# Chapter Three: Perceiving and Representing Shape and Depth

To determine how to represent the shape and depth of a layered transparent surface most effectively with sparse, opaque texture, we need both to understand the various potential sources of shape and depth information available in a two-dimensional representation of a three-dimensional scene and to have some sense of how effectively our visual system might use this information to accurately infer the shape and depth relations of the depicted objects. The distinction between shape cues and depth cues has historically been somewhat blurred, and it has not been until very recently that a coherent theory for the definition of object *shape* in a global sense, distinct from local measurements of slant and tilt, has emerged [Koenderink 1990], although the question of how our perception of the global shape of a surface might be related to our perception of the depths and local orientations of that surface at isolated points is still not exactly clear [Bülthoff and Mallot 1988]. While there is insufficient time and space here to do justice to the vast extent of current knowledge about shape and depth perception, I have attempted in this chapter to review selected aspects of these topics that I hope will be useful for providing insights into methods for more effectively representing shape and depth information in computer-generated images.

## 3.1: Pictorial depth cues

Our perception of the relative depth, or distance from the viewpoint, of the various objects in a static, two-dimensional image of a three-dimensional scene may be influenced to various extents by several different factors, which are sometimes referred to as *pictorial* cues to depth because of their use by artists to convey a greater sense of depth in a flat medium.

Of most relevance to this dissertation, of course, is the topic of texture and the relationship between texture properties and perceived depth and shape. Because of the central importance of this particular issue, the texture gradient depth cue will be discussed separately, in section 3 of this chapter. A very brief description of the other pictorial depth cues is now given.

### 3.1.1: Occlusion

Probably the most important of the pictorial depth cues is occlusion, the obscuration of a portion of a more distant object's surface by another surface that is closer to the viewpoint. There is evidence that our visual system makes fundamental use of occlusion information to encode the relative depths of superimposed surfaces at a relatively early stage in visual processing, and that the occlusion *boundaries* are the key elements in conveying the depth relationships [Nakayama *et al.* 1989]. Figure 3.1, based on the demonstration and discussion of this concept by [Nakayama *et al.* 1989], illustrates the effect that properties of the occluding object can have on the perceptibility of the occluded one.

**Figure 3.1: An illustration, based on figures and discussion in [Nakayama *et al. 1989*], showing the effect of occlusion boundaries on the perception of relative depth and figural continuity. Left: an upper-case letter "B" occluded by a figure whose color cannot be distinguished from the background. Center: the same arrangement of figures, but with the edge of the occluding figure clearly visible and the occluding line defined sharply toward the interior and in a more blurred fashion toward the exterior. Right: the same arrangement of figures, but with the entire occluding figure filled in. (As a consequence of this filling in, the sharp edges of the occluding figure are no longer "enhanced".)**

### 3.1.2: Linear perspective

When we look through our eyes at the world, the three-dimensional information in the scene gets projected onto the two-dimensional surface of our retina. Mathematically, this transformation can be described by a *perspective projection*, a 3D-to-2D mapping that has several important features. The first of these is what is often referred to in the literature as "linear perspective" or the observation that lines which are parallel in the three-dimensional model will appear in the projected image to converge toward a single vanishing point as they recede into the distance. This phenomenon is illustrated in figure 3.2. In general, the effect will be most pronounced when the lines originate close to the viewpoint and extend a considerable distance away.

**Figure 3.2: A photograph illustrating linear perspective (the apparent convergence, in the two-dimensional projection, of lines known to be parallel in the three-dimensional scene).**

Parallel lines are obviously not the only lines whose relative orientations are "remapped" by a perspective projection, but the effects of perspective convergence may be most clearly represented by them. It becomes increasingly difficult to appreciate the "distortion" due to perspective projection when the represented objects are more distant from the viewpoint, have a more limited extent in depth, and are smoothly curving or of irregular or unfamiliar shape.

A perspective projection is clearly not the only possible 3D-to-2D mapping, and in some instances it has been argued that, for the purposes of representing certain types of information about a three-dimensional model, it is not necessarily the best type of mapping to use. A common alternative is to employ an *orthographic projection*, in which parallel lines remain parallel. Such projections can be somewhat easier to construct manually and interpret intellectually, and certain algorithms for generating images from three-dimensional models derive their computational efficiency from relying on the regularities inherent in this approach. We are not constrained, when constructing a computer-generated image, to represent the data exactly as it would appear if it sat before us as a physical entity; however, it is important to consider both the advantages and disadvantages of stylizing the display.

