CHAPTER 7

Visual Objects and Data Objects

The object metaphor is pervasive in the way we think about abstract data. Object-oriented programming is one example; the body politic is another. Object-related concepts are also basic in modern systems design. A modular system is one that has easily understood and easily replaced components. Good modules are *plug-compatible* with one another; they are discrete and separate parts of a system. In short, the concept of a module has a lot in common with the perceptual and cognitive structures that define visual objects. This suggests that visual objects may be an excellent way to represent modular system components. A visual object provides a useful metaphor for encapsulation and cohesiveness, both important concepts in defining modular systems.

For our present purposes, an *object* can be thought of as any identifiable, separate, and distinct part of the visual world. Information about visual objects is cognitively stored in a way that ties together critical features, such as oriented edges and patches of color and texture, so that they can be identified, visually tracked, and remembered. Because visual objects cognitively group visual attributes, if we can represent data values as visual features and group these features into visual objects, we will have a very powerful tool for organizing related data.

Two radically different theories have been proposed to explain object recognition. The first is image-based. It proposes that we recognize an object by matching the visual image with something roughly like a snapshot stored in memory. The second type of theory is structure-based. It proposes that is analyzed in terms of primitive 3D forms and the structural interrelationships between them. Both of these models have much to recommend them, and it is entirely plausible that each is correct in some form. It is certainly clear that the brain has multiple ways of analyzing visual input. Certainly, both models provide interesting insights into how to display data effectively.

Image-Based Object Recognition

We begin with some evidence related to picture and image perception. People have a truly remarkable ability to recall pictorial images. In an arduous experiment, Standing et al. (1970) presented subjects with a list of 2560 pictures at a rate of one every 10 seconds. This was like the family slide show from hell, it took them more than seven hours spread over a four-day period. Amazingly, when subsequently tested, subjects were able to distinguish pictures from others not previously seen, with better than 90% accuracy.

People can also recognize objects in images that are presented very rapidly. Suppose you asked someone, "Is there a dog in one of the following pictures?" and then showed them a set of images, rapidly, all in the same place, at a rate of 10 per second. Remarkably, they will be able to detect the presence, or absence, of a dog in one of the images most of the time. This experimental technique is called *rapid serial visual presentation* (RSVP). Experiments have shown that the maximum rate for the ability to detect common objects in images is about 10 images per second (Potter and Levy, 1969; Potter, 1976).

A related phenomenon is *attentional blink*. If, in a series of images, a second dog were to appear in an image within 350 ms of the first, people do not notice it (or anything else). This moment of blindness is the attentional blink (Coltheart, 1999). It is conjectured that the brain is still processing the first dog, even though the image is gone, and this prohibits the identification of other objects in the sequence.

It is useful to make a distinction between recognition and recall. We have a great ability to recognize information that we have encountered before, as the picture memory experiment of Standing et al. shows. However, if we are asked to reconstruct visual scenes—for example, to recall what happened at a crime scene—our performance is much worse. Recognition is much better than recall. This suggests that a major use of visual images can be as an aid to memory. An image that we recognize can help us remember events or other information related to that image. This is why icons are so effective in user interfaces; they help us to recall the functionality of computer programs.

More support for image-based theories comes from studies showing that three-dimensional objects are recognized most readily if they are encountered from the same view direction as when they were initially seen. Johnson (2001) studied subjects' abilities to recognize bent pipe structures. Subjects performed well if the same viewing direction was used in the initial viewing and in the test phase; they performed poorly if a different view direction was used in the test phase. But subjects were also quite good at identification from exactly the opposite view direction. Johnson attributed this unexpected finding to the importance of silhouette information. Silhouettes would have been similar, although flipped left-to-right from the initial view.

Although most objects can easily be recognized independent of the size of the image on the retina, image size does have some effect. Figure 7.1 illustrates this. When the picture is seen from a distance, the image of the Mona Lisa face dominates; when it is viewed up close, smaller objects become dominant: a gremlin, a bird, and a claw emerge. Experimental work by Biederman and

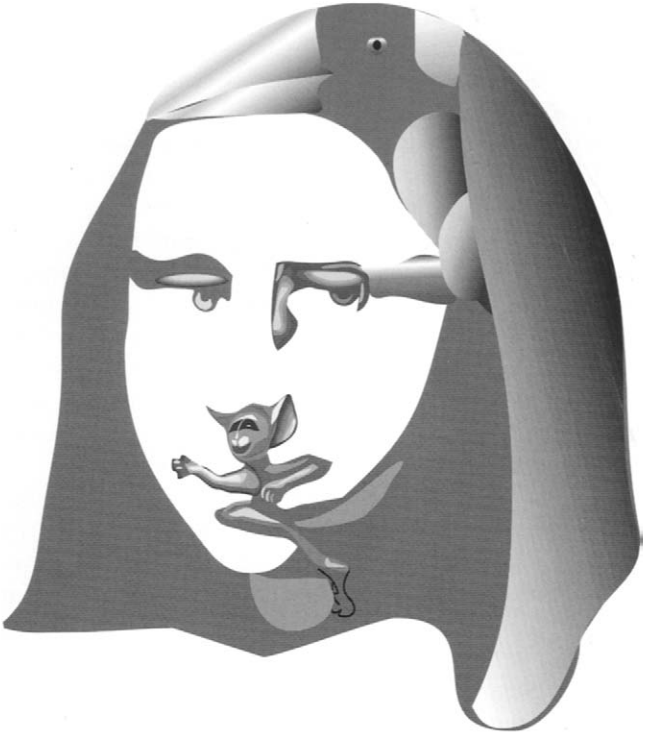

Figure 7.1 When the image of the Mona Lisa is viewed from a distance, the face dominates. But look at it from 30 cm, and the gremlin hiding in the shadows of the mouth and nose emerges. When component objects have a size of about 4 degrees of visual angle, they become maximally visible. *Adapted from the work of the Tel Aviv artist Victor Molev.*

Cooper (1992) suggests that the optimal size for recognizing a visual object is about 4 to 6 degrees of visual angle. This gives a useful rule of thumb for the optimal size for rapid presentation of visual images so that we can best see the visual patterns contained in them.

Another source of evidence for image-based object recognition comes from priming effects. The term *priming* refers to the fact that people can identify objects more easily if they are given prior exposure to some relevant information. Most priming studies have been carried out using verbal information, but Kroll and Potter (1984) showed that *pictures* of related objects, such as a cow and a horse, have a mutually priming effect. This is similar to the priming effect between the words *cow* and *horse*. However, they found little cross-modality priming; the word *cow* provided only weak priming for a picture of a horse. It is also possible to prime using purely visual information, that is, information with no semantic relationship. Lawson et al. (1994) devised a series of experiments in which subjects were required to identify a specified object in a series of briefly presented pictures. Recognition was much easier if subjects had been primed by visually similar images. They argued that this should not be the case if objects are recognized on the basis of a high-level, 3D structural model of the kind that we will discuss later in this chapter; only image-based storage can account for their results.

Priming effects can occur even if information is not consciously perceived. Bar and Biederman (1998) showed pictorial images to subjects, so briefly that it was impossible for them to identify the objects. They used what is called a *masking technique*, a random pattern shown immediately after the target stimulus to remove the target from the iconic store, and they rigorously tested to show that subjects performed at chance levels when reporting what they had seen. Nevertheless, 15 minutes later, this unperceived exposure substantially increased the chance of recognition on subsequent presentation. Although the information was not consciously perceived, exposure to the particular combination of image features apparently primed the visual system to make subsequent recognition easier. They found that the priming effect decreased substantially if the imagery was displaced sideways. They concluded that the mechanism of priming is highly image-dependent and not based on high-level semantic information.

Palmer et al. (1981) showed that not all views of an object are equally easy to recognize. They found that many different objects have something like a *canonical view* from which they are most easily identified. From this and other evidence, a theory of object recognition has been developed, proposing that we recognize objects by matching the visual information with internally stored viewpoint-specific exemplars, or "prototypes" (Edelman and Buelthoff, 1992; Edelman, 1995). According to this theory, the brain stores a number of key views of objects. These views are not simple snapshots; they allow recognition despite simple geometric distortions of the image that occur in perspective transformation. This explains why object perception survives the kinds of geometric distortions that occur when a picture is viewed and tilted with respect to the observer. However, there are strict limits on the extent to which we can change an image before recognition problems occur. For example, numerous studies show that face recognition is considerably impaired if the faces are shown upside down (Rhodes, 1995).

Adding support to the multiple-view, image-based theory of object recognition is neurophysiological data from recordings of single cells in the inferotemporal cortexes of monkeys. Perrett et al. (1991) discovered cells that respond preferentially to particular views of faces. Figure 7.2 shows some of their results. One cell (or cell assembly) responds best to a three-quarter view of a face; another, to profiles, either left or right; still another responds to a view of a head from any angle. We can imagine a kind of hierarchical structure, with the cell assemblies that respond to particular views feeding into higher-level cell assemblies that respond to any view of the object.

