

One of the mechanisms that can account for this lack of perceived distortion may be based on a built-in perceptual assumption that objects in the world are rigid. Suppose that the mesh in Figure 8.5 is smoothly rotated about a vertical axis, projected assuming viewpoint (a) but viewed from point (b). It should appear as a nonrigid, elastic body. But perceptual processing is constrained by a rigidity assumption, and this causes us to see a stable, nonelastic three-dimensional object.

Under extreme conditions, some distortion is still seen with off-axis viewing of moving pictures. Hagen and Elliott (1976) showed that this residual distortion is reduced if the projective geometry is made more parallel. This can be done by simulating long–focal length lenses, which may be a useful technique if displays are intended for off-axis viewing.

Various technologies exist that can track a user's head position with respect to a computer screen and thereby estimate the position of the eye(s). With this information, a 3D scene can be

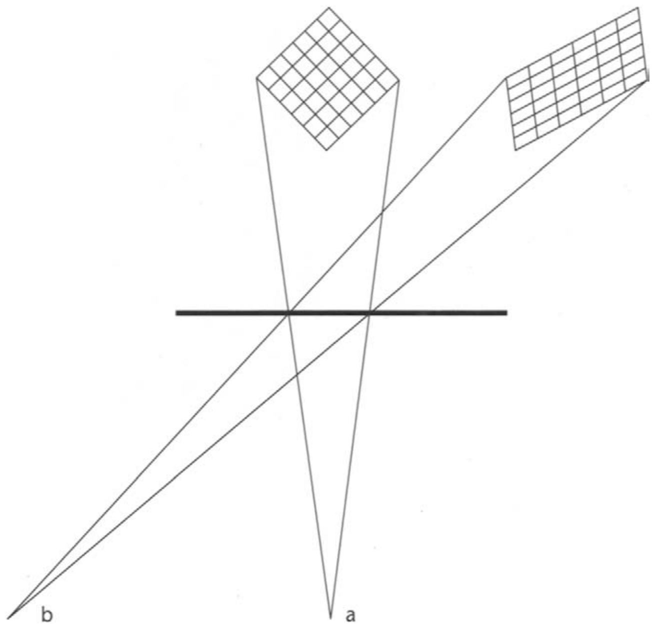


Figure 8.5 When a perspective picture is seen from the wrong viewpoint (point b), simple geometry predicts that large distortions should be seen. However, they are generally not seen or, when seen, are minimal.

computed and viewed so that the perspective is "correct" at all times by adjusting the viewpoint parameters in the computer graphics software (Deering, 1992; Ware et al., 1993). There are two reasons why this might be desirable, despite the fact that incorrect perspective viewing of a picture seems generally unimportant. The first reason is that as an observer changes position, the perspective image will change accordingly, resulting in motion parallax. Motion parallax is itself a depth cue, as discussed later in the structure-from-motion section. The second reason is that in some virtual-reality systems, it is possible to place the subject's hand in the same space as the virtual computer graphics imagery. When we make visually guided hand movements toward some object in the world, we are constantly correcting our movements based on visual feedback. If this were done using computer graphics imagery to represent a virtual object and a virtual image of the subject's hand, head-coupled perspective could be necessary to keep the subject's body sense (kinesthetic feedback) of hand position aligned with his or her visual feedback. An example of an experimental setup is shown in Figure 8.6. However, research has shown that as long as continuous visual feedback is provided, *without excessive lag*, people can adjust rapidly to simple changes in the eye–hand relationship (Held et al., 1966). The effects of lag on performance are discussed further in Chapter 10.

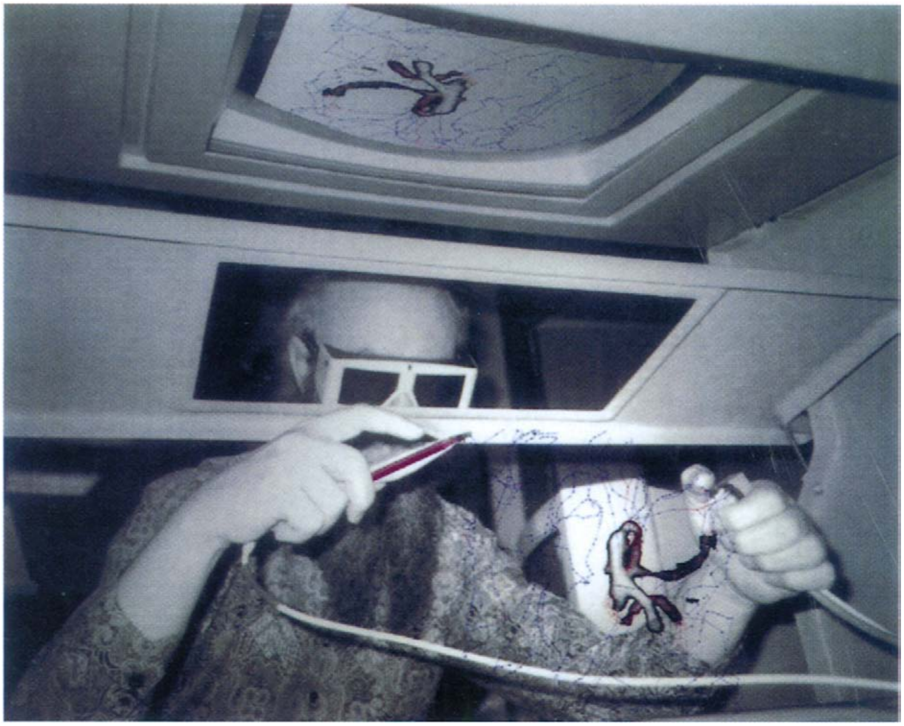


Figure 8.6 A user is attempting to trace 3D blood vessels in an interface that puts his hand in the same space as the virtual computer graphics imagery. *From Serra et al., 1997.*

When virtual-reality head-mounted displays are used, it is essential that the perspective be coupled to a user's head movement, because the whole point is to allow users to change viewpoint in a natural way. Experimental evidence supports the idea that head-coupled perspective enhances the sense of presence in virtual spaces more than stereoscopic viewing (Arthur et al., 1993; Pausch et al., 1996).

Occlusion

If one object overlaps or occludes another, it appears closer to the observer. See Figure 8.7. This is probably the strongest depth cue, but it provides only binary information. An object is either behind or in front of another; no information is given about the distance between them. A kind of partial occlusion occurs when one object is transparent or translucent. In this case, there is a color difference between the parts of an object that lie behind the transparent plane and the parts that are in front of it. This can be useful in positioning one object inside another (Zhai et al., 1994).

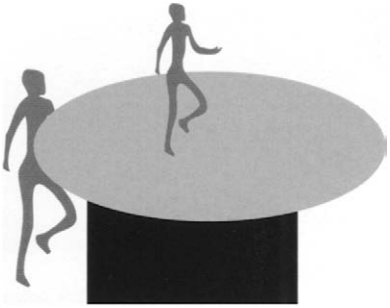


Figure 8.7 An object that occludes another appears closer to the viewer.

Occlusion can be useful in design; for example, the tabbed cards illustrated in Figure 8.8(a) use occlusion to provide rank-order information, in addition to rapid access to individual cards. Although modern graphical user interfaces (GUIs) are usually described as being 2D, they are actually 3D in a nontrivial way. Overlapping windows rely on our understanding of occlusion to be effective. See Figure 8.8(b).

Depth of Focus

When we look around, our eyes change focus to bring the images of fixated objects into sharp focus on the fovea. As a result, the images of both nearby and more distant objects become blurred. The equations that determine depth of focus are presented in Chapter 2. Focus effects are important in separating foreground objects from background objects, as shown in Figure 8.9. Perhaps because of its role as a depth cue, simulating depth of focus is an excellent way to highlight information by blurring everything except that which is critical. Unfortunately, the technique is computationally expensive and thus currently limited in utility.

Focus can be considered a pictorial depth cue only if the object of fixation can be predicted. In normal vision, our attention shifts and our eyes refocus dynamically depending on the distance of the object fixated. Chapter 2 describes a system designed to change focus information based on measured point of fixation in a virtual environment.

Cast Shadows

Cast shadows are a very potent cue to the height of an object above a plane, as illustrated in Figure 8.10(a). They can function as a kind of indirect depth cue—the shadow locates the object with respect to some surface in the environment. In the case of Figure 8.10, this surface is not present in the illustration but is assumed by the brain. In a multifactor experiment, Wanger et al. (1992) found that shadows provided the strongest "depth" cue when compared to texture, projection type, frames of reference, and motion. But it should be noted that they used a checkerboard as a base plane to provide the actual distance information. Cast shadows function

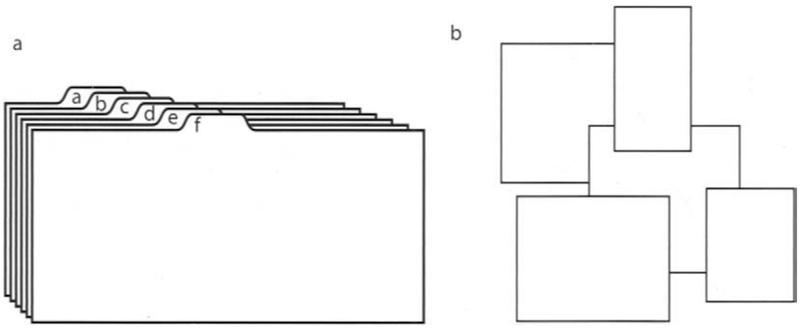


Figure 8.8 (a) Careful use of occlusion enables small tabs to provide access to larger objects. (b) Window interfaces use occlusion.

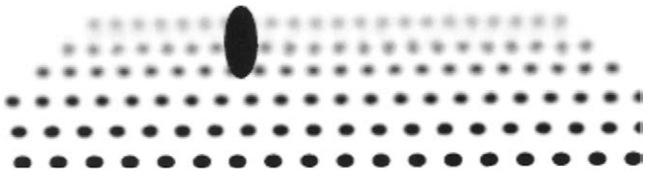


Figure 8.9 The eye adjusts to bring objects of interest into sharp focus. As a result, objects at different distances become blurred.