### 3.1.3: Relative familiar size

A second consequence of perspective (but not orthographic) projection is that as an object moves farther away from the viewpoint it will subtend a smaller visual angle on the retina. This means that more distant objects will have relatively smaller projected sizes, and this is the basic premise behind the pictorial depth cue referred to as "relative familiar size". The idea is that if one knows the actual relative sizes the depicted objects, then, in the absence of other cues to


**Figure 3.3: A pair of photographs illustrating appropriate and inappropriate situations for using the "relative familiar size" cue to depth. Left: The relative distances of the gondolas can be inferred from the differences in their relative sizes in this image, because their actual relative sizes are known a priori (or can be reasonably assumed). Right: The relative distances of the buildings in the foreground and the buildings in the background are not accurately reflected by their relative sizes in the image. (What we see in the foreground is a scale model of the city of Sevilla; the real city of Sevilla appears in the background.)**

depth, one can infer their relative depths from the differences in the sizes of their projections. This concept is illustrated in the lefthand image of figure 3.3, where the different sizes of the gondolas reflect their different distances from the viewpoint. Of course, if the actual relative sizes of the objects are not accurately known, no valid conclusions about their relative depths can be drawn from this size information, as the righthand image of figure 3.3 clearly shows.

Numerous psychophysical experiments and geometrical illusions provide evidence of an intrinsic connection between our perception of relative size and relative depth. Ittelson [1951a,b] conducted a series of experiments that quantitatively measured the effects of relative familiar size on perceived relative distance in the absence of other depth cues and verified that these relative size differences evoked an impression of differences in relative depth. Although he argues that the relationship between size and depth is a reciprocal one, other researchers have offered evidence that when alternative depth cues are present, particularly perspective convergence, they tend to dominate the perception, and to such an extent that an illusory misperception of both actual and relative size differences can result, as illustrated in figure 3.4.

**Figure 3.4: A photocollage, modeled after an illustration by Lehmann [1966] (cited in [Metzger 1975]), that attempts to demonstrate the subordinate role of relative size in the judgment of relative depth. (My aim was to enable a perception of the man on the left as either a giant standing in the depth plane of the farthest woman, or as a "twin" of the rightmost man, standing on a hill in the foreground. When such a picture is correctly composited, the former impression will dominate and the repeated man will appear to be of different sizes in his multiple locations in the image as well as in the scene it represents; these illusory size differences are not perceived when the latter impression is attained.)**

### 3.1.4: Relative height

When objects resting on a common horizontal ground plane are viewed from above, the objects that are more distant from the viewpoint will project to positions that are relatively higher in the view. This feature of 3D-to-2D projection, which can be observed in figures 3.2, 3.4 and 3.3-right (but not in figure 3.3-left) has been referred to as the "relative height" cue to depth, and is intrinsic to both perspective and orthographic projections. Painters have historically used relative height as a device for portraying depth in pictures, and a series of experiments by

McGurk and Jahoda [1974] showed that youngsters aged 4–10 appeared to interpret elevation as indicative of depth in simple pictures. Because relative height will *not* be a reliable cue to relative depth for objects that do not share a common, horizontal ground plane, it seems reasonable to assume that an estimation of the strength of this cue might be given by a measure of the extent to which observers are predisposed to making this figural assumption (that objects rest on a common ground plane) when it is not otherwise indicated or contradicted. Gibson [1950a] used the experimental set-up shown in figure 3.5 to illustrate the propensity of observers, in absence of information to the contrary, to perceive an object as resting on the ground plane over which it is superimposed.