Applications of Images in User Interfaces

The fact that visual images are easily recognized after so little exposure suggests that icons in user interfaces should make excellent memory aids, helping us recall the functionality of parts of complex systems. Icons that are readily recognized may trigger activation of related concepts

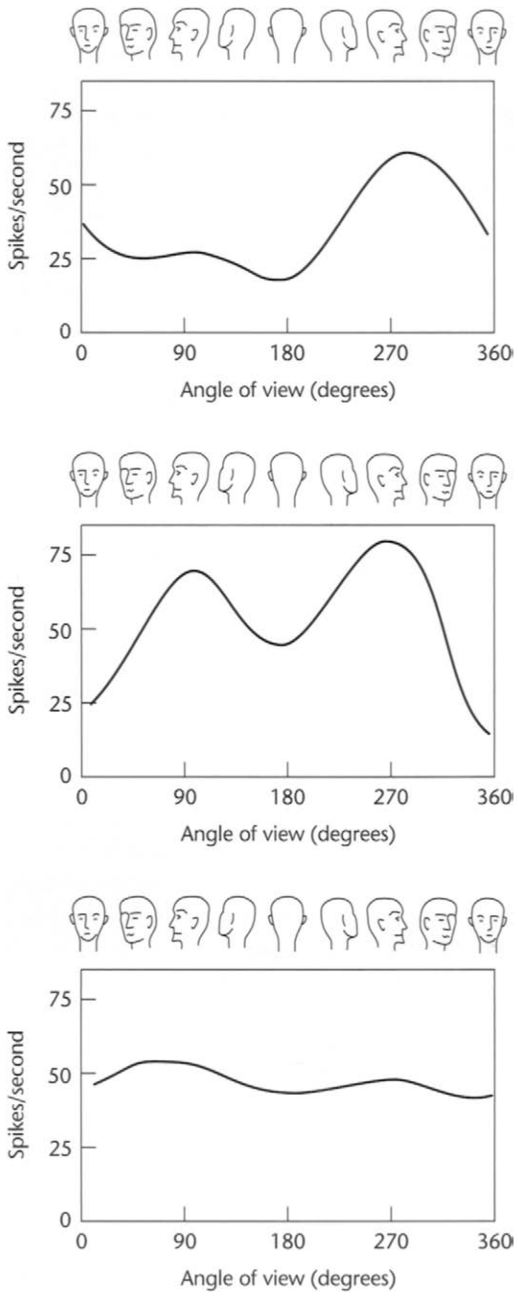

Figure 7.2 The responses of three cells in the temporal cortex of a monkey to faces in different orientations. At the top is a cell most sensitive to a right profile. The middle cell responds well to either profile. The cell at the bottom responds well to a face irrespective of orientation. *Adapted from Perrett et al. (1991).*

in the semantic network of long-term memory. Icons are also helpful because to some extent they can represent pictorially the things they are used to reference.

Priming may be useful in helping people search for particular patterns in data. The obvious way of doing this is to provide sample images of the kind of pattern being sought and repeating the samples at frequent intervals during the search process. An example would be the use of images of sample viruses in a medical screening laboratory.

Searching an Image Database

Presenting images rapidly in sequence may be a useful way to allow users to scan picture databases (Wittenburg et al., 1998; de Bruijn et al., 2000). The fact that people can search rapidly for an image in a sequence of up to 10 pictures per second suggests that presenting images using RSVP may be efficient. Contrast this with the usual method of presenting image collections in a regular grid of small thumbnail images. If it is necessary to make an eye movement to fixate each thumbnail image, it will not be possible to scan more than three to four images per second.

Even though RSVP is promising, there are a number of design problems that must be solved in building a practical interface. Once a likely candidate image is identified as being present in an RSVP sequence, it must still be found. By the time a user responds with a mouse click several images will have passed, more if the user is not poised to press the stop button. Thus, either controls must be added for backing up through the sequence, or part of the sequence must be fanned out in a conventional thumbnail array to confirm that candidate's presence and study it further (Spence, 2002; Wittenburg et al., 1998).

Personal Image Memory Banks

Based on straightforward predictions about the declining cost and increasing capacity of computer memory, it will soon be possible to have a personal memory data bank containing video and sound data collected during every waking moment of a person's lifetime. This could be achieved with an unobtrusive miniature camera, perhaps embedded in a pair of eyeglasses, and assuming continuing progress in solid-state storage, the data could be stored in a device weighing a few ounces and costing a few hundred dollars. Storing speech information will be even more straightforward. The implications of such devices are staggering. Among other things, it would be the ultimate memory aid—the user would never have to forget anything. However, a personal visual memory device of this kind would need a good user interface. One way of searching the visual content might be by viewing a rapidly presented sequence of selected frames from the video sequence. Perhaps 100 per day would be sufficient to jog the user's memory about basic events. Video data compressed in this way might make it possible to review a day in a few seconds, and a month in a few minutes.

Structure-Based Object Recognition

Image-based theories of object recognition imply a rather superficial level of analysis of visual objects. However, there is evidence that a much deeper kind of structural analysis must also occur. Figure 7.3 shows two novel objects, probably never seen by the reader before. Yet despite the fact that the *images* of these two objects are very different from one another, they can be rapidly recognized as representations of the same object. No image-based theory can account for this result.

Geon Theory

Figure 7.4 provides a somewhat simplified overview of a neural-network model of structural object perception, developed by Hummel and Biederman (1992). This theory proposes a hierarchical set of processing stages leading to object recognition. Visual information is decomposed first into edges, then into component axes, oriented blobs, and vertices. At the next layer, three-dimensional primitives such as cones, cylinders, and boxes, called *geons*, are identified. A selection of geons is illustrated in Figure 7.5. Next, the structure is extracted that specifies how the geon components interconnect; for example, in a human figure, the arm cylinder is attached near the top of the torso box. Finally, object recognition is achieved.

Silhouettes

Silhouettes appear to be especially important in determining how we perceive the structure of objects. The fact that simplified line drawings are often silhouettes may, in part, account for our ability to interpret them. At some level of perceptual processing, the silhouette boundaries of objects and the simplified line drawings of those objects excite the same neural contour-extraction mechanisms. Halverston (1992) noted that modern children tend to draw objects on the basis of the most salient silhouettes, as did early cave artists. Many objects have particular

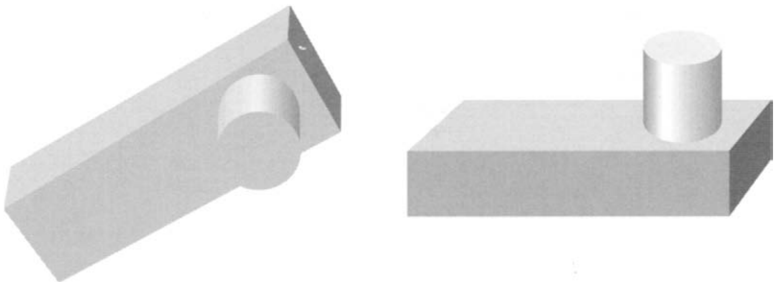

Figure 7.3 These two objects are rapidly recognized as identical, or at least very similar, despite the very different visual images they present.

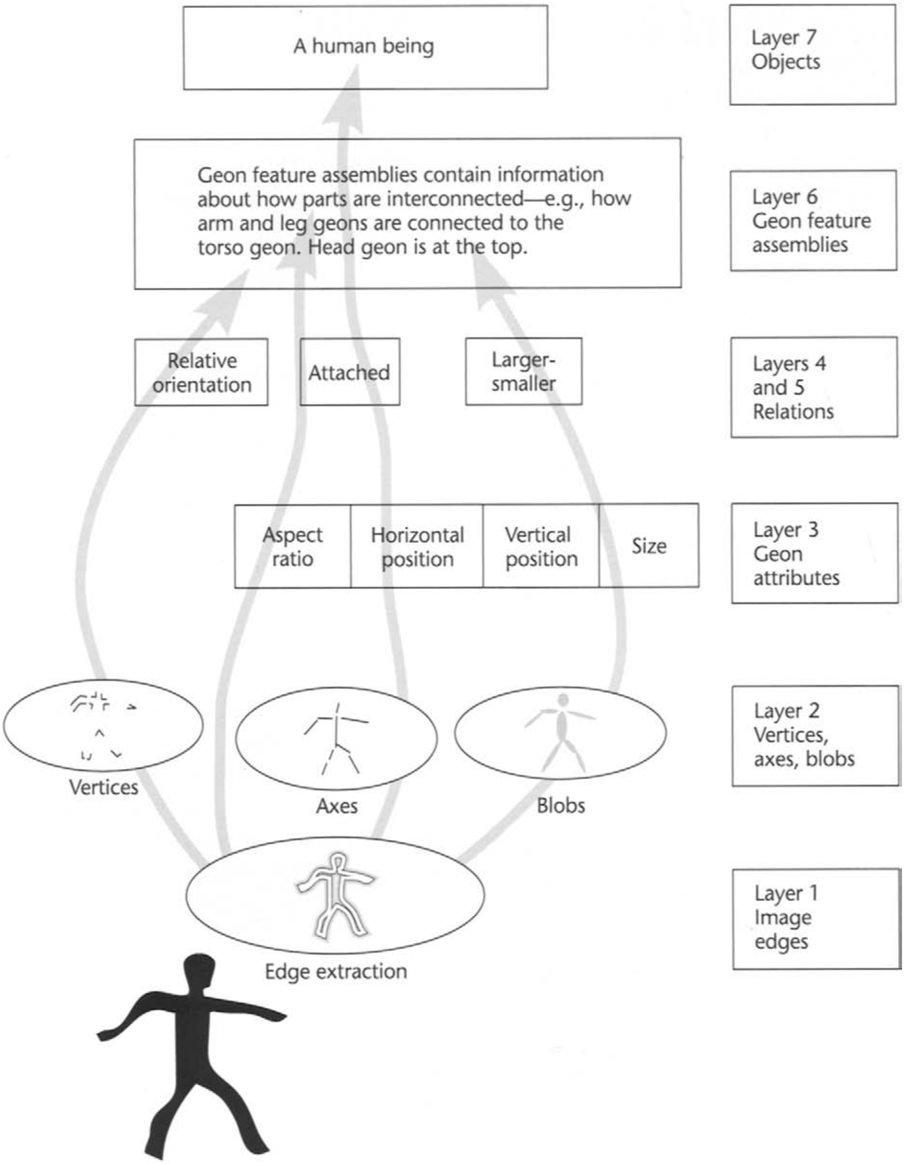

Figure 7.4 A simplified view of Hummel and Biederman's (1992) neural-network model of form perception.