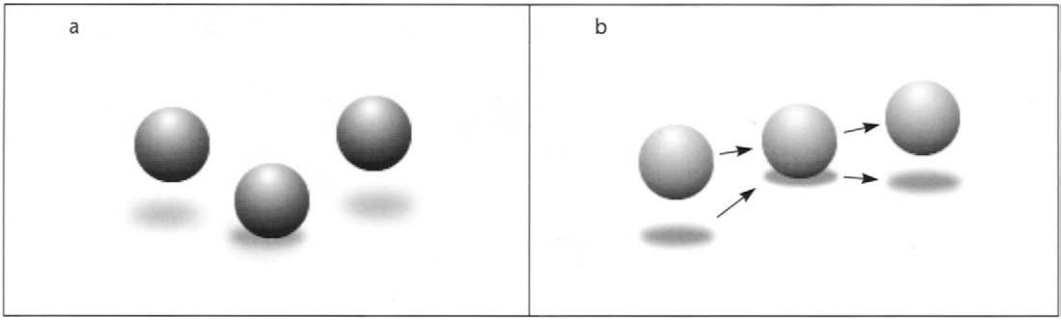


Figure 8.10 (a) Shadows can provide a strong cue for the relative height of objects above a plane. (b) The effect becomes even stronger with motion. The ball actually appears to bounce when the ball and shadow are animated to follow the trajectories shown.

best as a height-above-surface cue when there is a relatively small distance between the object and the surface, and they can be especially effective in showing when an object is very close to the point of contact (Madison et al., 2001).

Cast shadows are useful in distinguishing information that is layered a small distance above a planar surface, as illustrated in Figure 8.11. This technique can be applied to layered map displays of the type used in geographical information systems (GISs). In complex environments, where objects are arranged throughout 3D space, cast shadows can be confusing rather than helpful, because it may not be possible to determine perceptually which object cast a particular shadow.

Kersten et al. (1997) showed that cast shadows are especially powerful when objects are in motion. One of their demonstrations is illustrated in Figure 8.10(b). In this case, the *apparent* trajectory of a ball moving in 3D space is caused to change dramatically depending on the path of the object's *shadow*. The image of the ball actually travels in a straight line, but the ball appears to bounce because of the way the shadow moves. In this study, shadow motion was shown to be a stronger depth cue than change in size with perspective.

It seems likely that shadows can be correctly interpreted without being realistic. Kersten et al. (1996) found no effect of shadow quality in their results. However, one of the principal cues in distinguishing shadows from nonshadows in the environment is the lack of sharpness in shadow edges. Fuzzy shadows are likely to lead to less confusing images.

Shape-from-Shading

See Chapter 7 for a discussion of the perception of surface shape-from-shading information. We can add one more point here. Shading information can be useful in emphasizing the affordances of display widgets such as buttons and sliders, even in displays that are very flat. Figure 8.12 illustrates a slider enhanced with shading. This technique is widely used in today's GUIs.

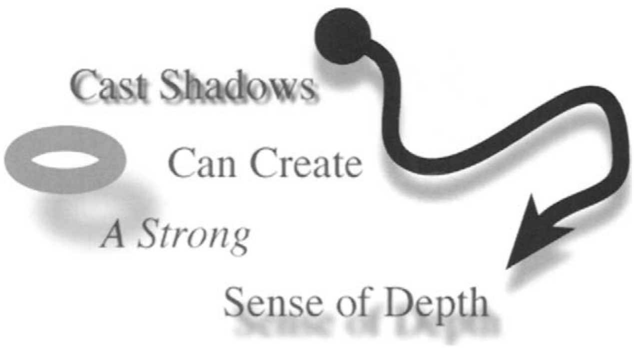


Figure 8.11 Cast shadows can be useful in making data appear to stand out above an opaque plane.

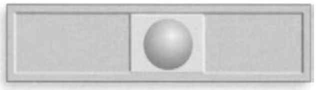


Figure 8.12 Even with mostly 2D interfaces, subtle shading can make sliders and other widgets look like objects that can be manipulated.

Eye Accommodation

The eye changes focus to bring attended objects into sharp focus on the retina. However, because we are only capable of focusing to one-half of a diopter, this means that accommodation can provide limited information about the distance to nearby objects (Hochberg, 1971). Accommodation does not appear to be used to judge distance directly, but it is used in computing the size of nearby objects (Wallach and Floor, 1971).

Structure-from-Motion

When an object is in motion or when we ourselves move through the environment, the result is a dynamically changing pattern of light on the retina. Structure-from-motion information is generally divided into two different classes: motion parallax and the kinetic depth effect.

An example of *motion parallax* occurs when we look sideways out of a car or train window. Things nearby appear to be moving very rapidly, whereas objects close to the horizon only appear to move gradually. Overall, there is a velocity gradient, as illustrated in Figure 8.13(a). When we move forward through a cluttered environment, the result is a very different expanding pattern of motion, like that shown in Figure 8.13(b). Wann et al. (1995) showed that subjects were able to control their headings with an accuracy of 1 to 2 degrees when they were given feedback from a wide-screen field of dots through which they had to steer. There is also evidence for specialized neural mechanisms sensitive to the time to contact with visual moving targets. These may enable animals to become aware of objects on a collision course (Wang and Frost, 1992).

The *kinetic depth effect* can be demonstrated with a wire bent into a complex 3D shape and projected onto a screen, as shown in Figure 8.13(c). Casting the shadow of the wire will suffice for the projection. The result is a two-dimensional line, but if the wire is rotated, the three-dimensional shape of the wire immediately becomes apparent (Wallach and O'Connell, 1953). The kinetic depth effect dramatically illustrates a key concept in understanding space perception. The brain generally assumes that objects are rigid in 3D space, and the mechanisms of object perception incorporate this constraint. The moving shadow of the rotating bent wire is perceived as a rigid 3D object, not as a wiggling 2D line. It is easy to simulate this in a computer graphics system by creating an irregular line, rotating it about a vertical axis, and displaying it using standard graphics techniques.

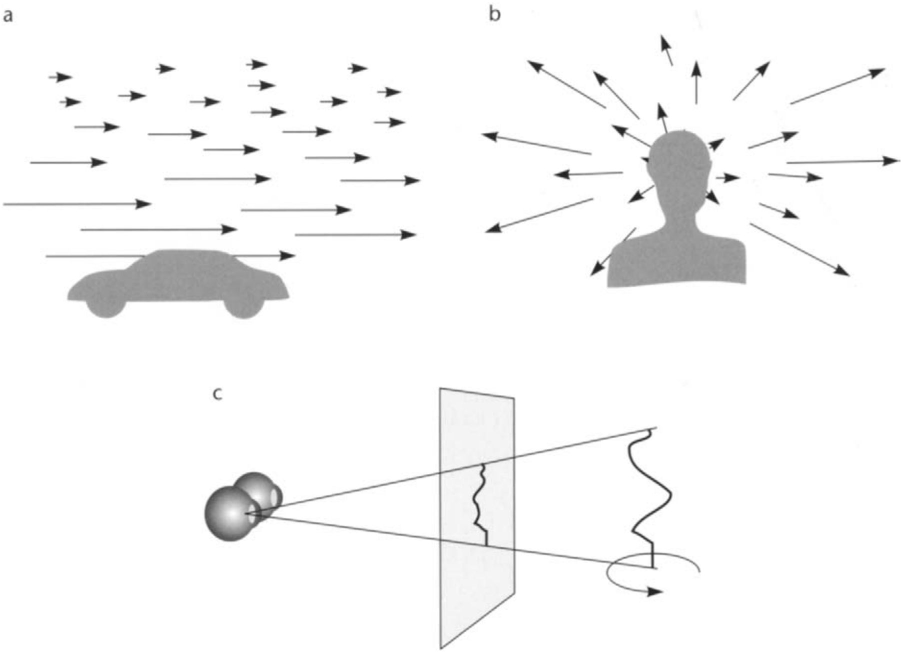


Figure 8.13 Three different kinds of structure-from-motion information. (a) The velocity gradient that results when the viewer is looking sideways out of a moving vehicle. (b) The velocity field that results when the viewer is moving forward through the environment. (c) The kinetic depth information that results when a rotating rigid object is projected onto a screen.

For many tasks, structure-from-motion information is at least as important as stereoscopic depth in providing us with information about the spatial layout of objects in space (Rogers and Graham, 1979). It helps us determine both the 3D shapes of objects and the large-scale layout of objects in space. Structure-from-motion is the reason for the effectiveness of fly-through animated movies that take an observer through a data space.

Eye Convergence

When we fixate an object with both eyes, they must converge to a degree dictated by the distance of the object. This *vergence* angle is illustrated in Figure 8.14. Given the two line-of-sight vectors, it is a matter of simple trigonometry to determine the distance to the fixated object. However, the evidence suggests that the human brain is not good at this geometric computation except for objects within arm's length (Viguier et al., 2001). The vergence sensing system appears capable of quite rapid recalibration in the presence of other spatial information (Fisher and Cuiffreda, 1990).

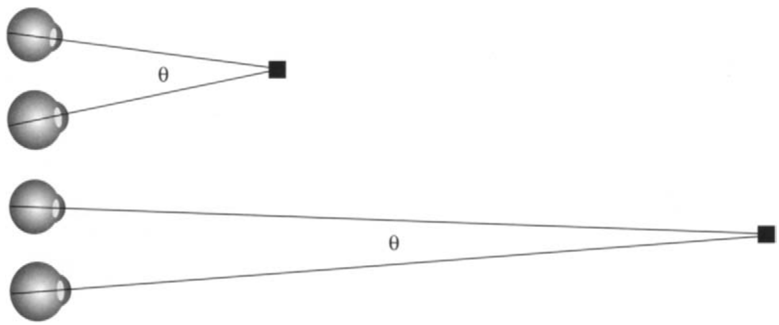


Figure 8.14 The vergence angle θ varies as the eyes fixate on near and far objects.