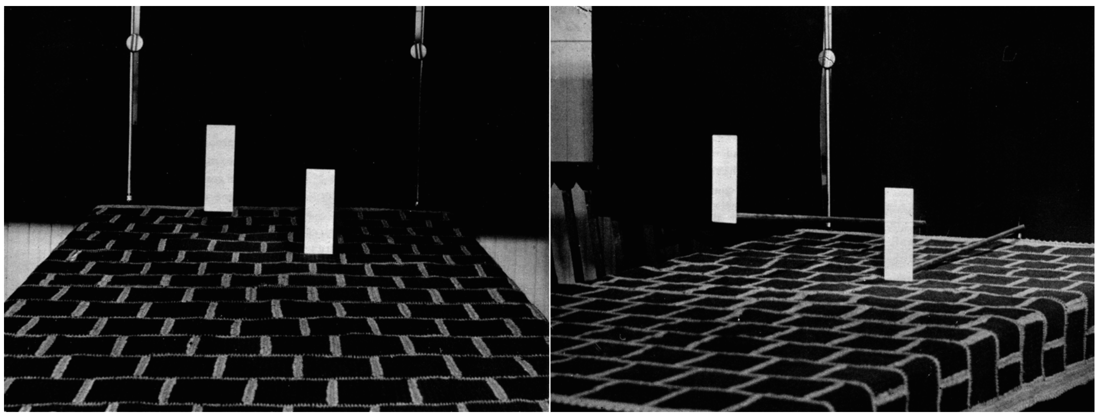


**Figure 3.5: Experimental set-up constructed by Gibson [1950a] to demonstrate the tendency we have to perceive an object as resting on the ground plane over which it is superimposed in a projected view. When the rectangles are viewed monocularly from a stationary, forward vantage point as illustrated in the image on the left, they appear to be of the same size but located at different depth distances. The actual arrangement of the objects is made clear in the image on the right. Both pictures are from [Gibson 1950a].**

### 3.1.5: Atmospheric attenuation

In many parts of the world the air is not clear but is filled with multitudes of tiny particles, from pollutants or from moisture. The accumulation of these particles over a distance of several miles (or, in some cases, yards) compromises the visibility of distant objects, causing them to appear somewhat "washed-out" and to have less contrast, both internally and in comparison with the sky. This phenomenon, also referred to as "aerial perspective", is illustrated in figure 3.6. O'Shea *et al.* [1994] and Egusa [1983] both cite numerous psychophysical experiments confirming that stimuli which have a lower luminance contrast with the background are perceived to be more distant, and Egusa [1983] also reports that objects with equal luminance but contrasting hues or saturations may be interpreted to lie in different depth planes. (I was not able to find a report specifically documenting an effect on depth perception of differences in internal luminance contrast, between the various regions of a pattern on a surface for example, but I did empirically verify this concept for myself.)

Livingstone and Hubel [1987] conducted a series of experiments in which they found that each of the classical depth cues (some of which we have already discussed and some of which will be discussed in later sections) ceased to function as such under conditions of equiluminance. Depth from stereo, shape from shading, shape from texture, depth from linear perspective, depth from motion, even depth from occlusion, all proved to be difficult or impossible to perceive when the images were defined by color rather than luminance differences. Since Livingstone and

**Figure 3.6: A photograph illustrating the "aerial perspective" cue to depth. One can observe the effects of atmospheric attenuation here in the decreased visibility of both the distant mountains (which are barely visible in the very center of the image) and the relatively nearer hillsides, in contrast with the vegetation in the immediate foreground. (This particular image, taken in Honduras during the dry season, represents a fairly extreme situation. By way of comparison, it should be noted that the aerial perspective cue is almost nonexistent in the landscape of figure 3.4. )**

Hubel report that many of these effects are visible only at equiluminance and that the exact colors that are perceived as equiluminant will vary between individual observers, I will not attempt to repeat these demonstrations in print. Nevertheless this finding probably more than any other underlines the central importance of contrast as a depth cue.

In a very similar vein, Cavanagh and Leclerc [1989] showed that shadow perception appears to be completely dependent on luminance information and remarkably impervious to unnatural perturbations of color, texture or binocular disparity. In a long series of experiments, they showed that in order for the shadow regions of an image to be perceived as such they had to be of relatively darker luminance than the surface on which they fell, both globally (in terms of the mean) and locally (across the shadow borders), and that, when these conditions were not met, the identical image subregions defined by any other means — including equiluminant color differences, texture motion (including the accretion and deletion of texture elements along the borders of the region), binocular disparity and equiluminant texture pattern differences — while clearly discriminable, could not be interpreted as shadows.