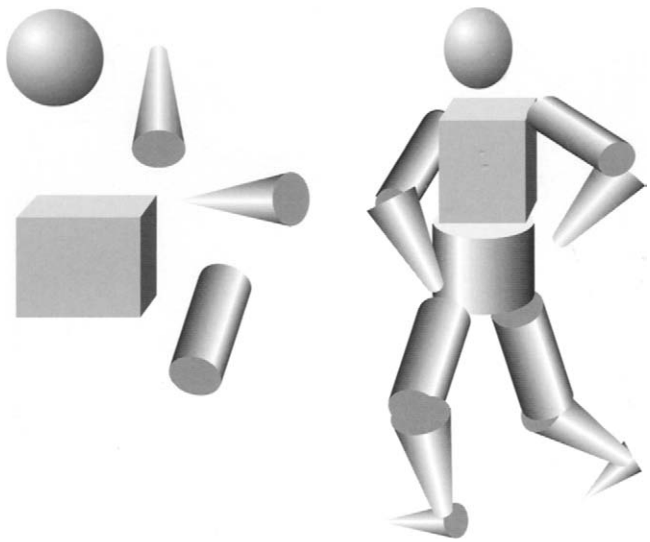

Figure 7.5 According to Biederman's geon theory, the visual system interprets 3D objects by identifying 3D component parts called geons.

silhouettes that are easily recognizable; think of a teapot, a shoe, a church, a person, or a violin. These *canonical* silhouettes are based on a particular view of an object, often from a point at right angles to a major plane of symmetry. Figure 7.6 illustrates canonical views of a teapot and a person.

David Marr suggested ways in which the brain might use silhouette information to extract the structures of objects (Marr, 1982). He argued that "buried deep in our perceptual machinery" are mechanisms that contain constraints determining how silhouette information is interpreted. Three rules are embedded in this perceptual machinery:

1. Each line of sight making up a silhouette grazes the surface exactly once. The set of such points is the *contour generator*. The idea of the contour generator is illustrated in Figure 7.7.

2. Nearby points on the contour of an image arise from nearby points on the contour generator of the viewed object.

3. All the points on the contour generator lie on a single plane.

Under Marr's default assumptions, contour information is used in segmenting an image into its component solids. Marr and Nishihara (1978) suggested that concave sections of the silhouette contour are critical in defining the ways different solid parts are perceptually defined. Figure 7.8 illustrates a crudely drawn animal that we nevertheless readily segment into head, body, neck,

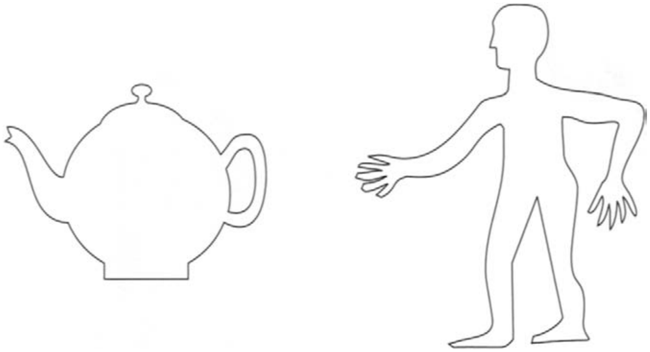

Figure 7.6 Many objects have canonical silhouettes, defined by the viewpoints from which they are most easily recognized. In the case of the man, the overall posture is unnatural, but the component parts—hands, feet, head, and so on—are all given in canonical views.

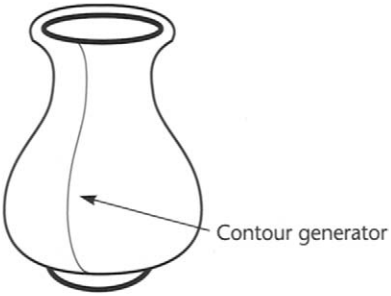

Figure 7.7 According to Marr, the perceptual system makes assumptions that occluding contours are smoothly connected and lie in the same plane. *Adapted from Marr (1982).*

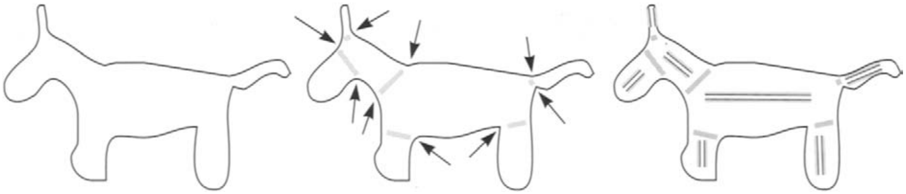

Figure 7.8 Concave sections of the silhouette define subparts of the object and are used in the construction of a structural skeleton. *Adapted from Marr and Nishihara (1978).*

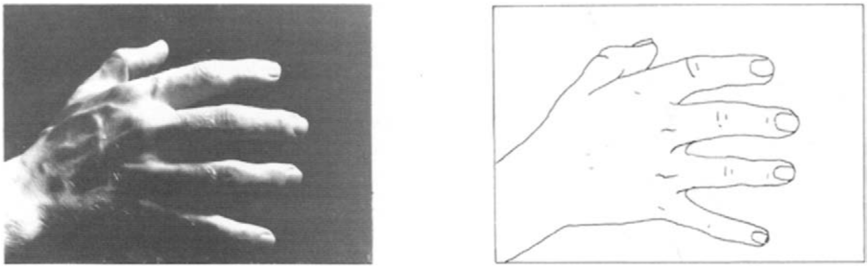

Figure 7.9 A photograph of a hand and a simplified line drawing of the hand. Ryan and Schwartz (1956) showed that a cartoon image was recognized more rapidly than a photograph.

legs, and so on. Marr and Nishihara also suggested a mechanism whereby the axes of the parts become cognitively connected to form a structural skeleton.

One of the consequences of structural theories of perception is that certain simplified views should be easier to read. There are practical advantages to this. For example, a clear diagram may sometimes be more effective than a photograph. This is exactly what Ryan and Schwartz (1956) showed when they found that a hand could be perceived more rapidly in the form of a simplified line drawing than in the form of a photograph (see Figure 7.9).

But this result should not be overgeneralized. Other studies have shown that time is required for detailed information to be perceived (Price and Humphreys, 1989; Venturino and Gagnon, 1992). Simplified line drawings may be most appropriate only when rapid responses are required.

Although image-based theories and structure-based theories of object recognition are usually presented as alternatives, it may be that both kinds of processes occur. If geons are extracted based on concavities in the silhouette, certain views of a complex object will be much easier to recognize. Further, it may well be that viewpoint-dependent aspects of the visual image are stored in addition to the 3D structure of the object. Indeed, it seems likely that the brain is capable of storing many kinds of information about an object or scene if they have some usefulness. The implication is that even though 3D objects in a diagram may be more effective in some cases, care should be taken to provide a good 2D layout.

Faces

Faces are special objects in human perception. Infants learn about faces faster than other objects. It is as if we are born with visual systems primed to learn to recognize important humans, such as our own mothers (Morton and Johnson, 1991; Bruce and Young, 1998; Bushnell et al., 1989). A specific area of our brains, the right middle fusiform gyrus, is especially important in face perception (Puce et al., 1995; Kanwisher et al., 1999; Kanwisher et al., 1997). This area is

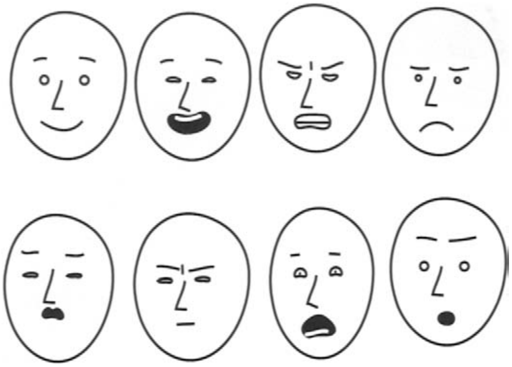

Figure 7.10 Happiness, elation, anger, sadness, disgust, determination, fear, surprise.

also useful for recognizing other complex objects, such as automobiles; although it is not essential as a Volkswagen detector, we cannot recognize faces without it.

Faces have an obvious importance in communication, because we use facial expression to communicate our emotion and degree of interest. Cross-cultural studies by Paul Ekman and coworkers strongly suggests that certain human expressions are universal communication signals, correctly interpreted across cultures and social groups (Ekman and Friesen, 1975; Ekman, 2003). Ekman identified six universal expressions: anger, disgust, fear, happiness, sadness and surprise. These are illustrated in Figure 7.10, along with determination and elation (a variation on happiness). The motion of facial features is also important in conveying emotion. Animated images are necessary to convey a full range of nuanced emotion; it is especially important to show motion of the eyebrows (Basilli, 1978; Sadr et al., 2003).

Facial expressions are produced by the contractions of facial muscles. The *facial action coding system* (FACS) is a widely applied method of measuring and defining groups of facial muscles and their effect on facial expression (Ekman et al., 1988). The eyebrows and mouth are particularly significant in emotion signaling, but the shape of the eyes is also important. There is evidence that false smiles can be distinguished from true smiles from the particular expression around the eyes that occurs with contraction of a muscle that orbits the eye (Ekman et al., 1988; Ekman, 2003). This muscle contracts with true smiles but not with false ones. According to Ekman (2003) it is difficult, if not impossible, to control this voluntarily and thus fake a "true" smile.

The main application of FACS theory in computer displays has been in the creation of computer avatars that convey human emotion (Kalra et al., 1993; Ruttkay et al., 2003). Appropriate emotional expression may help make a virtual salesperson more convincing. In computer-aided instruction, the expression on a human face could reward or discourage.

The Object Display and Object-Based Diagrams

Wickens (1992) is primarily responsible for the concept of an *object display* as a graphical device employing "a single contoured object" to integrate a large number of separate variables. Wickens theorized that mapping many data variables onto a single object will guarantee that these variables are processed together, in parallel. This approach, he claimed, has two distinct advantages. The first is that the display can reduce visual clutter by integrating the variables into a single visual object. The second is that the object display makes it easier for an operator to integrate multiple sources of information.