Stereoscopic Depth

There is an often-expressed opinion that stereoscopic displays allow "truly" three-dimensional images. In advertising literature, potential buyers are urged to buy stereoscopic display equipment and "see it in 3D." As should be plain from this chapter, stereoscopic disparity is only one of many depth cues that the brain uses to analyze 3D space, and it is by no means the most useful one. If fact, as much as 20% of the population may be stereo-blind, yet they function perfectly well and in fact are often unaware that they have a disability. Nevertheless, stereoscopic displays can provide a particularly compelling sense of a three-dimensional virtual space, and for certain tasks they can be extremely useful.

The basis of stereoscopic depth perception is forward-facing eyes with overlapping visual fields. On average, human eyes are separated by about 6.4 centimeters; this means that the brain receives slightly different images, which can be used to compute relative distances of pairs of objects. Stereoscopic depth is a technical subject, and we therefore begin by defining some of the terms.

Figure 8.15 illustrates a simple stereo display. Both eyes are fixated on the vertical line (a for the right eye, c for the left eye). A second line d in the left eye's image is fused with b in the right eye's image. The brain resolves the discrepancy in line spacing by perceiving the lines as being at different depths, as shown.

Angular disparity is the difference between the angular separation of a pair of points imaged by the two eyes (disparity = $\alpha - \beta$). *Screen disparity* is the distance between parts of an image on the screen (screen disparity = $(c - d) - (a - b)$).

If the disparity between the two images becomes too great, double vision, called *diplopia,* occurs. Diplopia is the appearance of the doubling of part of a stereo image when the visual system fails to fuse the images. The 3D area within which objects can be fused and seen without double images is called *Panum's fusional area.* In the worst case, Panum's fusional area has remarkably little depth. At the fovea, the maximum disparity before fusion breaks down is only

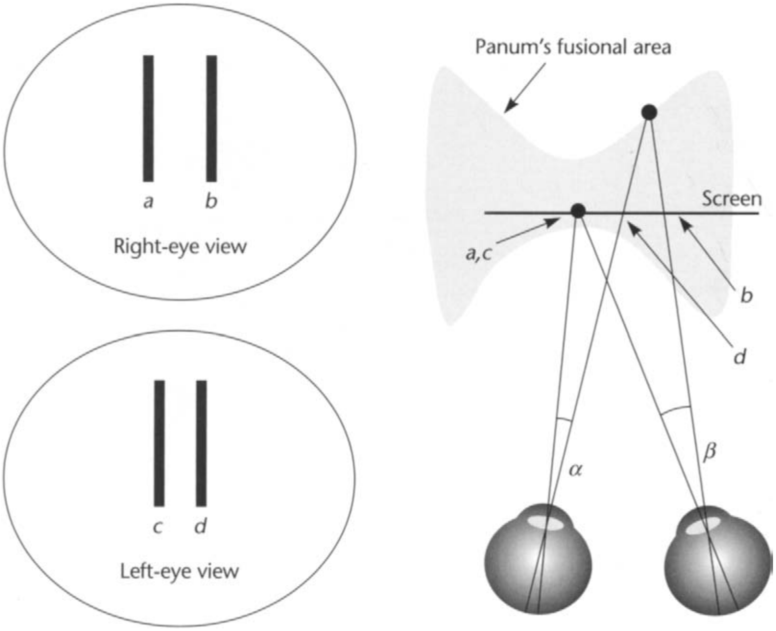


Figure 8.15 A simple stereo display. Different images for the two eyes are shown on the left. On the right, a top-down view shows how the brain interprets this display. The vertical lines *a* and *b* in the right-eye image are perceptually fused with *c* and *d*, respectively, in the left-eye image.

1/10 degree, whereas at 6 degrees eccentricity (of the retinal image from the fovea), the limit is 1/3 degree (Patterson and Martin, 1992).

It is worthwhile to consider what these numbers imply for monitor-based stereo displays. A screen with 30 pixels/cm, viewed at 57 cm, will have 30 pixels per degree of visual angle. The 1/10-degree limit on the visual angle before diplopia occurs translates into about three pixels of screen disparity. This means that we can only display three whole-pixel-depth steps before diplopia occurs, either in front of or behind the screen. It also means that in the worst case, it will only be possible to view a virtual image that extends in depth a fraction of a centimeter from the screen (assuming an object on the screen is fixated). However, it is important to emphasize that this is a worst-case scenario. It is likely that antialiased images will allow better-than-pixel resolution, for exactly the same reason that vernier acuities can be achieved to better-than-pixel resolution (discussed in Chapter 2). In addition, the size of Panum's fusional area is highly dependent on a number of visual display parameters, such as the exposure duration of the images and the size of the targets. Both moving targets and blurred images can be fused at greater disparities, and the fusional area becomes larger, with lateral separation of the image components (Patterson and Martin, 1992). Depth judgments can also be made outside the fusional area, although these are less accurate.

Stereopsis is a superacuity. We can resolve disparities of only 10 seconds of arc at better than chance. This means that we should be able to see a depth difference between an object at 1 kilometer and an object at infinity, under optimal viewing conditions.

Problems with Stereoscopic Displays

It is common for users of 3D visualization systems with stereoscopic display capabilities to disable stereo viewing once the novelty has worn off, and view the data using a monocular perspective. There are a number of reasons that stereoscopic displays are disliked. Double-imaging problems tend to be much worse in stereoscopic computer displays than in normal viewing of the 3D environment. One of the principal reasons for this is that in the real world, objects farther away than the one fixated are out of focus on the retina. Because we can fuse blurred images more easily than sharply focused images, this reduces diplopia problems in the real world. In addition, focus is linked to attention and foveal fixation. In the real world, double images of nonattended peripheral objects generally will not be noticed. Unfortunately, in present-day computer graphics systems, particularly those that allow for real-time interaction, depth of focus is never simulated. All parts of the computer graphics image are therefore equally in focus, even though some parts of the image may have large disparities. Thus, the double images that occur in stereoscopic computer graphics displays are very obtrusive.

Frame Cancellation

Valyus (1966) coined the phrase *frame cancellation* to describe a common problem with stereoscopic displays. If the stereoscopic depth cues are such that a virtual image should appear in front of the screen, the edge of the screen appears to occlude the virtual object, as shown in Figure 8.16. Occlusion overrides the stereo depth information, and the depth effect collapses. This is typically accompanied by a double image of the object that should appear in front.

The Vergence–Focus Problem

When we change our fixation between objects placed at different distances, two things happen: the convergence of the eyes changes (vergence), and the focal lengths of the lenses in the eyes accommodate to bring the new object into focus. The vergence and the focus mechanisms are coupled in the human visual system. If one eye is covered, the vergence and the focus of the *covered* eye change as the uncovered eye accommodates objects at different distances. This illustrates vergence being driven by focus. The converse also occurs: a change in vergence can drive a change in focus.

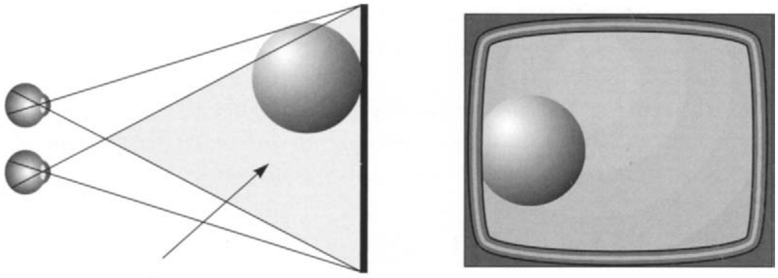


Figure 8.16 Frame cancellation occurs when stereoscopic disparity cues indicate that an object is in front of the monitor screen. Because the edge of the screen clips the object, this acts as an occlusion depth cue and the object appears to be behind the window, canceling the stereo depth effect. Because of this, the usable working volume of a stereoscopic display is restricted as shown.

In a stereoscopic display, all objects lie in the same focal plane, regardless of their apparent depth. However, accurate disparity and vergence information may fool the brain into perceiving them at different depths. Screen-based stereo displays provide disparity and vergence information, but no focus information. The failure to present focus information correctly, coupled with vergence, may cause a form of eyestrain (Wann et al., 1995; Mon-Williams and Wann, 1998). This problem is present in both stereoscopic head-mounted systems and monitor-based stereo displays. Wann et al. concluded that vergence and focus cross-coupling "prevents large depth intervals of three-dimensional visual space being rendered with integrity through dual two-dimensional displays." This may account for the common reports of eyestrain occurring with dynamic stereoscopic displays.

Distant Objects

The problems with stereoscopic viewing are not always related to disparities that are too large. Sometimes disparities may be too small. The stereoscopic depth cue is most useful for 30 meters or less from the viewer. Beyond this, disparities are too small to be resolved. For practical purposes, most useful stereoscopic depth is obtained within distances of less than 10 meters from the viewer and may be optimal for objects held roughly at arm's length.

Making Effective Stereoscopic Displays

Because stereoscopic depth perception is a superacuity, the ideal stereoscopic display should have very high resolution, much higher than the typical desktop monitor. On current monitors, the fine detail is produced by pixels, and in a stereoscopic display the pixelation of features such as fine lines will generate false binocular correspondences. High-resolution displays enable the presentation of fine texture gradients and hence disparity gradients that are the basis for stereoscopic surface shape perception.

There are also ways of mitigating the diplopia, frame cancellation, and vergence–focus problems described previously, although they will not be fully solved until true 3D displays become commercially viable. All the solutions involve reducing screen disparities by artificially bringing the computer graphics imagery into the fusional area. Valyus (1966) found experimentally that the diplopia problems were acceptable if no more than 1.6 degrees of disparity existed in the display. Based on this, he proposed that the screen disparity should be less than 0.03 times the distance to the screen. However, this provides only about ±1.5 cm of useful depth at normal viewing distances. Using a more relaxed criterion, Williams and Parrish (1990) concluded that a practical viewing volume falls between −25% and +60% of the viewer-to-screen distance. This provides a more usable working space.