### 3.1.6: Spatial frequency, focus, depth of field

When we perceive objects in our everyday experience, we are rarely conscious of anything appearing to be "out of focus". Under ordinary circumstances, when we direct our attention to an object we fixate it in our fovea and automatically adjust the lens of our eye to bring the object into focus. Because of the great disparity between the depth of field attainable by our visual system and the depth of field available with a camera, however, it is not at all uncommon to observe photographs in which some part of the image is blurred due to its distance in depth from the focal point. As the image in figure 3.7 illustrates, we seem to be able to interpret relative differences in focus as indicating a disparity in relative depth.

**Figure 3.7: A photograph illustrating how an impression of distance can be intuitively understood from differences in relative focus, in an image with few complementary depth cues. The fact that the child in this picture is waving at an iguana in the distance and not reaching out to touch an iguana perched on a rock is immediately obvious, although the quantitative amount of depth distance between the two is not (it was a few yards).**

Experimental evidence substantiating the usefulness of focus as a depth cue was provided by Pentland [1985], who performed experiments verifying that people perceived a greater separation in depth between overlapping objects in images in which there was a larger gradient of focus between them. Further support for the perceptual significance of focal differences can be drawn from experiments by Klymenko and Weisstein [1986], who found that when the separate parts of an ambiguously organized image were filled with sine wave gratings of two different spatial frequencies but equivalent contrasts, the region filled with the higher spatial frequency pattern was seen as figure (rather than ground) a significantly greater percentage of the time. Although there is generally not enough information available in a limited-depth-of-field image of two non-overlapping surfaces to determine the sign of the depth distance between them purely on the basis of focus information, Marshall *et al.* [1996] described how it should be theoretically possible for observers to disambiguate the depth order of overlapping textured surface on the basis of the amount of blur along the occluding edge, and conducted experiments verifying this hypothesis. (About half of the subjects in these experiments consistently judged the blurred half-plane to be closer when the occluding edge was blurred and the focused half-plane to be closer when the occluding edge was sharp; for the other half of the subjects this trend, while detectable, was fairly modest.)

## 3.2: Oculomotor Cues to Depth

In addition to deriving information about the relative depths or depth order of surfaces from pictorial information in static two-dimensional images, our visual system is able to obtain direct information about the absolute distances of objects in our immediate environment from the physical sensations associated with the actions of the muscles of our eyes.

### 3.2.1: Accommodation

Typically described as a monocular depth cue because it is not explicitly dependent on the involvement of both eyes, accommodation refers to the action of changing the curvature of the lens of the eye so that an attended object can be viewed in focus. When the ciliary muscle, which surrounds the lens of the eye, is fully relaxed, the lens is at its flattest. This is the normal condition for viewing objects from a distance of more than about 10 feet [Kaufman 1974]. When we focus on closer objects, the ciliary muscle is contracted and the lens bulges outward with greater curvature. (A simple and effective way to elicit the sensation of changing accommodation is to close one eye, hold up two open fingers at arm's length, and shift attention back and forth from the fingers to whatever is visible between them in the background.) While the exact mechanism by which we interpret absolute depth from the sensations associated with accommodation is not fully understood, psychophysical research has confirmed that at least some human observers are able to make reasonable judgments about the absolute distances of appropriate target stimuli in situations in which accommodation is the only available cue to depth and the objects are located 16–50cm from the viewpoint. Fisher and Ciuffreda [1988] found that the amount of perceived depth was linearly related to the magnitude of the accommodative response, but with a slope of less than 1, and also showed that accommodative response closely matched accommodative stimulus only for targets with certain spatial frequency and contrast characteristics, explaining why some researchers have found accommodation to be a relatively weak source of depth information in other experiments. Although depth differences can be detected under some conditions on the basis of accommodative responses to stimuli located at distances greater than 50cm [Kaufman 1974], the relationship between accommodation and perceived distance will no longer be linear for distances in this far range and larger errors in the accuracy of the perceived distance can be expected [Fisher and Ciuffreda 1988].