Among the earlier examples of object displays are Chernoff faces, named after their inventor, Herman Chernoff (1973). In this technique, a simplified image of a human face is used as a display. Examples are shown in Figure 7.11. To turn a face into a display, data variables are mapped to different facial features, such as the length of the nose, the curvature of the mouth, the size of the eye, the shape of the head, etc. There are good psychological reasons for choosing what might seem to be a rather whimsical display object. Faces are probably the most important class of objects in the human environment. Even newborn babies can rapidly distinguish

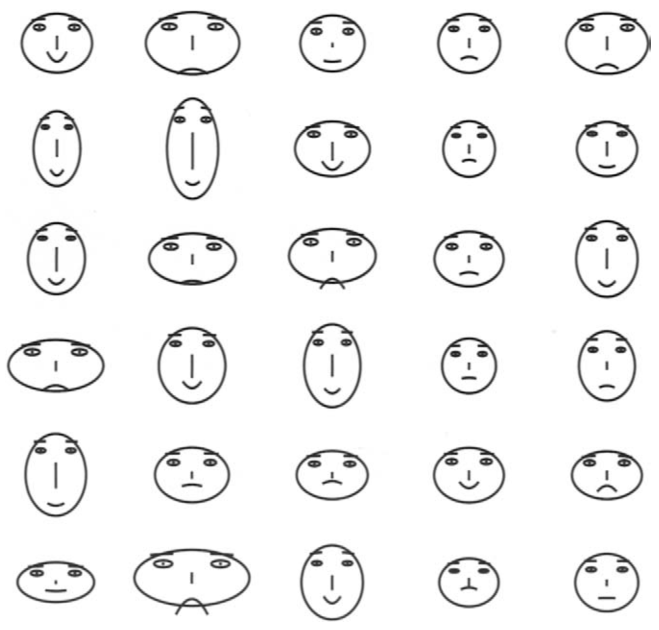

Figure 7.11 Chernoff faces. Different data variables are mapped to the sizes and shapes of different facial features.

faces from nonfaces with scrambled features, suggesting that we may have special neural hardware for dealing with faces. Jacob et al. (1976) carried out a classification task using a series of displays that were progressively more objectlike. The displays included Chernoff faces, tables, star plots, and the whisker plots described in Chapter 5. They found that the more objectlike displays, including Chernoff face plots, enabled faster, more accurate classification.

Chernoff faces have not generally been adopted in practical visualization applications. The main reason for this may be the idiosyncratic nature of faces. When data is mapped to faces, many kinds of perceptual interactions can occur. Sometimes the combination of variables will result in a particular stereotypical face, perhaps a happy face or a sad face, and this will be identified more readily. In addition, there are undoubtedly great differences in our sensitivity to the different features. We may be more sensitive to the curvature of the mouth than to the height of the eyebrows, for example. This means that the perceptual space of Chernoff faces is likely to be extremely nonlinear. In addition, there are almost certainly many uncharted interactions between facial features, and these are likely to vary from one viewer to another.

Often, object displays will be most effective when the components of the objects have a natural or metaphorical relationship to the data being represented. For example, Figure 7.12 illustrates how a storage vessel in a chemical plant might be represented using both a conventional bar chart and a customized object display. The variables in the object diagram are represented as follows:

- Size of cylinder represents tank capacity.
- Height of liquid represents volume of material stored.
- Texture of liquid represents the chemical composition.
- Color of liquid represents liquid temperature.
- Diameter of pipe represents outflow capacity.
- Status of the valve and thickness of the outgoing fluid stream represent rate at which liquid is being drawn from the tank.

In this example, the object display has a number of clear advantages. It can reduce accidental misreadings of data values. Mistakes are less likely because components act as their own descriptive icons. In addition, the structural architecture of the system and the connections between system components are always visible, and this may help in diagnosing the causes and effects of problems. Conversely, the disadvantage of object displays is that they lack generality. Each display must be custom-designed for the particular application and, ideally, should be validated with a user population to ensure that the data representation is clear and properly interpreted. This requires far more effort than displaying data as a table of numbers or a simple bar chart.

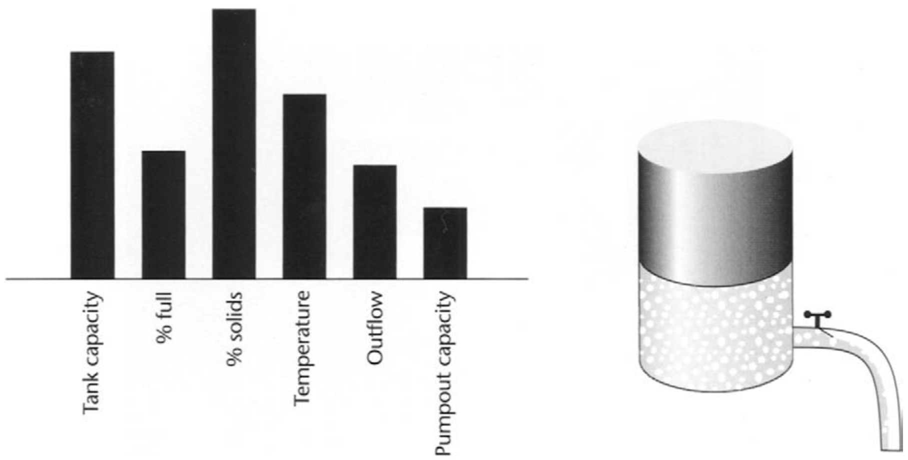

Figure 7.12 Two representations of the same data. The object diagram on the right combines six variables in an easily interpreted, cohesive representation.

The Geon Diagram

Biederman's geon theory, outlined earlier, can be applied directly to object display design. If cylinders and cones are indeed perceptual primitives, it will make sense to construct diagrams using these geon elements. This should make the diagrams easy to interpret if a good mapping can be found from the data to a geon structure. The geon diagram concept is illustrated in Figure 7.13(a). Geons are used to represent the major components of a compound data object, whereas the architecture of the data object is represented by the structural skeleton linking the geons. The size of a geon becomes a natural metaphor for the relative importance of a data entity, or its complexity or relative value. The strength of the connections between the components is given by the necklike linking structures. Additional attributes of entities and relationships can be coded by coloring and texturing them.

We evaluated the geon diagram concept in a comparison with Unified Modeling Language (UML) diagrams (Irani et al., 2001). UML is a widely used, standardized diagramming notation for representing complex systems. Equivalent diagrams were constructed by matching geon elements to UML elements (see Figure 7.13). We found that when the task involved rapid identification of substructures in a larger diagram, participants performed both faster and with only half the errors using the geon diagrams. Another experiment showed that geon diagrams were easier to remember.

In Biederman's theory, surface properties of geons, such as their colors and textures, are secondary characteristics. This makes it natural to use the surface color and texture of the geon to represent data attributes of a data object. The important mappings between data and a geon diagram are as follows:

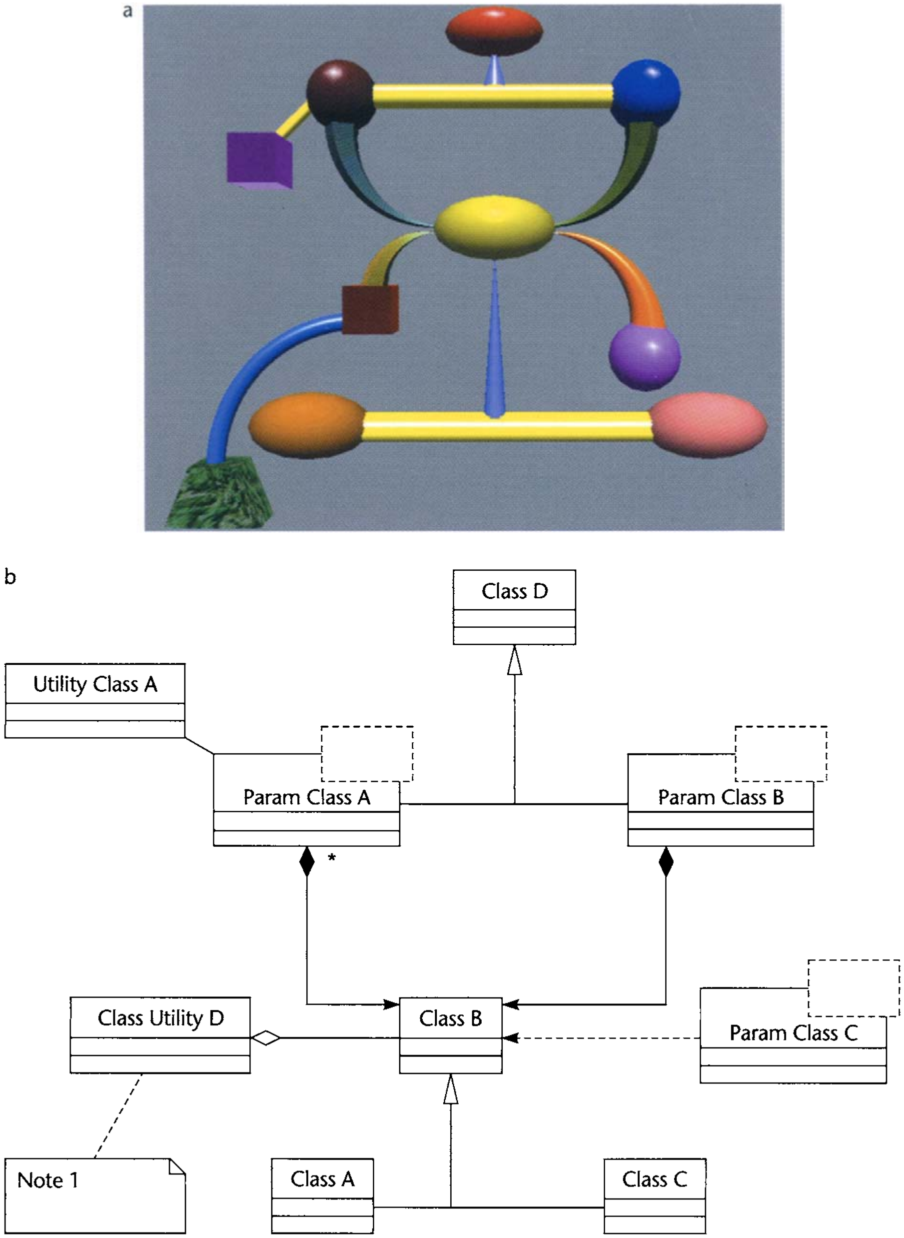

Figure 7.13 (a) A geon diagram constructed using a subset of Biederman's geon primitives. The primitive elements can also be color-coded and textured. (b) A Unified Modeling Language (UML) equivalent.