One obvious solution to the problem of creating useful stereoscopic displays is simply to create small virtual scenes that do not extend much in front of or behind the screen. However, in many situations this is not practical—for example, if we wish to make a stereoscopic view of extensive terrain. A more general solution is to compress the range of stereoscopic disparities so that they lie within a judiciously enlarged fusional area, such as that proposed by Williams and Parrish. A method for doing this is described in the next two sections.

But before going on, we must consider a potential problem. We should be aware that tampering with stereoscopic depth may cause us to misjudge distance. There is conflicting evidence as to whether this is likely. Some studies have shown stereoscopic disparity to be relatively unimportant in making absolute depth judgments. For example, Wallach and Karsh (1963) found that when they rotated a wireframe cube viewed in stereo, only half the subjects they were trying to recruit were even aware of a doubling in their eye separation. Because increasing eye separation increases stereo disparities, this should have resulted in a grossly distorted cube. The fact that distortion was not perceived indicates that kinetic depth-effect information and rigidity assumptions are much stronger than stereo information. Ogle (1962) argued that stereopsis gives us information about the *relative* depths of objects that have small disparities; when it comes to judging the overall layout of objects in space, other depth cues dominate. Yet, under certain circumstances, accurate depth may be made on the basis of stereoscopic disparities (Durgin et al., 1995). More research will be needed before we have a really clear picture of the way stereoscopic depth is combined with other depth information in the brain. Also, many experiments show large individual differences in how we use the different kinds of depth information, so we will never have a simple "one-size-fits-all" account.

Overall, we can conclude that the brain is very flexible in weighing evidence from the different depth cues and that disparity information can be scaled by the brain depending on other available information. Thus, it should be possible to manipulate artificially the overall pattern of stereo disparities and enhance local 3D space perception without distorting the overall sense of space if other strong cues to depth, such as linear perspective, are provided. We (Ware et al., 1998) investigated dynamically changed disparities by smoothly varying the stereoscopic eye separation parameter. We found that a subject's disparity range could be changed by about 30% over two seconds, without them even noticing, as long as the change was smooth.

Cyclopean Scale

One simple method that we have developed to deal with diplopia problems is called a *cyclopean scale* (Ware et al., 1998). As illustrated in Figure 8.17, this manipulation involves scaling the virtual environment about the midpoint between the observer's estimated eye positions. The scaling variable is chosen so that the nearest part of the scene comes to a point just behind the monitor screen. To understand the effects of this operation, it is worthwhile to consider first that scaling a virtual world about a single viewpoint does not result in any change in computer graphics imagery (assuming depth of focus is not taken into account). Thus, the cyclopean scale does not change the overall sizes of objects as they are represented on a computer screen. The cyclopean scale has a number of benefits for stereo viewing:

- More distant objects, which would normally not benefit from stereo viewing because they are beyond the range where significant disparities exist, are brought into a position where usable disparities are present.
- The vergence–focus discrepancy is reduced. At least for the part of the virtual object that lies close to the screen, there is no vergence—focus conflict.
- Virtual objects that are closer to the observer than to the screen are also scaled so that they lie behind the screen. This removes the possibility of frame cancellation.

Virtual Eye Separation

The cyclopean scale, although useful, does not remove the possibility of disparities that result in diplopia. In order to do so, it is necessary to compress or expand the disparity range. To understand how this can be accomplished, it is useful to consider a device called a *telestereoscope*. This

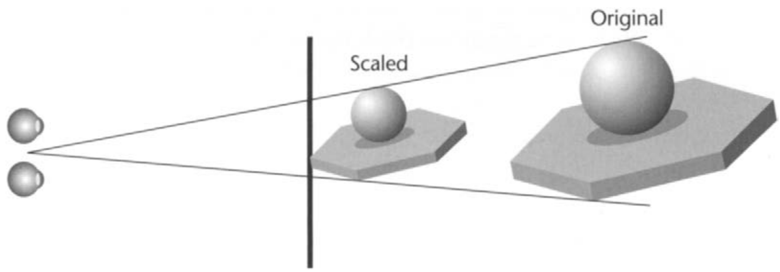


Figure 8.17 Cyclopean scale: A virtual environment is resized about a center point midway between the left and right viewpoints.

uses a system of mirrors to increase the effective separation of the eyes, as shown in Figure 8.18. A telestereoscope is generally used to increase disparities when distant objects are viewed. However, the same principle can also be used to decrease the range of disparities by optically moving the eyes closer together.

Figure 8.19 illustrates the concept of virtual eye separation and demonstrates how the apparent depth of an object decreases if the virtual viewpoints have a wider eye separation than the actual viewpoint. We consider only a single point in the virtual space. If E_v is the virtual eye separation and E_a is the actual eye separation of an observer, the relationship between depth in the virtual image (z_v) and in the viewed stereo image (z_s) is a ratio:

$$\frac{E_v}{E_a} = \frac{z_S(z_v + z_e)}{z_v(z_S + z_e)} \tag{8.1}$$

where z_e represents the distance to the screen. By rearranging terms, we can get the stereo depth expressed as a function of the virtual depth and the virtual eye separation.

$$z_S = \frac{z_e z_v E_v}{E_a Z_v + E_a z_e - E_v z_v} \tag{8.2}$$

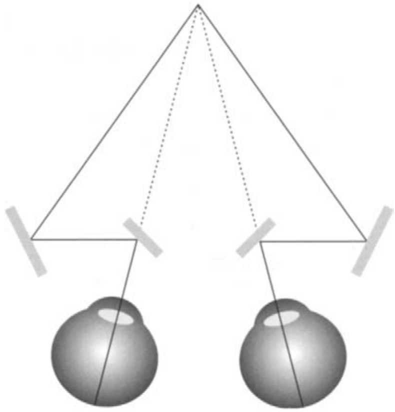


Figure 8.18 A telestereoscope is a device that increases the effective eye separation, thereby increasing stereoscopic depth information (disparities).

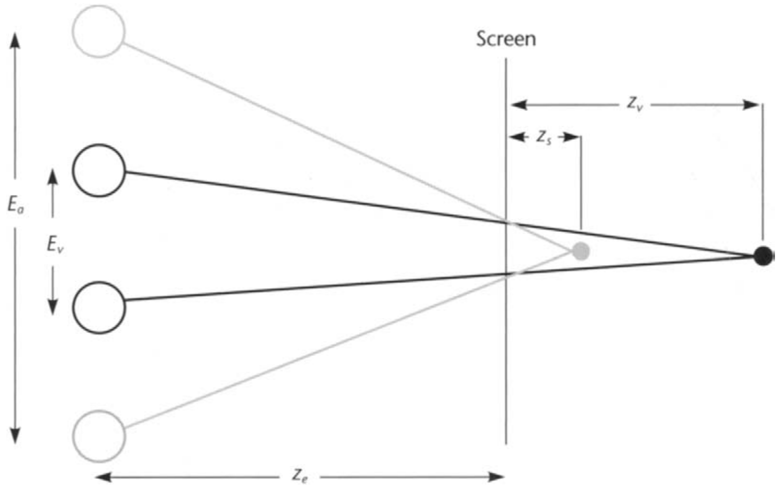


Figure 8.19 The geometry of virtual eye separation. In this example, the stereoscopic depth is decreased by computing an image with a virtual eye separation that is smaller than the actual eye separation. Stereoscopic depth can just as easily be increased.

If the virtual eye separation is smaller than the actual eye separation, stereo depth is decreased. If the virtual eye separation is larger than the actual eye separation, stereo depth is increased. $E_v = E_a$ for "correct" stereoscopic viewing of a virtual scene, although for the reasons stated, this may not be useful in practice. When $E_v = 0.0$, both eyes get the same image, as in single-viewpoint graphics. Note that stereo depth and perceived depth are not always equal. The brain is an imperfect processor of stereo information, and other depth cues may be much more important in determining the perceived depth.

Experimental evidence shows that subjects given control of their eye-separation parameters have no idea of what the "correct" setting should be (Ware et al., 1998). When asked to adjust the virtual eye-separation parameter, subjects tended to decrease the eye separation for scenes in which there was a lot of depth, but actually increased eye separation beyond the normal (enhancing the sensation of stereoscopic depth) when the scene was flat. This behavior can be mimicked by an algorithm designed to test automatically the depth range in a virtual environment and adjust the eye-separation parameters appropriately (after cyclopean scale). We have found the following function to work well for a large variety of digital terrain models. It uses the ratio of the nearest point to the farthest point in the scene to calculate the virtual eye separation in centimeters.

$$\text{EyeSeparation} = 2.5 + 5.0 * (\text{NearPoint}/\text{FarPoint})^2 \tag{8.3}$$

This function increases the eye separation to 7.5 cm for shallow scenes (as compared to a normal value of 6.4 cm) and reduces it to 2.5 cm for very deep scenes.

Artificial Spatial Cues

There are effective ways to provide information about space that are not based directly on the way information is provided in the normal environment, although the best are probably effective because they make use of existing perceptual mechanisms.One common technique that is used to enhance 3D scatter plots is illustrated in Figure 8.20. A line is dropped from each data point to the ground plane. Without these lines, only a 2D judgment of spatial layout is possible. With the lines, it is possible to estimate 3D position. Kim et al. (1991) showed that this artificial spatial cue can be at least as effective as stereopsis in providing 3D position information.

It should be understood that although the vertical line segments in Figure 8.20 can be considered artificial additions to the plot, there is nothing artificial about the way they operate as depth cues. Gibson (1986) pointed out that one of the most effective ways to estimate the sizes of objects is with reference to the ground plane. Adding the vertical lines creates a link to the ground plane and the rich texture size and linear perspective cues embedded in it. They function in the same way as cast shadows, only they are generally easier to interpret, given that cast shadows can be confusing with certain lighting directions.