### 3.2.2: Vergence

When we look at a nearby object with both of our eyes, the two eyes do not each point straight ahead but rather turn slightly inward, with the directions of sight from each eye converging at a focal point on the object being attended to. The depth distance of an object can be derived, mathematically, from the amount of inward turning of the eyes, or *binocular convergence*, together with the interocular distance, when both of these quantities are known. As was the case with accommodation, it is not exactly known how the kinesthetics of convergence are interpreted by the visual system to indicate absolute depth, but evidence cited by Kaufman [1974] appears to indicate that this phenomenon is mediated by the signals sent from the brain to the eye to control their direction of gaze rather than by sensations reported from the eye muscles themselves. A number of researchers, including von Hofsten [1976] and Komoda and Ono [1974], have reported psychophysical evidence indicating that subjects are able in a fairly consistent manner to infer the egocentric distances of objects purely on the basis of vergence cues, although it is a viable cue to relative distance only within a depth range of approximately 6m and, within this range, distances of less than about 40cm appear to be consistently overestimated and distances greater than about 40cm consistently underestimated Grant [1942]. It has long been observed that a more powerful perception of depth can be derived from accurately rendered perspective images if one eye is closed. ([Pirenne 1975] even cites a quote to this effect by Leonardo da Vinci in the 15th century.) It is possible that this phenomenon can be partially explained by the fact that binocularly-derived evidence for the flatness of the image is eliminated under the conditions of monocular viewing. Recent experiments by Enright [1987] support this conclusion, indicating that appropriately directed vergence responses can be evoked from perspective stimuli viewed monocularly. Grant [1942] found that in “incomplete” monocular stimuli (in which stimulation is presented to one eye only) although both eyes respond in a unified fashion, the response of the unstimulated eye is somewhat incomplete.

A strong connection has been shown between vergence and accommodation responses to stimuli. Under typical viewing conditions, a stimulus to one will evoke a coordinated response by both [Kersten and Legge 1983], and when they are both stimulated but in a contradictory fashion, the perceived depth tends towards an average of the stimulated depths [Grant 1942].

Although psychophysical evidence appears to indicate that accommodation and vergence, either separately or together, can provide useful depth information, it is at present unclear how we might hope to implement a visualization system in which an appropriate correspondence between the displayed image and the accommodation and vergence states of the eyes is maintained. As data representation technology improves and other sources of depth information are more faithfully represented, the impact on perceived depth of accommodative cues to the depth of the screen may assume increasing significance.

## 3.3: Shading

While shading has sometimes been described as a "pictorial depth cue", and some of its particular aspects, such as shadows, certainly fall under that category, I have chosen to discuss all of the phenomena of shading together, under a separate heading, because of the important role that shading information can play in communicating shape as well as depth. Following Yonas [1979], I am dividing my discussion of "shading cues" into two sections, treating the luminance attenuation on a surface due to light source occlusion ("shadows") separately from the luminance distribution on a surface due to illumination by a non-occluded light source ("shading"). The former is primarily useful as a cue to depth while the latter can provide supplementary information about surface orientation, curvature and shape.

### 3.3.1: Depth from shadows

In our everyday visual experience, many if not most of the objects that we encounter tend to be positioned so that they rest on top of an underlying surface rather than being suspended freely in the air. This precept does not always apply, however, to objects represented in computer-generated (or hand-drawn) images, which means that the exact location of an object in relation to the ground plane can be ambiguous in these pictures. As was mentioned earlier,

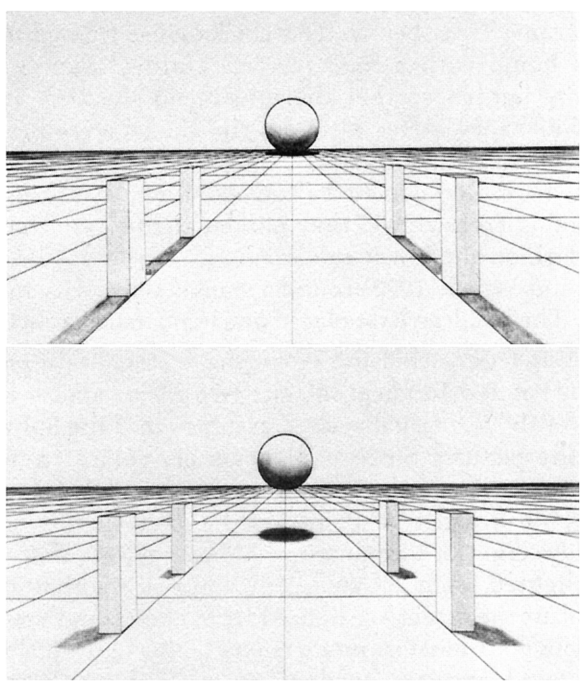


**Figure 3.8: An example of sample stimuli used by Yonas *et al.* [1978] to examine the influence of cast shadows on the perception of relative depth, height and size in pictorial representations.**