Major components of a complex data object	→	Geons
Architectural links between data object components	→	Limbs consisting of elongated geons—connections between limbs reflect architectural structure of data
Minor subcomponents	→	Geon appendices—small geon components attached to larger geons
Component attributes	→	Geon color, texture, and symbology mapped onto geons

Although the geon diagram is a 3D representation, there are reasons to pay special attention to the way it is laid out in 2D in the *x,y* plane. As discussed earlier, some silhouettes are especially effective in allowing the visual system to extract object structure. Thus, a common-sense design rule is to lay out structural components principally on a single plane. A diagramming method resembling the bas-relief stone carvings common in classical Rome and Greece may be optimal. Such carvings contain careful 3D modeling of the component objects, combined with only limited depth and a mainly planar layout.

Abstract semantics may be expressible, in a natural way, through the way geons are interconnected. In the everyday environment, there is meaning to the relative positioning of objects that is understood at a deep, possibly innate level. Because of gravity, *above* is different from *below*. If one object is inside another, it is perceived as either contained by that other object or a part of it. Irani et al. (2001) suggested that the semantics inherent in the different kinds of relationships of real-world objects might be applied to diagramming abstract concepts. Based on this idea, the researchers developed a set of graphical representations of abstract concepts. Some of the more successful of these mappings are illustrated in Figure 7.14 and listed as follows.

- Sometimes we wish to show different *instances* of the same generic object. Geon theory predicts that having the same shape should be the best way of doing this. Geon shape is dominant over color, which is a secondary attribute. Thus the elbow shapes in Figure 7.14(a) are seen as two instances of the same object, whereas the two green objects are not.
- Having an object inside another transparent object is a natural representation of a *part-of* relationship. The inside objects seem part of the outside objects, as in Figure 7.14(b).
- One object above and touching another, as shown in Figure 7.14(c), is easily understood as representing a *dependency* relationship.
- A thick bar between two objects is a natural representation of a *strong* relationship between two objects; a thinner, transparent bar represents a *weak* relationship. See Figure 7.14(d).

Perceiving the Surface Shapes of Objects

Not all things in the world are made up of closed, discrete components like geons. For example, there are undulating terrains that have no clearly separable components. Although to some extent

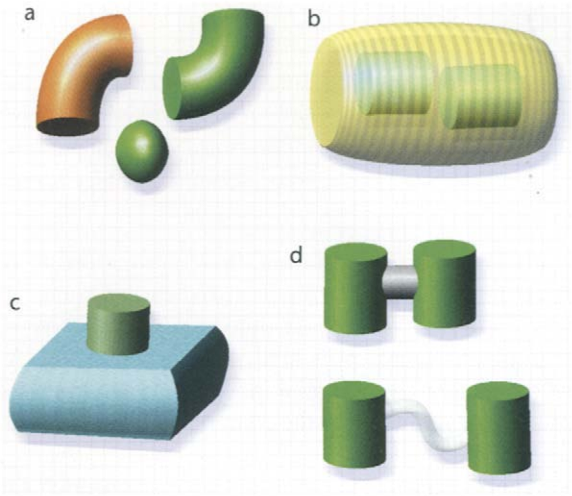

Figure 7.14 Certain spatial relationships between objects can readily represent abstract concepts. (a) That objects belong to the same class is better shown by shape than by color. (b) A part-of relationship. (c) A dependency relationship. (d) Strong and weak relationships.

we can decompose such a landscape into features such as hills and valleys, these are not essential to perceiving the shape of any given area of the surface. Examples of continuous surfaces that are important in visualization include digital elevation maps representing the topography of the land or the ocean floor, maps of physical properties of the environment, such as pressure and temperature, and maps representing mathematical functions that are only distantly related to the raw data. The general terms for this class of data object are *two-dimensional scalar field* and *univariate map*. The two traditional methods for displaying scalar field information are the contour map, which originated in cartography, and the pseudocolor map, discussed in Chapter 4.

Spatial Cues for Representing Scalar Fields

From a Gibsonian point of view, the obvious way to represent a univariate map is to make it into a physical surface in the environment. Some researchers occasionally do just this; they construct plaster or foam models of data surfaces. But the next best thing may be to use computer graphics techniques to shade the data surface with a simulated light source and give it a simulated color and texture to make it look like a real physical surface. Such a simulated surface can be viewed using a stereoscopic viewing apparatus, by creating different perspective images, one for each eye. These techniques have become so successful that the auto industry is using them to design car bodies in place of the full-sized clay models that were once constructed by hand to show the curves of a design. The results have been huge cost savings and a considerably accelerated design process.

An important issue in the creation of univariate maps is determining how to represent surface shape most effectively. Four principal sets of visual cues for surface shape perception have been studied: shading models, surface texture, stereoscopic depth, and motion parallax.

Shading Models

The basic shading model used in computer graphics to represent the interaction of light with surfaces has already been discussed in Chapter 2. It has four basic components, as follows:

Lambertian shading: Light reflected from a surface equally in all directions

Specular shading: The highlights reflected from a glossy surface

Ambient shading: Light coming from the surrounding environment

Cast shadows: Shadows cast by an object, either on itself or on other objects

Figure 7.15 illustrates the shading model, complete with cast shadows, applied to a digital elevation map of San Francisco Bay. As can be seen, even this simple lighting model is capable of producing a dramatic image of a surface topography. A key question in choosing a shading model for data visualization is not its degree of realism, but how well it reveals the surface shape. There is some evidence that more sophisticated lighting may be harmful in representing surfaces.

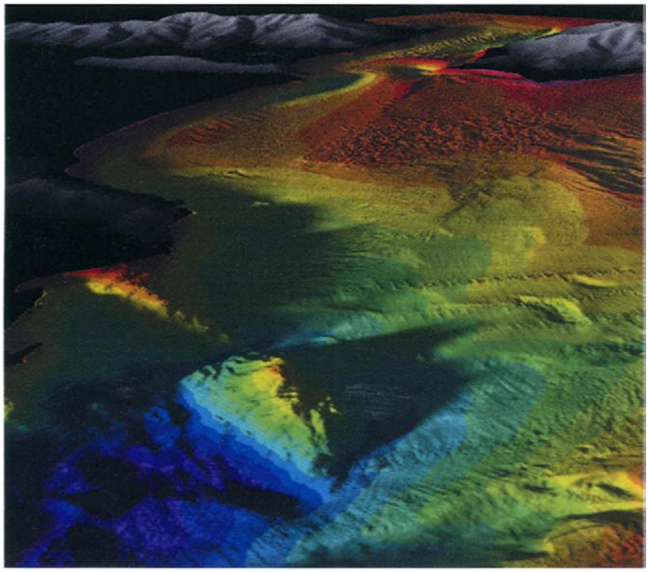

Figure 7.15 A shaded representation of San Francisco Bay, shown as if the water had been drained out of it. *Data courtesy of Jim Gardiner, U.S. Geological Survey. Image constructed using IVS Fledermaus software.*

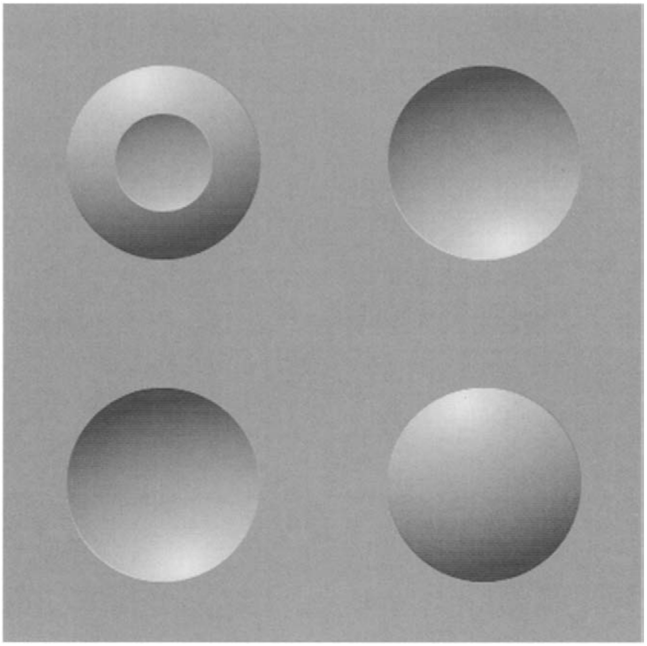

Figure 7.16 The brain generally assumes that lighting comes from above. The bumps in this image become hollows when the picture is turned upside down.

Experiments by Ramachandran (1988) suggest that the brain assumes a *single* light source from *above* in determining whether a particular shaded area is a bump or a hollow. (See Figure 7.16.) The kinds of complex shadows that result from multiple light sources and radiosity modeling may be visually confusing rather than helpful. Chapter 8 presents evidence that cast shadows provide spatial information relevant to the layout of objects in space rather than their surface shapes.

Surface Texture

Surfaces in nature are generally textured. Gibson (1986) took the position that surface texture is an essential property of a surface. A nontextured surface, he said, is merely a patch of light. The way in which textures wrap around surfaces can provide valuable information about surface shape.