Computer graphics systems sometimes provide a facility for what vision researchers call *proximity luminance covariance* (Dosher et al., 1986), which is simply (but rather confusingly) called *depth cueing* by computer graphics texts. Depth cueing in computer graphics is the ability to vary the color of an object depending on its distance from the viewpoint, as illustrated in Figure 8.21. Normally, this is done so that more distant objects are faded toward

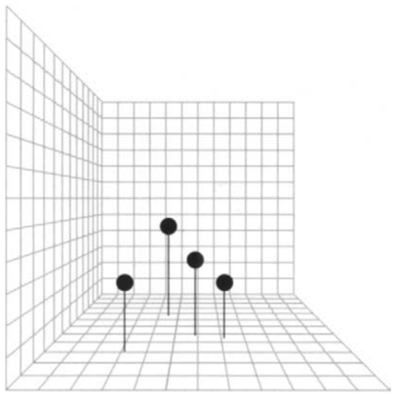


Figure 8.20 Dropping lines to a ground plane is an effective artificial spatial cue.

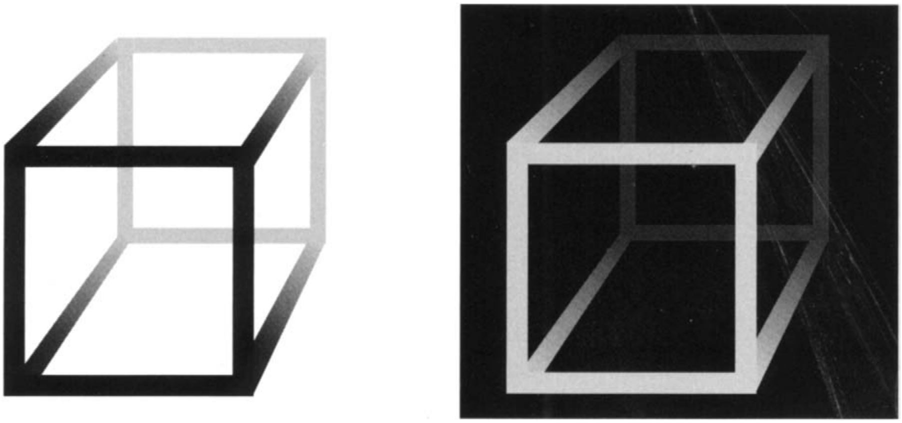


Figure 8.21 Proximity luminance covariance as a depth cue. Object color is altered with distance in the direction of the background color. This simulates extreme atmospheric effects.

the background color, becoming darker if the background is dark and lighter if the background is light.

Proximity luminance covariance mirrors an environmental depth cue sometimes called *atmospheric depth*. This refers to the reduction in contrast of distant objects in the environment, especially under hazy viewing conditions. However the depth cueing is used in computer graphics, it is generally much more extreme than any atmospheric effects that occur in nature, and for this reason it can be considered an "artificial" cue. Dosher et al. (1986) showed that proximity luminance covariance could function as an effective depth cue but was weaker than stereo for static displays. With moving displays, however, proximity luminance covariance became a relatively stronger cue in making an ambiguous 3D scene unambiguous.

Depth Cues in Combination

In computer graphics–based data displays, the designer has considerable freedom about which depth cues to include in a data visualization and which to leave out. One approach would be to simply include all of them. However, this is not always the best solution. There can be considerable costs associated with creating a stereoscopic display or with using real-time animation to take advantage of structure-from-motion cues. Other cues, such as depth-of-focus information, are difficult or impossible to compute in the general case, because without knowing what object the observer is looking at, it is impossible to determine what should be shown in focus and what should be shown out of focus. A general theory of space perception should make it possible to determine which depth cues are likely to be most valuable. Such a theory would provide information about the relative values of different depth cues when they are used in combination.

Unfortunately, there is no single, widely accepted unifying theory of space perception, although the issue of how depth cues interact has been addressed by a number of studies. For example, the weighted-average model assumes that depth perception is a weighted linear sum of the depth cues available in a display (Bruno and Cutting, 1988). Alternatively, depth cues may combine in a geometric sum (Dosher et al., 1986). Young et al. (1993) proposed that depth cues are combined additively, but are weighted according to their apparent reliability in the context of other cues and relevant information. However, there is also evidence that some depth cues—in particular, occlusion—work in a logical binary fashion rather than contributing to an arithmetic or geometric sum. For example, if one object overlaps another in the visual image, it is perceived as closer to the observer.

Most of the work on the combination of spatial information implicitly contains the notions that spatial information is combined into a single cognitive model of space and that this model is used as a resource in performing all spatial tasks. This theoretical position is illustrated in Figure 8.22. However, evidence is accumulating that this unified model of cognitive space is fundamentally flawed.

The alternative model is that depth cues are combined expeditiously, depending on task requirements (Bradshaw et al., 2000; Fine and Jacobs, 1999). For example, Wanger et al. (1992)

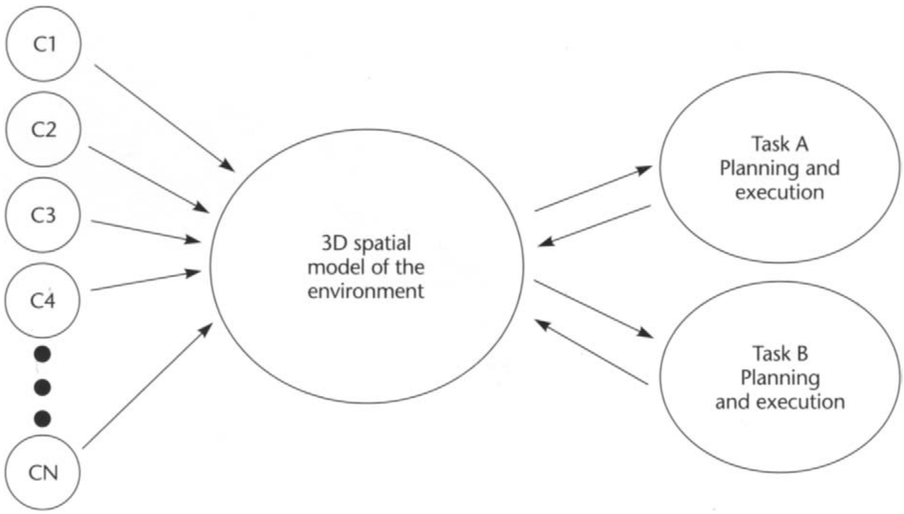


Figure 8.22 Most models of 3D space perception assume that depth cues (C1 . . . CN) feed into a cognitive 3D model of the environment. This, in turn, is used as a resource for task planning and execution.

showed that cast shadows and motion parallax cues both helped in the task of orienting one virtual object to match another. Correct linear perspective (as opposed to parallel orthographic perspective) actually increased errors; thus, it acted as a kind of negative depth cue for this particular task. However, when the task was one of translating an object, linear perspective was the most useful of the cues, and motion parallax did not help at all. Bradshaw et al. (2000) showed that stereopsis is critical in setting objects at the same distance from the observer, but motion parallax is more important for a different layout task involving the creation of a triangle laid out in depth. This alternative model is illustrated in Figure 8.23. Depending on whether the task is threading a needle or running through a forest, different depth cues are most informative, and judgments are made depending on the best available evidence.

An application designer's choice is not whether to design a 3D or 2D interface, but rather how much 3D to use, because depth cues can be applied somewhat independently. For example, in a static picture we use all the monocular pictorial depth cues, but not motion parallax or stereoscopic disparity. If we add structure-from-motion information, we get what we see at the movie theater. If we add stereo to a static picture, the result is the kind of stereoscopic viewer

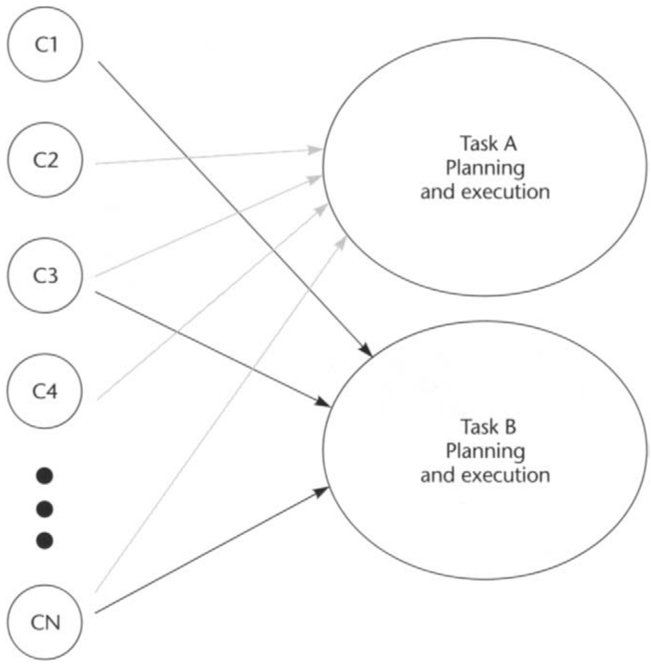


Figure 8.23 Experimental evidence shows that depth cues (C1 . . . CN) are weighted very differently for different tasks, suggesting that there is no unified cognitive spatial model.

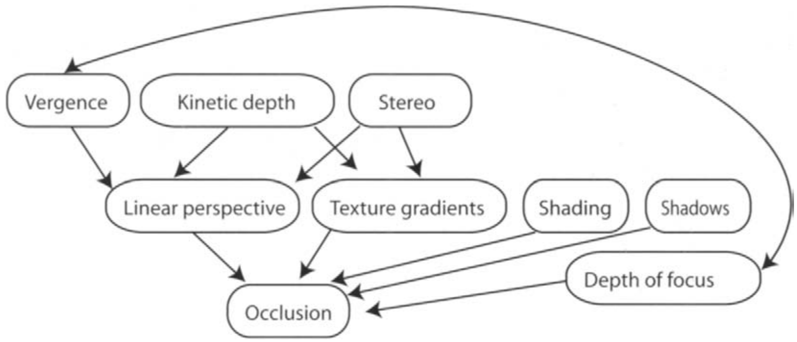


Figure 8.24 A dependency graph for depth cues. Arrows indicate how depth cues depend on each other for undistorted appearance.

popular in Victorian times. We can also use far fewer depth cues. Modern desktop GUIs only use occlusion for windows, some minor shading information to make the menus and buttons stand out, and a cast shadow for the cursor.