Texturing surfaces is especially important when they are viewed stereoscopically. This becomes obvious if we consider that a uniform nontextured polygon contains no *internal* stereoscopic information about the surface it represents. Under uniform lighting conditions, such a surface also contains no orientation information. When a polygon is textured, every texture

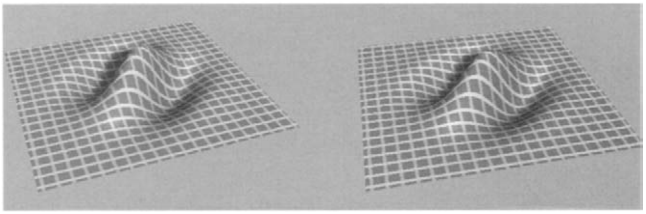

Figure 7.17 A stereo pair showing a textured surface.

element provides stereoscopic depth information relative to neighboring points. Figure 7.17 shows a stereoscopic pair of images representing a textured surface.

Without texture, it is usually impossible to distinguish one transparent curved surface from another transparent curved surface lying beneath it. Figure 7.18 shows an illustration from Interrante et al. (1997) containing experimental see-through textures designed to reveal one curved surface lying above another. The concept of *laciness*, discussed in Chapter 6, is relevant here, because it tells us something about how to make layers visually distinct. Stereoscopic viewing considerably enhances our ability to see one surface through another, semitransparent one.

Integration of Cues for Surface Shape

Given the many factors that may be involved in surface shape perception, the question arises as to which of them are most helpful. To study this problem, Norman et al. (1995) used computer graphics to render smoothly shaded rounded objects like the one shown in Figure 7.19. They manipulated the entire list of variables given above—specular shading, Lambertian shading, texture, stereo, and motion parallax—in a multifactor experiment. Stereo and motion were studied only in combination with the other cues because without shading or texture, neither stereo nor motion cues can be effective. The subjects' task was to indicate surface orientation at a number of selected points by manipulating the 3D glyph shown in Figure 7.20.

Norman et al. found *all* of the cues they studied to be useful in perceiving surface orientation, but the relative importance of the cues differed from one subject to another. For some subjects, motion appeared to be the stronger cue; for others, stereo was stronger. A summary of their results with motion and stereo data combined is given in Figure 7.21. Motion and stereo both reduced errors dramatically when used in combination with *any* of the surface representations. Overall, the combination of shading (either specular or Lambertian) with either stereo or motion was either the best or nearly the best combination for all the subjects.

There have been other studies of the relative importance of different cues to the perception of surface shape. Todd and Mingolla (1983) found surface texture to be more effective in determining surface shape than either Lambertian shading or specular shading. However, because of the lack of a convincing general theory for the combination of spatial cues, it is difficult to generalize from experiments such as this. Many of the results may be valid only for specific textures

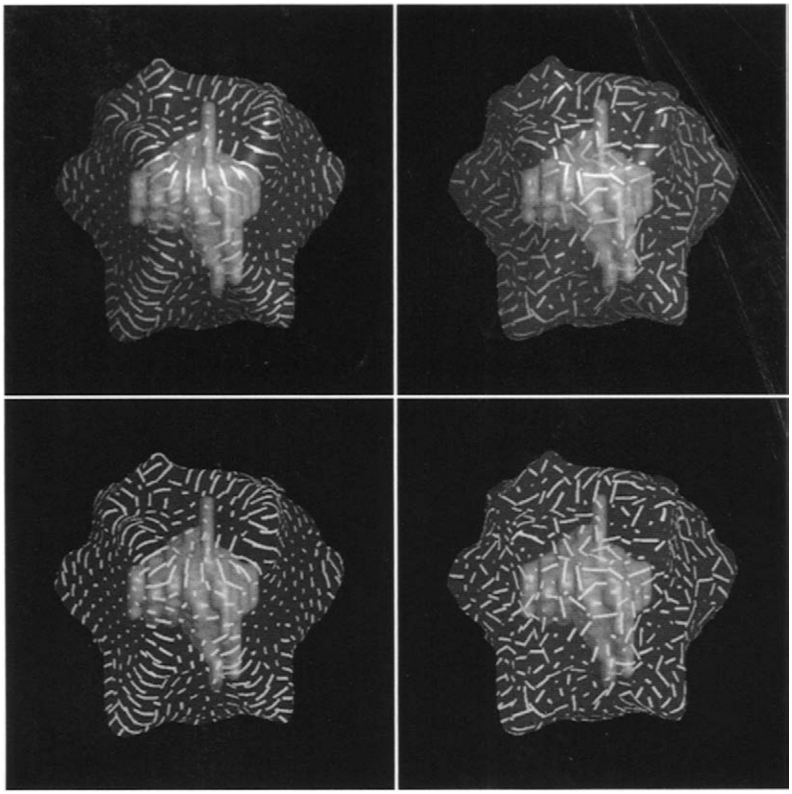

Figure 7.18 Texture designed to reveal surface shape. *From Interrante et al. (1997).*

used, for example. The fact that there are large individual differences is another barrier to reaching general conclusions. Random textures, such as those used by Norman et al. (1995), may not be as effective in revealing shape as texture that follows the surface in some way (Interrante et al., 1997; Kim et al., 1993). For these reasons, it is not meaningful to make general statements such as "Lambertian shading is more useful than texture." The values of the different cues will also depend on the specific task. For example, specular highlights can be extremely useful in revealing fine surface details, as when a light is used to show scratches on glass. At other times, highlights will obscure patterns of surface color.

Interaction of Shading and Contour

The boundary contours of objects can interact with surface shading to change dramatically the perception of surface shape. Figure 7.22 is adapted from Ramachandran (1988). It shows two

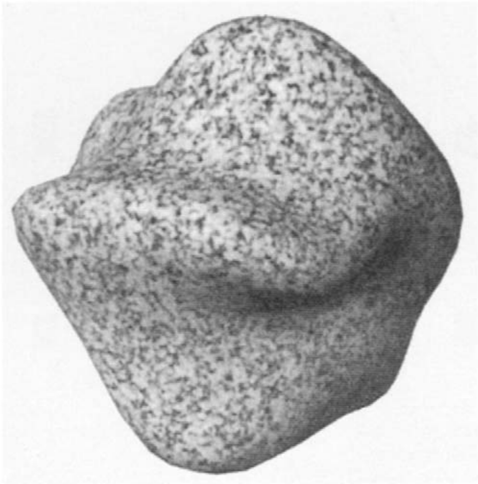

Figure 7.19 Textured, shaded, irregular objects were used by Norman et al. (1995) in experiments to determine which visual information contributes most to the perception of surface shape.

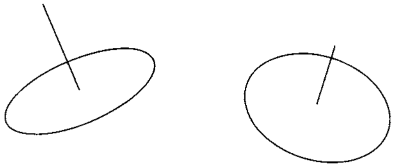

Figure 7.20 Interactive glyph used by Norman et al. (1995) to measure perception of surface orientation.

shapes that have exactly the same shading but different silhouette contours. The combination of silhouette contour information with shading information is convincing in both cases, but the surface shapes that are perceived are very different. This tells us that shape-from-shading information is inherently ambiguous; it can be interpreted in different ways, depending on the contours.

Contours that are drawn on a shaded surface can also drastically alter the perceived shape of that surface. Figure 7.23 has added shaded bands that provide internal contour information. As in Figure 7.22, the actual pattern of shading within each of the two images, and within the bands, is the same. It is the contour information that makes one surface shape appear so different from the other. This technique can be used directly in displaying shaded surfaces to make a shape easier to perceive.

One of the most common ways to represent surfaces is to use a contour map. A contour map is a plan view representation of a surface with isoheight contours, usually spaced at regular

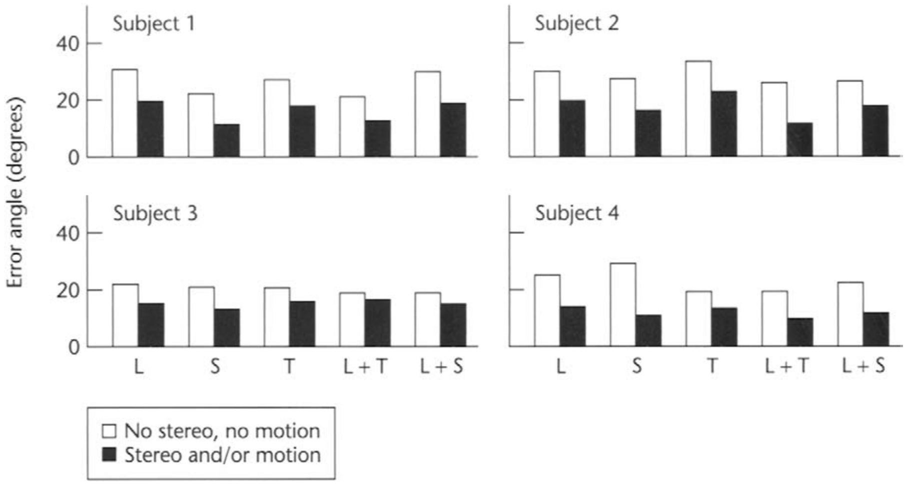

Figure 7.21 Results of the study of shape perception by Norman et al. (1995). The average errors in adjusted orientation are shown for five different surface representations. The different representations are labeled as follows: (L) Lambertian shading, (S) specular highlight shading, (T) texture with no shading, (L + T) Lambertian shading with texture, and (L + S) Lambertian shading with specular highlights. The four sets of histograms represent results from four different subjects.

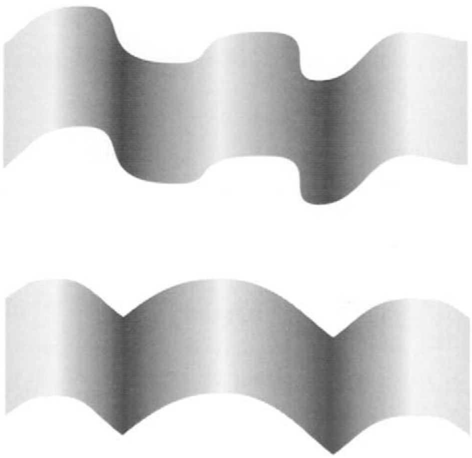

Figure 7.22 When scanned from left to right, the sequences of gray values in these two patterns are identical. The external contour interacts with the shading information to produce the perception of two very differently shaped surfaces.