However, there are some restrictions on our freedom to arbitrarily choose combinations of depth cues. Figure 8.24 shows a dependency graph for depth cues. An arrow means that a particular cue depends on another cue to appear correctly. This graph does not show absolute rules that cannot be broken, but it does imply that breaking the rules will have undesirable consequences. For example, the graph shows that kinetic depth depends on correct perspective. It is possible break this rule and show kinetic depth with a parallel (orthographic) perspective. The undesirable consequence is that a rotating object will appear to distort as it rotates. This graph is transitive; all of the depth cues depend on occlusion being shown properly, because they all depend on something that in turn depends on occlusion. Thus, occlusion is, in a sense, the most basic depth cue; it is difficult to break the occlusion dependency rule and have a perceptually coherent scene.

Task-Based Space Perception

The obvious advantage of a theory of space perception that takes the task into account is that it can be directly applied to the design of interactive 3D information displays. The difficulty is that the number of tasks is potentially large, and many tasks that appear at first sight to be simple and unified are found, upon more detailed examination, to be multifaceted. Nevertheless, taking the task into account is unavoidable; perception and action are intertwined. If we are to understand space perception, we must understand the purpose of perceiving. The best hope for progress lies in identifying a small number of elementary tasks that are as common as possible. Once this is done, informed design decisions can be made. The remainder of this chapter is devoted to analyzing the following tasks:

- Tracing data paths in 3D graphs
- Judging the morphology of surfaces and surface target detection
- Finding patterns of points in 3D space
- Judging the relative positions of objects in space
- Judging the relative movement of self within the environment
- Reaching for objects
- Judging the "up" direction
- Feeling a sense of presence

This list of eight tasks is at best only a beginning; each has many variations. One additional task, navigation (or *wayfinding*), is discussed in Chapter 10.

Tracing Data Paths in 3D Graphs

Many kinds of information structures can be represented as networks of nodes and arcs, technically called *graphs*. Figure 8.25 shows an example of object-oriented computer software represented using a 3D graph. Nodes in the graph stand for various kinds of entities, such as modules, classes, variables, and methods. The 3D spars that connect the entities represent various kinds of relationships characteristic of object-oriented software, such as inheritance, function calls, and variable usage.

Information structures are becoming so complex that there has been considerable interest in the question of whether a 3D visualization will reveal more information than a 2D visualization.

One special kind of graph is a tree, illustrated in Figure 8.26. Trees are a standard technique for representing hierarchical information, such as organizational charts or the structure of information in a computer directory. The *cone tree* is a graphical technique for representing tree graph information in 3D (Robertson et al., 1993). It shows the tree branches arranged around a series of circles, as illustrated in Figure 8.27. The inventors of the cone tree claim that as many as 1000 nodes may be displayable without visual clutter using cone trees—clearly more than could be contained in a 2D layout. However, 3D cone trees require more complex user interactions to access some of the information than are necessary for 2D layouts.

Empirical evidence also exists that shows that the number of errors in detecting paths in 3D tree structures is substantially reduced if a 3D display method is used. Sollenberger and Milgram (1993) investigated a task involving two 3D trees with intermeshed branches. The task was to discover to which of two tree roots a highlighted leaf was attached. Subjects carried out the task both with and without stereo depth, and with and without rotation to provide kinetic depth. Their results showed that both stereo and kinetic depth viewing reduced errors, but that kinetic depth was the more potent cue.

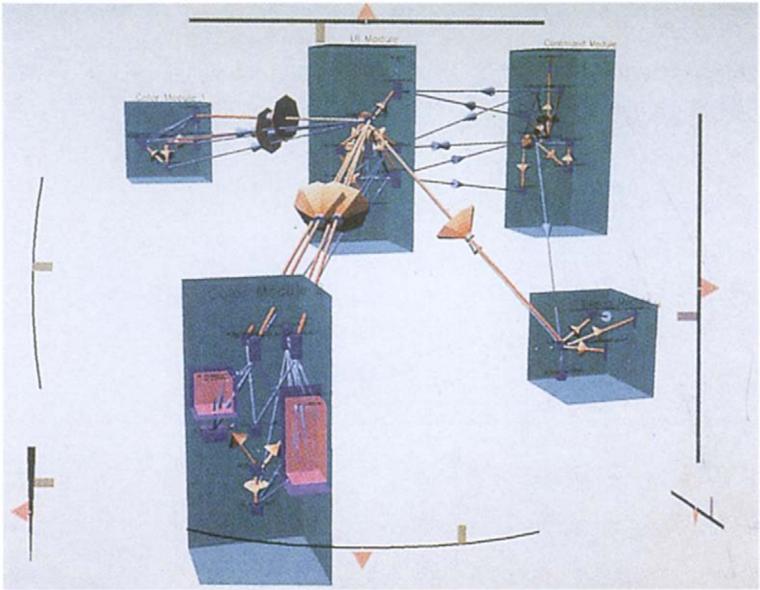


Figure 8.25 The structure of object-oriented software code is represented as a graph in 3D.

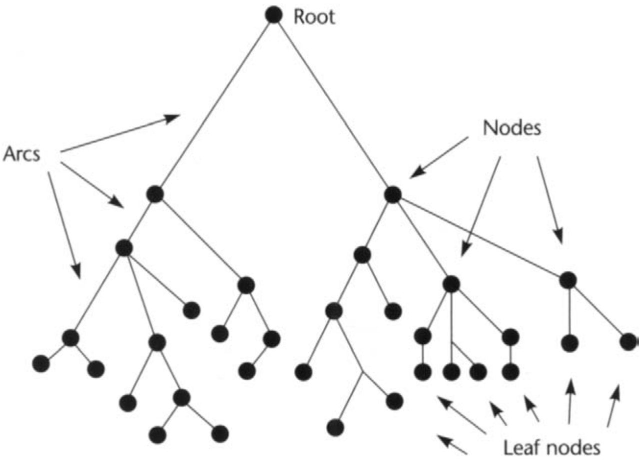


Figure 8.26 A tree is one of the most common ways of structuring information.

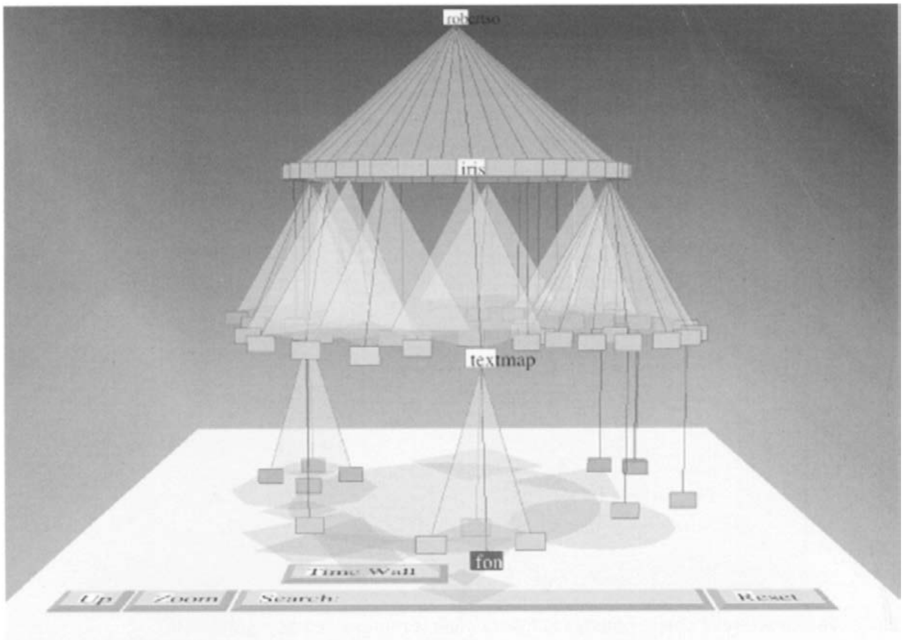


Figure 8.27 The cone tree developed by Robertson et al. (1993).

However, an abstract tree structure is not necessarily a good candidate for 3D visualization, for the reason that a tree data structure can always be laid out on a 2D plane in such a way that none of the paths cross (path crossings are the main reason for errors in path-tracing tasks). Conversely, more general graph structures, such as that illustrated in Figure 8.28, usually cannot be laid on a plane without some paths crossing and therefore would benefit more from 3D viewing techniques.

To study the effects of stereo and kinetic depth cues on 3D visualization of graphs, we systematically varied the size of a graph laid out in 3D and measured path-tracing ability with both stereoscopic and motion depth cues (Ware and Franck, 1996). Our results, illustrated in Figure 8.28, showed a factor-of-1.6 increase in the complexity that could be viewed when stereo was added to a static display, but a factor-of-2.2 improvement when kinetic depth cues were added. A factor-of-3.0 improvement occurred with both stereo and kinetic depth cues. These results held for a wide range of graph sizes. A subsequent experiment showed that the advantage of kinetic depth cues applied whether the motion was coupled to movements of the head or movements of the hand, or consisted of automatic oscillatory rotation of the graph.