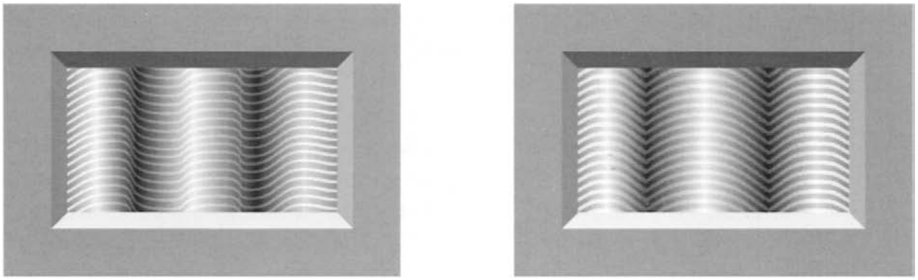

Figure 7.23 The left-to-right gray sequence in these patterns is identical. The internal contours interact with the shading information to produce the perception of two very differently shaped surfaces.

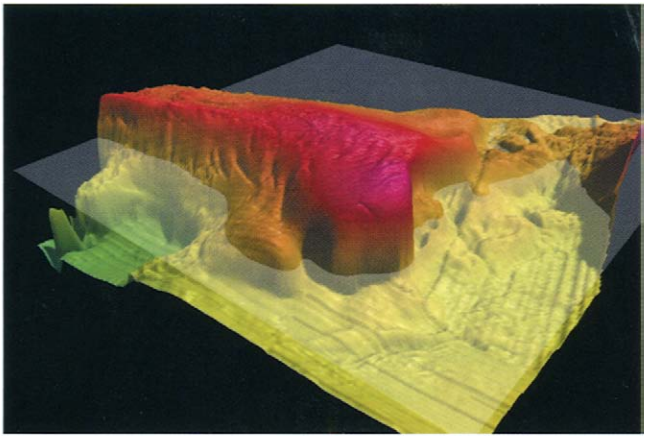

Figure 7.24 A contour is created by the intersection of a plane with a scalar field.

intervals. Conceptually, each contour can be thought of as the line of intersection of a horizontal plane with a particular value in a scalar height field, as illustrated in Figure 7.24. Although reading contour maps is a skill that requires practice and experience, contour maps should not necessarily be regarded as entirely arbitrary graphical conventions. Contours are visually ambiguous with respect to such things as degree of slope and direction of slope; this information is given only in the printed labels that are attached to them. However, it is likely that the contours in contour maps get at least some of their expressive power because they provide a limited perceptual code. As we have seen, both occluding (silhouette) contours and surface contours are effective in providing shape information. Although contour-map contours are not silhouettes, they obey one of the cognitive restrictions that Marr (1982) proposed for occluding contours, namely, that contours are assumed to be planar. They also provide texture gradient information. Thus, contour maps are a good example of a hybrid code; they make use of a perceptual mechanism, and they are also partly conventional.

There are many ways to make oriented textures conform to a surface. Texture lines can be constructed to follow the fall-line (down slope), to be horizontal contours, to be at right angles to maximum curvature direction, or to be orthogonal to the line of site of a viewer, to present a few examples. Kim et al. (1993) investigated combinations of first and second principal directions of curvature contours, as illustrated in Figure 7.25. All of the textured surfaces were artificially lit using standard computer graphics shading algorithms. Subjects made smaller errors in surface orientation judgments when two contour directions were used to form a mesh, as in Figure 7.25(a). Nevertheless, this study and Norman et al. (1995) found that errors *averaged* 20 degrees. This is surprisingly large and suggests that further gains are possible.

Guidelines for Displaying Surfaces

Taken together, the evidence suggests that to represent a surface clearly it may be possible to do better than simply create a photorealistic rendering of a scene using the most sophisticated techniques of computer graphics. A simplified lighting model—for example, a single light source located at infinity—may be more effective than complex rendering using multiple light sources. The importance of contours and the easy recognizability of cartoon representation suggest that an image may be enhanced for display purposes by using techniques that are nonrealistic.

Taking all these caveats into consideration, some guidelines may be useful for the typical case:

1. A simple lighting model, based on a single light source, should normally be used. The light source should be from above and to one side and infinitely distant.

2. Both Lambertian and moderate specular surface reflection should be modeled. More sophisticated lighting modeling, such as the interreflection of light between surfaces, should be avoided for reasons of clarity.

3. Specular reflection is especially useful in revealing fine surface detail. Because specular reflection depends on both the viewpoint and the position of the light source, the user should be given interactive control of the lighting direction, and the amount of specular reflection to specify where the highlights will appear.

4. Cast shadows should be used if possible, but only if the shadows do not interfere with other displayed information. The shadows should be computed to have blurred edges to make a clear distinction between shadow and surface pigment changes.

5. Surfaces should be textured, especially if they are to be viewed in stereo. However, the texturing should ideally be low-contrast so as not to interfere with shading information. Textures that have linear components are more likely to reveal surface shape than textures with randomly stippled patterns.

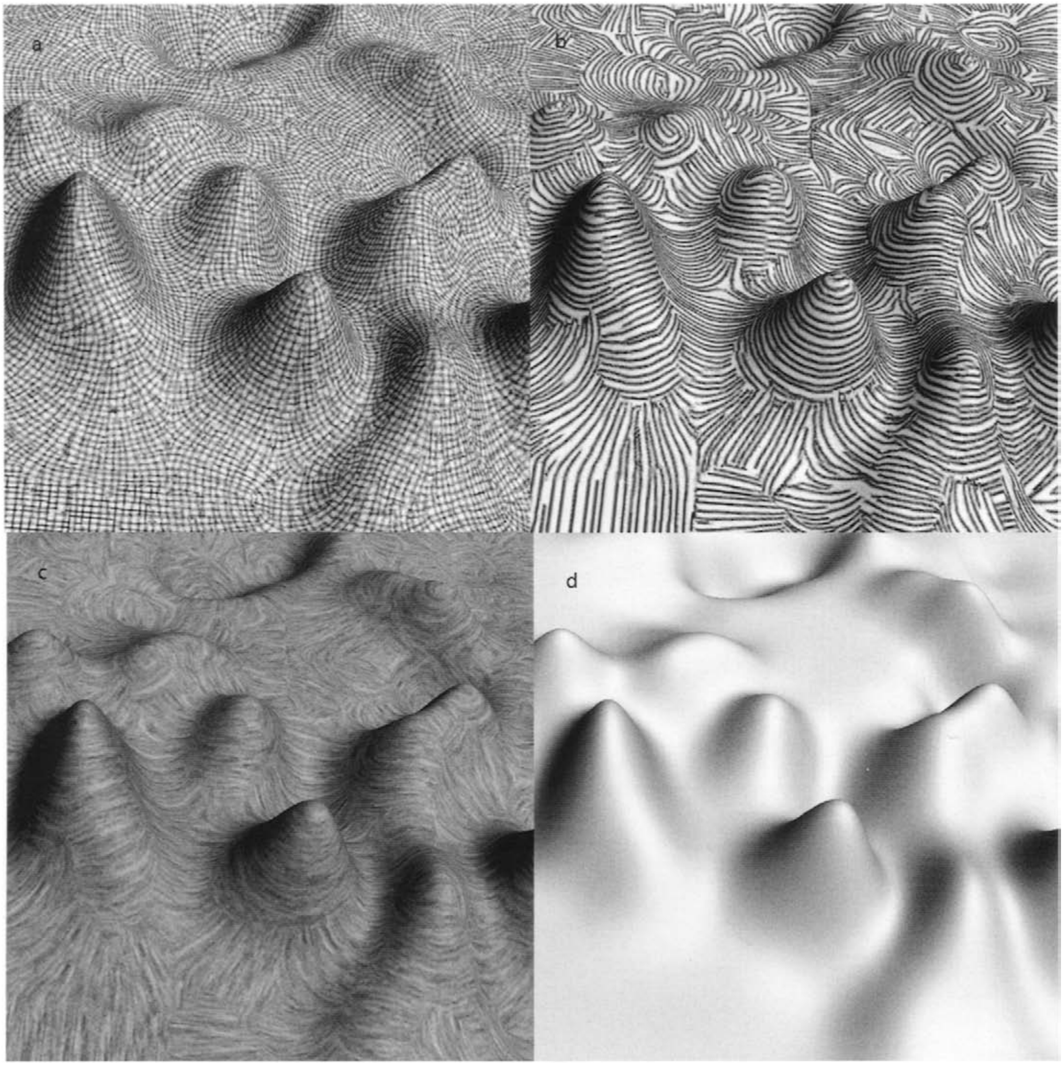

Figure 7.25 Surface-revealing texture patterns (Kim et al., 2003). (a) Two-directional texture pattern following first and second principal directions. (b) One-directional texture pattern following first principal curvature direction. (c) One-directional line-integral convolutions texture following first principal curvature direction. (d) No texture. *Reproduced with permission.*

6. Where appropriate hardware is available, both structure-from-motion (by rotating the surface) and stereoscopic viewing will enhance the user's understanding of 3D shape. These may be especially useful when one textured transparent surface overlays another.

Bivariate Maps: Lighting and Surface Color

In many cases, it is desirable to represent more than one continuous variable over a plane. This representation is called a *bivariate* or *multivariate map*. From the ecological optics perspective discussed in Chapter 1, the obvious bivariate map solution is to represent one of the variables as a shaded surface and the other as color coding on that surface. A third variable might use variations in the surface texture. These are the patterns we have evolved to perceive. An example is given in Figure 7.26, where one variable is a height map of the ocean floor and the surface color represents sonar backscatter strength. In this case, the thing being visualized is actually a physical 3D surface. However, the technique also works when both variables are abstract. For example, a radiation field can be expressed as a shaded height map and a temperature field can be represented by the surface color.

If this colored and shaded surface technique is used, some obvious tradeoffs must be observed. Since luminance is used to represent shape-from-shading by artificially illuminating the surface, we should not use luminance (at least not much) in coloring the surface. Therefore, the surface coloring must be done mainly using the chromatic opponent channels discussed in Chapter 4. But because of the inability of color to carry high–spatial frequency information, only rather gradual changes in color can be perceived. Therefore, in designing a multivariate surface display, rapidly changing information should always be mapped to luminance. For a more

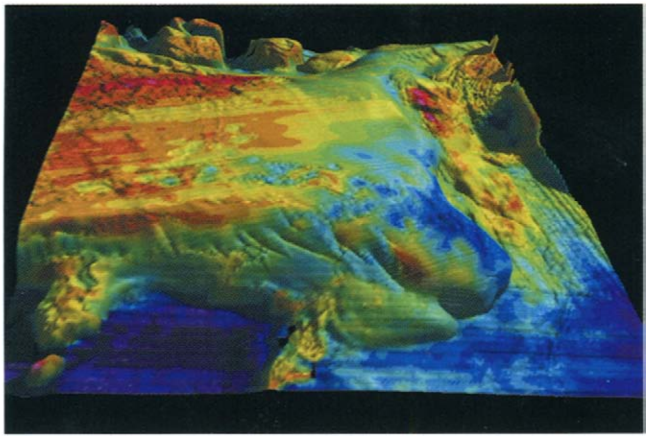

Figure 7.26 A bivariate map showing part of the Stellwagen Bank National Marine Sanctuary (Mayer et al., 1997). One variable shows angular response of sonar backscatter, color-coded and draped on the depth information given through shape-from-shading. *Courtesy of Larry Mayer.*

detailed discussion of these spatial tradeoffs, see Robertson and O'Callaghan (1988), Rogowitz and Treinish (1996), and Chapter 4 of this book.

A similar set of constraints applies to the use of visual texture. Normally it is advisable to use luminance contrast in displaying texture, but this will also tend to interfere with shape-from-shading information. Thus, if we use texture to convey information, we have less available visual bandwidth to express surface shape and surface color. We can gain a relatively clear and easily interpreted trivariate map, but only so long as we do not need to express a great deal of detail. Using color, texture, and shape-from-shading to display different continuous variables does not increase the total amount of information that can be displayed per unit area, but it does allow multiple map variables to be independently perceived.

Cushion Maps

The treemap visualization technique was introduced in Chapter 6 and illustrated in Figure 6.36 (Johnson and Shneiderman, 1991). As discussed, a problem with treemaps is that they do not convey tree structure well, although they are extremely good at showing the sizes and groupings of the leaf nodes. An interesting solution devised by van Wijk and van de Wetering (1999) makes use of shading. They applied a hierarchical shading model to the treemap so that areas representing large branches are given an overall shading. Regions representing smaller branches are given their own shading within the overall shading. This is repeated down to the leaf nodes, which have the smallest scale shading. The illustration shown in Figure 7.27 shows a computer file system. As can be seen, the hierarchical structure of the system is quite clear.

Integration

In this chapter, we have seen a number of ways in which different spatial variables interact to help us recognize objects, their structures, and their surfaces. However, there has been no discussion of how the brain organizes these different kinds of information. What is the method by which the shape, color, size, texture, structure, and other attributes of an object are stored and indexed? Unfortunately, this is still a largely open question. Only some highly speculative theories exist.

One suggestion is the theoretical concept of the *object file*, introduced by Kahneman and Henik (1981) to account for human perceptual organization. An object file is a temporary cognitive structure that stores or indexes all aspects of an object: its color, size, orientation, texture, and even its name and other semantic links (Kahneman et al., 1992). An object file can be thought of as a cognitive data structure that maintains links to all the perceived attributes of an object. An object file allows us to keep track of objects in the visual field, and from one fixation to the

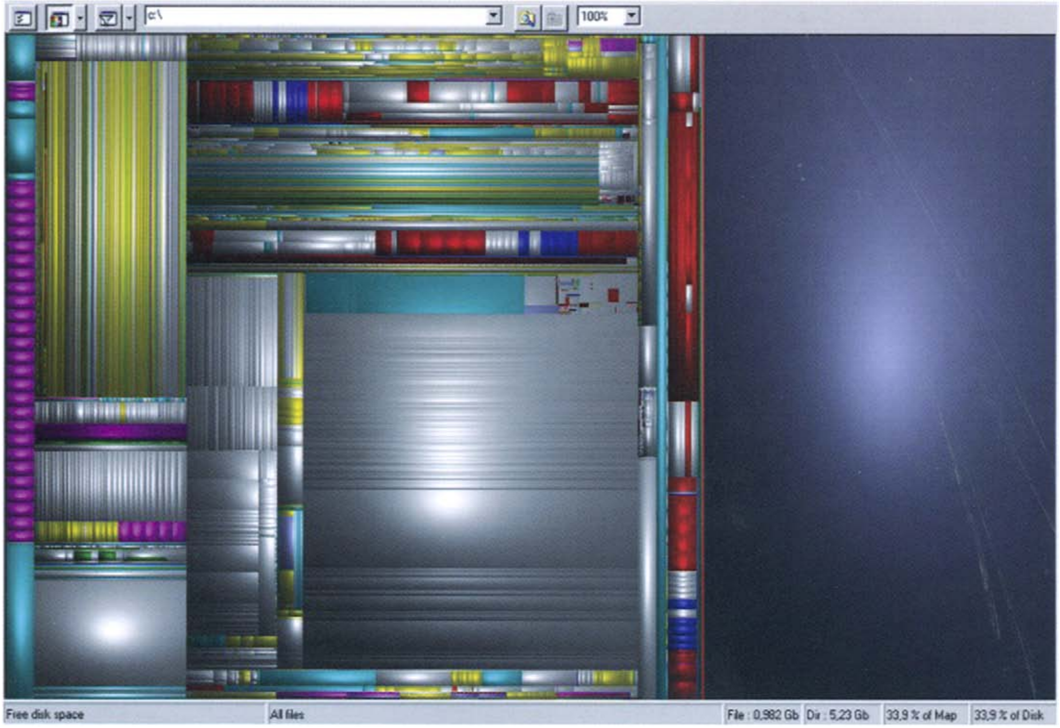

Figure 7.27 The cushion map is a variation on a treemap that uses shape-from-shading information to reveal hierarchical structure. *Courtesy of Jack van Wijke.*

next when they temporarily disappear behind other objects. Work by Pylyshyn and Storm (1988) and Yantis (1992) suggests that only a small number of visual objects, somewhere between two and four, can be maintained simultaneously in this way. Because of this, the display designer should drastically restrict the number of complex objects that are required simultaneously for any complex decision-making task.

Because both linguistic and visual information is included in the object file, it explains a number of well known psychological effects. One effect is that almost any information about an object, either visual or verbal, can be used to prime for that object. If there were a strong separation between visual and verbal information, we would not expect a verbal priming cue, for example, the word *bark*, to make it easier to identify a picture of a dog. But in fact, verbal priming does improve object recognition, at least under certain circumstances.

However, as discussed earlier in this chapter, there are also many priming effects that are strongest within a sensory modality; this is part of the evidence for separate verbal and visual processing centers. The concept of the object data structure also accounts for interference effects.

In the Stroop effect, subjects read a list of color words such as *red*, *yellow*, *green* and *blue* (Stroop, 1935). If the words themselves are printed with colored inks and the colors do not match the word meanings—for example, the word *blue* is colored red—people read more slowly. This shows that visual and verbal information must be integrated at some level, perhaps in something like Kahneman and Henik's object file.

Speculating further, the cognitive object file also provides an explanation of why object displays can be so effective. Essentially, the object display is the graphical analog of a cognitive object file. However, the strong grouping afforded by an object display can be a double-edged sword. A particular object display may suit one purpose but be counterproductive for another. Object-based displays are likely to be most useful when the goal is to give an unequivocal message about the relationship of certain data variables. For example, when the goal is to represent a number of pieces of information related to a part of a chemical plant, the object display can be clear and unambiguous. Conversely, when the goal is information discovery, the object display may not be useful because it will be strongly biased toward a particular structure. Other, more abstract representations will be better because they more readily afford multiple interpretations. Chapter 9 offers more discussion of the relationship between verbal and visual information and presents a number of rules for integrating the two kinds of information.

Conclusion

The notion of a visual object is central to our understanding of the higher levels of visual processing. In a sense, the object can be thought of as the point at which the image becomes thought. Objects are units of cognition as well as things that are recognized in the environment.

There is strong evidence to support both viewpoint-dependent recognition of objects and the theory that the brain creates 3D structural models of objects. Therefore, in representing information as objects, both kinds of perceptually stored information should be taken into account. Even though data may be represented as a 3D structure, it is critical that this structure be laid out in such a way that it presents a clear 2D image. Special attention should be paid to silhouette information, and if objects are to be rapidly recognized, they should be presented in a familiar orientation.

Visual processing of objects is very different from the massive processing of low-level features described in Chapter 5. Only a very small number of complex visual objects, perhaps only one or two, can be held in mind at any given time. This makes it difficult to find novel patterns that are distributed over multiple objects. However, there is a kind of parallelism in object perception. Although only one visual object may be processed at a time, all the features of that object are processed together. This makes the object display the most powerful way of grouping disparate data elements together. Such a strong grouping effect may not always be desirable; it may inhibit the perception of patterns that are distributed across multiple objects. However, when strong visual integration is a requirement, the object display is likely to be the best solution.

Once we choose to represent visual objects in a data display, we encounter the problem of what degree of abstraction or realism should be employed. There is a tradeoff between literal realism, which leads to unequivocal object identification, and abstraction, which leads to more general-purpose displays. Most interesting is the possibility that we can create a kind of hyper-realism through our understanding of the mechanisms of perception. By using simplified lighting models and enhanced contours, together with carefully designed colors and textures, the important information in our data may be brought out with optimal clarity.