Occlusion is one additional depth cue that should make it easier to differentiate arcs if they are colored differently, because occlusion makes it easier to see which arcs lie above and beneath. It seems unlikely that other depth cues will contribute much to a path-tracing task. There is no

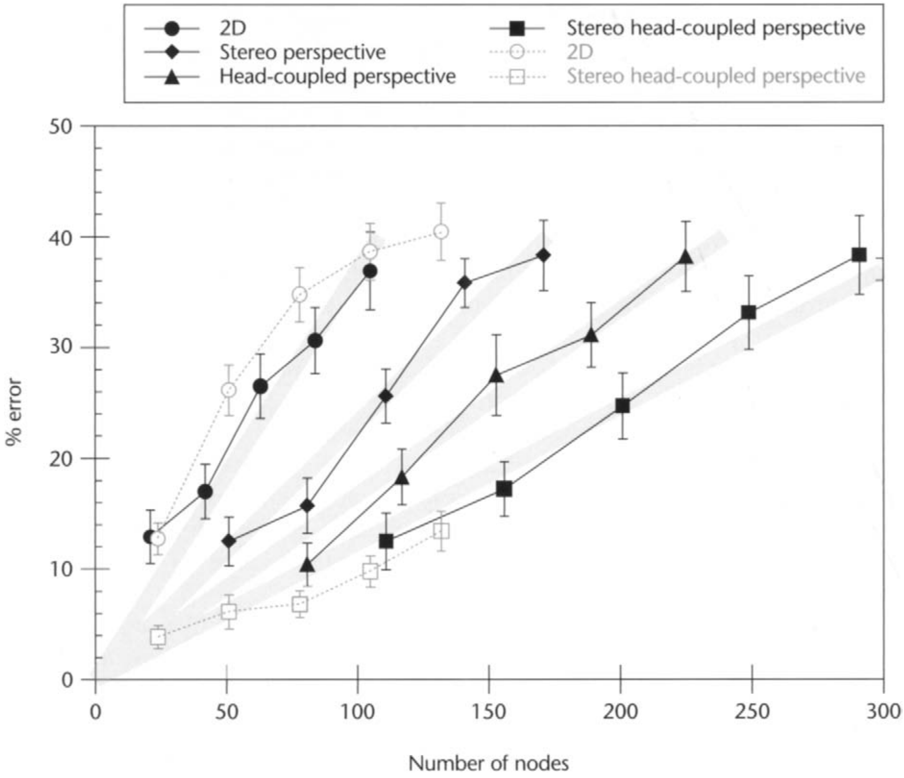


Figure 8.28 The plot shows how the errors increase as the number of nodes increases in a 3D graph representing stereo and motion depth cues.

obvious reason that we should expect perspective viewing to aid the comprehension of connections between nodes in a 3D graph, and this was confirmed empirically by our study (Ware and Franck, 1996). There is also no reason to suppose that shading and cast shadows would provide any significant advantage in a task involving connectivity, although shading might help in revealing the orientation of the arcs.

Judging the Morphology of Surfaces and Surface Target Detection

Shape-from-shading and texture cues are extremely important in revealing surface shape, as discussed in Chapter 7. Here is some additional information on the value of stereoscopic and motion parallax information.

Experimental evidence suggests that the relative contribution of structure-from-motion and stereoscopic depth depends on very specific task-related factors. Surface shape detection is not a simple problem. A study of the judged *heights* of cones showed that stereo depth was much more effective than structure-from-motion (Durgin et al., 1995). Conversely, Tittle et al. (1995) showed that structure-from-motion information was more important than stereo information in judging the *gradient* of a textured surface. Disparity curvature information may be considerably more important than absolute disparities in judging the shapes of surfaces, because this information is relatively invariant with viewing distance. Rogers and Cagnello (1989) showed that the kind of curvature matters. In a stereoscopic display, we are approximately twice as sensitive to the curvature of a horizontally oriented cylinder as we are to that of a vertically oriented cylinder.

There are also temporal factors to be taken into consideration. When we are viewing stereoscopic displays, it can take several seconds for the impression of depth to build up. However, stereoscopic depth and structure-from-motion information interact strongly. With moving stereoscopic displays, the time to fusion can be considerably shortened (Patterson and Martin, 1992). In determining shape from surfaces made from random dot patterns, using both stereoscopic and motion depth cues, Uomori and Nishida (1994) found that kinetic depth information dominated the initial perception of surface shape, but after an interval of four to six seconds, stereoscopic depth came to dominate.

Overall, it is clear that the way different depth cues combine in judgments of surface shape is highly complex. The relative values of stereo and structure-from-motion depend on the viewing distance, the texture of the surface, the kind of surface shape, and the viewing time. Because of this, when arbitrary surface shapes are being viewed, stereoscopic depth, kinetic depth, shape-from-shading, and surface textures can all add to our understanding of surface shape. The most important cues for any particular surface will vary, but including them all will ensure that good shape information is always presented.

Stereoscopic depth can also be used to enhance real-world imagery. Kalaugher (1985) developed an intriguing technique that enabled a fusion of real-world imagery and photographic imagery. His method is simply to take a slide viewer out into the field, to the same place where a photographic slide image of the scene was previously taken. One eye is then used to view the photographic image while the other eye views the actual scene. Using this technique, it is possible to either enhance or reduce stereoscopic depth simply by moving laterally. Kalaugher reported that with this viewing technique, otherwise invisible features in the real world, such as ledges on distant cliffs, could be seen. A variation of the technique can also be used to view changes in a landscape, such as landslides. When the eyes are alternately covered, these appear as anomalous depth or as movement effects.

Patterns of Points in 3D Space

The scatter plot is probably the most effective method for finding unknown patterns in 2D discrete data. In a 3D scatter plot, three data variables are used to position a point with respect to the XYZ axes. The resulting 3D scatter plot is usually rotated around a vertical axis, exploiting structure-from-motion to reveal its structure (Donoho et al., 1988). This technique can be added to the color- and shape-enhanced scatter plots discussed in Chapters 4 and 5.

There has been little or no empirical work on the role of depth cues in perceiving structures such as clusters and correlations in 3D. Nevertheless, a number of conclusions can be deduced from our understanding of the way depth cues function.

Perspective cues will not help us perceive depth in a 3D scatter plot, because a cloud of small, discrete points has no perspective information. If the points all have a constant and relatively large size, weak depth information will be produced by the size gradient. Similarly, with small points, occlusion will not provide useful depth information, but if the points are larger, some ordinal depth information will be perceivable. If there are a large number of points, cast shadows will not provide information, because it will be impossible to determine the association between a given point and its shadow. Shape-from-shading information will be missing, because a point has no orientation information. Each point will reflect light equally, no matter where it is placed and no matter where the light source is placed.

Hence, it is likely that the only important depth cues that will be useful in a 3D scatter plot are stereoscopic depth and structure-from-motion. There seems to be little doubt that using both will be advantageous. As with the perception of surfaces, discussed above, the relative advantages of the different cues will depend on a number of factors. Stereo depth will be optimal for fine depth discriminations between points that lie near one another in depth. Structure-from-motion will be more important for points that lie farther apart in depth.

One of the problems with visualizing clouds of data points is that the overall shape of the cloud cannot easily be seen, even when stereo and motion cues are provided. One way to add extra shape information to a cloud of discrete points is to add shape-from-shading information artificially. It is possible to treat a cloud of data points as though each point were actually a small, flat oriented object. These flat particles can be artificially oriented, if they lie near the boundary of the point cloud, to reveal the shape of the cloud when shading is applied. In this way, perception of the cloud's shape can be considerably enhanced, and shape information can be perceived without additional stereo and motion cues. At the same time, the positions of individual points can be perceived. Figure 8.29 illustrates this.

Judging Relative Positions of Objects in Space

Judging the relative positions of objects is a complex task, performed very differently depending on the overall scale and the context. When very fine depth judgments are made in the near vicinity, as in the task of threading a needle, stereopsis is the strongest single cue. Stereoscopic depth perception is a superacuity and is optimally useful for objects held at about arm's length. For these fine tasks, motion parallax is not very important, as evidenced by the fact that people hold their heads still when threading needles.

In larger environments, stereoscopic depth can play no role at all at distances beyond 30 m. Conversely, when we are judging the overall layout of objects in a larger environment, motion

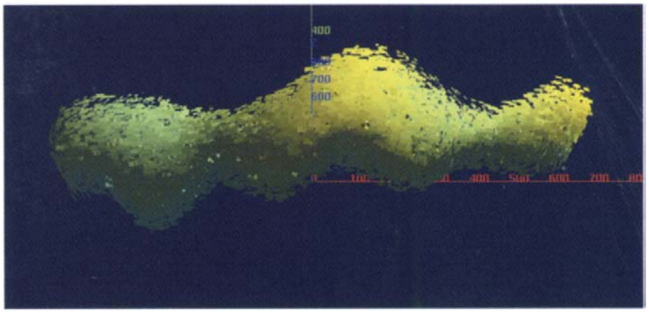


Figure 8.29 A cloud of discrete points is represented by oriented particles. The orientation is determined by using an inverse-square law of attraction between the particles. When the cloud is artificially shaded, its shape is revealed (Li, 1997).

parallax, linear perspective, cast shadows, and texture gradients all contribute, depending on the exact spatial arrangement.

Gibson (1986) noted that much of size constancy can be explained by a referencing operation with respect to a textured ground plane. The sizes of objects that rest on a uniformly textured ground plane can be obtained by reference to the texture element size. Objects slightly above the ground plane can be related to the ground plane through the shadows they cast. In artificial environments, a very strong artificial reference can be provided by dropping a vertical line to the ground plane. A practical aid to visualizing spatial layout is a regular grid or checkerboard on the floor and walls, as illustrated in Figure 8.20. A grid provides a strong linear perspective cue, as well as a reference texture that may be optimal for many applications.

Judging the Relative Movement of Self within the Environment

When we are navigating through a virtual environment representing an information space, there are a number of frames of reference that may be adopted. For example, an observer may feel she is moving through the environment or that she is stationary and the world is moving past. In virtual-environment systems that are either helmet-mounted or monitor-based, the user rarely actually moves physically any great distance, because real-world obstacles lie in the way. If self-movement is perceived, it is generally an illusion. Note that this applies only to linear motion, not to rotations; users with helmet-mounted displays can usually turn their heads quite freely.

A sensation of self-movement can be strongly induced even when the subject is not moving. This phenomenon, called *vection*, has been studied extensively. When observers are placed inside a large moving visual field—created either by a physical drum or by means of computer graphics within a virtual-reality helmet—they invariably feel that they are moving, even though they are not. A number of visual parameters influence the amount of vection that is perceived:

Field size: In general, the larger the area of the visual field that is moving, the stronger the experience of self-motion (Howard and Heckman, 1989).

Foreground/background: Much stronger vection is perceived if the moving part of the visual field is perceived as background more distant from the observer than foreground objects (Howard and Heckman, 1989). In fact, vection can be perceived even with quite a small moving field, if that field is perceived to be relatively distant. The classic example occurs when someone is sitting in a train at a station and the movement of an adjacent vehicle (seen through a window) causes that person to feel he or she is moving even though this is not the case.

Frame: Vection effects are considerably increased if there is a static foreground frame between the observer and the moving background (Howard and Childerson, 1994).

Stereo: Stereoscopic depth can determine whether a moving pattern is perceived as background or foreground, and thereby increase or decrease vection (Lowther and Ware, 1996).

In aircraft simulators and other vehicle simulators, it is highly desirable that the user experiences a sense of motion, even though the simulator's actual physical motion is relatively small or non-existent. One of the unfortunate side effects of this perceived motion is simulator sickness. The symptoms of simulator sickness can appear within minutes of acute exposure to perceived extreme motion. Kennedy et al. (1989) report that between 10 and 60% of users of immersive displays report some symptoms of simulator sickness. This high incidence may ultimately be a major barrier to the adoption of fully immersive display systems.

Simulator sickness is thought to be caused by conflicting cues from the visual system and the vestibular system of the inner ear. When most of the visual field moves, the brain usually interprets this as a result of self-motion. But if the observer is in a simulator, no corresponding information comes from the vestibular system. According to this theory, the contradictory information results in nausea.

There are ways to ensure that simulator sickness does occur, and ways of reducing its effects. Turning the head repeatedly while moving in a simulated virtual vehicle is almost certain to induce nausea (DiZio and Lackner, 1992). This means that a virtual ride should never be designed in which the participant is expected to look from side to side while wearing a helmet-mounted display. Simulator sickness in immersive virtual environments can be mitigated by initially restricting the participant's experience to short periods of exposure, lasting only a few minutes each day. This allows the user to build up a tolerance to the environment, and the periods of exposure can gradually be lengthened (McCauley and Sharkey, 1992).

Reaching for Objects

A number of researchers have investigated how eye–hand coordination changes when there is a mismatch between feedback from the visual sense and the proprioceptive sense of body position.

A typical experiment involves subjects pointing at targets while wearing prisms that displace the visual image relative to the proprioceptive information from their muscles and joints. Subjects adapt rapidly to the prism displacement and point accurately. Work by Rossetti et al. (1993) suggests that there may be two mechanisms at work, a long-term, slow-acting mechanism that is capable of spatially remapping misaligned systems, and a short-term mechanism that is designed to realign the visual and proprioceptive systems within a fraction of a second. These results have been confirmed in studies with fish-tank VR systems, showing that a large translational offset between the hand position and the object being manipulated with the hand has only a small effect on performance (Ware and Rose, 1999).

Rotational mismatches between what is seen and what is held may have a much greater negative impact on eye–hand coordination than translational mismatches. Experiments with prisms that invert the visual field have shown that it can take months to reach behavior approaching normal performance under this condition, and adaptation may never be complete (Harris, 1965).

Designers of 3D display systems must make choices about which depth cues to include. In a full-blown virtual reality system, the goal is to include all of the depth cues at the highest fidelity possible, but in practical systems for molecular modeling or 3D computer-aided design, various tradeoffs must be made. Two of the options are whether to use a stereoscopic display and whether to provide motion parallax through perspective coupled to head position. Both require an investment in technology not normally provided with computer workstations. The evidence suggests that having a stereoscopic display is more important than the motion parallax that occurs through the motion of the user's head with respect to the objects of attention (Boritz and Booth, 1998; Arsenault and Ware, in press). It appears that users can adapt rapidly to a stereoscopic view from an incorrect viewpoint.

Actually providing a sense of physical contact with nearby objects is also important in calibrating the proprioceptive system, especially for grasping (Mackenzie and Iberall, 1994). Unfortunately, this component of natural object interaction is proving very difficult to simulate. Although VR displays can produce excellent 3D sound and reasonable simulation of visual space, the simulation of touch is still very poor. There is no technology that can produce a physically touchable virtual object at any desirable location within a reasonably large volume of space, although such simulations can be made for small volumes of space by devices such as the PHANTOM (Massie and Salisbury, 1994). This means that it is possible to create small-scale virtual environments that allow for touch and high-resolution stereo display, but not to create large-scale data spaces with the same haptic affordances.

Judging the "Up" Direction

In abstract 3D data spaces (for example, molecular models), there is often no sense of an "up" direction, and this can be confusing. The "up" direction is defined both by gravity, sensed by the vestibular system in the inner ear, and by the presence of the ground on which we walk. Much of the research that has been done on perceived "up" and "down" directions has been done as part of space research, to help us understand how people can best orient themselves in a gravity-free environment.

Nemire et al. (1994) showed that linear perspective provides a strong cue in defining objects perceived at the same horizontal level. They showed that a linear grid pattern on the virtual floor and walls of a display strongly influenced what the participants perceived as horizontal; to some extent, this overrode the perception of gravity. Other studies have shown that placing recognizable objects in the scene very strongly influences a person's sense of self-orientation. The presence of recognizable objects with a known normal orientation with respect to gravity, such as a chair or a standing person, can strongly influence which direction is perceived as up (Howard and Childerson, 1994). Both of these results can easily be adapted to virtual environments. Providing a clear reference ground plane and placing recognizable objects on it can define, to some extent, a vertical polarity for a data space.

The Aesthetic Impression of 3D Space (Presence)

One of the most nebulous and ill-defined tasks related to 3D space perception is achieving a sense of *presence.* What is it that makes a virtual object or a whole environment seem vividly three-dimensional? What is it that makes us feel that we are actually present in an environment?

Much of presence has to do with a sense of engagement, and not necessarily with visual information. A reader of a powerfully descriptive novel may *visualize* (to use the word in its original cognitive sense) himself or herself in a world of the author's imagination—for example, watching Ahab on the back of the great white whale, Moby-Dick.

Presence is somewhat anomalous in a task-based classification of spatial information, because presence as such does not have a clear task associated with it. It is simply the sense of being there. Nevertheless, a number of practical applications require a sense of presence. For an architect designing a virtual building to present to a client, the feeling of spaciousness and the aesthetic quality of that space may be all-important. In virtual tourism, where the purpose is to give a potential traveler a sensation of what the Brazilian rain forest is really like, presence is also crucial.

A number of studies have used virtual-reality techniques for phobia desensitization. In one study by North et al. (1996), patients who had a fear of open spaces (agoraphobia) were exposed to progressively more challenging virtual open spaces. The technique of progressive desensitization involves taking people closer and closer to the situations that cause them fear. As they overcome their fears at one level of exposure, they can be taken to a slightly more stressful situation. In this way, they can overcome their phobias, one step at a time. The reason for using VR simulations in phobia desensitization is to provide control over the degree of presence and to reduce the stress level by enabling the patient to exit the stressful environment instantaneously. After treatment in a number of virtual environments, the experimental subjects of North et al. scored lower on a standardized Subjective Units of Discomfort test.

In developing a virtual-reality theme park attraction for Disneyland, Pausch et al. (1996) observed that high frame rate and high level of detail were especially important in creating a sense of presence for users "flying on a magic carpet." Presenting a stereoscopic display did not enhance the experience. Empirical studies have also shown that high-quality structure-from-motion information contributes more to a sense of presence than does stereoscopic display (Arthur et al., 1993). However, the sense of presence may also be divided into subtasks. Hendrix and Barfield (1996) found stereoscopic viewing to be very important when subjects were asked to rate the extent to which they felt they could reach for and grasp virtual objects, but it did not contribute at all to the sense of the overall realism of the virtual condition. Hendrix and Barfield also found that having a large field of view was important to creating a sense of presence.

Conclusion

High-quality, interactive 3D displays are now becoming cheap, although even mediocre-quality VR systems are still expensive. But creating a 3D visualization environment is considerably more difficult than creating a 2D system with similar capabilities. We still lack design rules for 3D environments, and many interaction techniques are competing for adoption.

The strongest argument for the ultimate ascendancy of 3D visualization systems, and 3D user interfaces in general, must be that we live in a 3D world and our brains have evolved to recognize and interact with 3D. The 3D design space is self-evidently richer than the 2D design space, because a 2D space is a part of 3D space. It is always possible to flatten out part of a 3D display and represent it in 2D.

Nevertheless, it also should be cautioned that going from 2D to 3D adds far less visual information than might be supposed. Consider the following simple argument. On a line of a computer display, we can perceive 1000 distinct pixels. On a plane of the same display, we can display $1000 \times 1000 = 1{,}000{,}000$ pixels. But going to a stereoscopic display only increases the number of pixels by a factor of 2. Even this is an overestimate, because it assumes that the images presented to the two eyes are completely independent, whereas in fact they must be highly correlated for us to perceive stereoscopic depth. We may only be able to fuse stereoscopically images that differ by 10% or so. Of course, as we have shown in this chapter, motion parallax can enable us to see more information, and in the case of 3D networks, a network about three times as large can be perceived with stereo and motion parallax. The other depth cues, such as occlusion and linear perspective, certainly help us perceive a coherent 3D space, but as the study of Cockburn and McKenzie (2001) suggests, we should not automatically assume that 3D provides more readily accessible information.

This chapter has been about the use of 3D spaces to display information. It should not be assumed; however, that a 3D display is automatically superior to a 2D solution. Deciding whether or not to use a 3D display must involve deciding whether there are sufficient important subtasks for which 3D is clearly beneficial. The complexity and the consistency of the user interface for the whole application must be weighed in the decision. Even if 3D is better for one or two subtasks, the extra cost involved and the need for nonstandard interfaces for the 3D components may suggest that a 2D solution would be better overall. In terms of overall assessment, the cost of navigation is an essential component, and many 3D navigation methods are considerably slower than 2D alternatives. Even if we can show somewhat more information in 3D, the rate of information access may be slower. Issues relating to overall system costs are dealt with in Chapters 10 and 11.