however, observers do appear to assume, in the absence of information to the contrary, that an object rests directly on top of whatever ground plane is indicated in the illustration [Gibson 1950a]. Artists have for many years used cast shadows as a pictorial device for enhancing the three-dimensional appearance of a scene (Yonas [1979] cites several historical examples), and psychophysical experiments have confirmed that the information provided by cast shadows can influence observers' perception of the position of an object relative to both the ground plane and the viewpoint. When Yonas *et al.* [1978] presented observers with variations of the stimuli shown in figure 3.8 and asked them to indicate which sphere was higher off the ground, adult subjects chose the sphere with the detached shadow about 80% of the time. They also chose this sphere as being the closest to them in about 80% of the cases, and as the smaller of the two in almost 90% of the cases.

Figure 3.9 shows an example of the usefulness of cast shadows for indicating depth in a "real world situation". In this case, the shadows indicate the relatively greater relief of the carving in the center panel.

**Figure 3.9: A photograph [Réunion des musées nationaux] of a leaf of a 6th century ivory diptych, in which variations in the extent of the relief are made apparent by the presence or absence of cast shadows.**

### 3.3.2: Shape from shading

The importance of shading as a source of information about surface shape in two-dimensional images is undeniable. Figure 3.10 illustrates an example of a stone relief in which shading is the only available cue to shape and depth. The importance of shading has been repeatedly demonstrated in the computer graphics literature, where considerable effort continues to be devoted to the problem of formulating the equations defining the illumination so that the subjectively most important characteristics of realistic lighting can be effectively and efficiently portrayed in computer-generated images.

In practice, the distribution of incident light reflected from the surface of an object and visible from a particular viewpoint will depend on a number of factors including the direction of view, the material properties of the object (whether the surface is smooth or rough, and in what

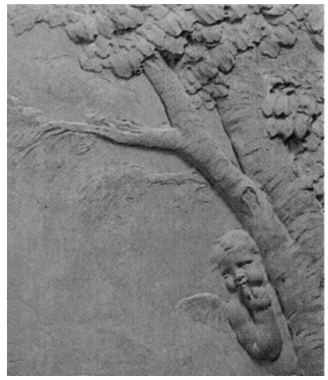


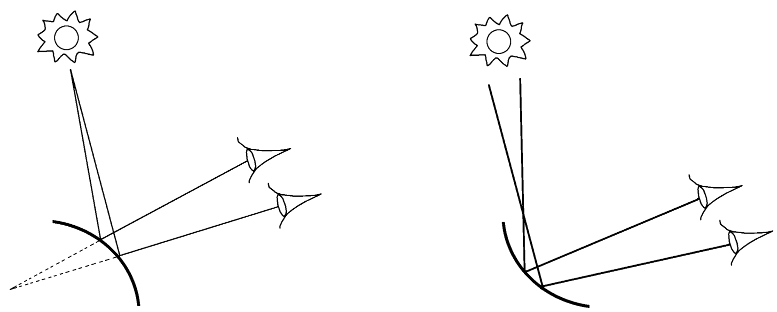
**Figure 3.10: A photograph [Réunion des musées nationaux] of an 18th century stone relief [Clodion 1782] in which shading is the only available cue to shape.**

various proportions it absorbs, transmits or reflects incident light) and the intensity, position (near or far) and direction of the light source(s), both primary and secondary. Because so many different variables combine in so many different ways to affect the pattern of illumination on objects that we see, interpreting shape from patterns of shading can get very complicated.

A great deal of effort has been invested by the computer vision community in coming up with algorithms to compute shape information from patterns of shading [cf. Horn and Brooks 1989]. Computational shape-from-shading methods typically solve for the direction of the surface normal at any given point on an object by "inverting" the illumination equation, relying on a combination of simplifying assumptions about the nature of the surface and the source of the illumination to constrain the possible set of solutions. The human visual system does not appear to "solve" the shape-from-shading problem in this way, however [Mingolla and Todd 1986]. Although simpler shading models are undeniably easier to deal with computationally, it doesn't necessarily follow from this that surface shape and depth are communicated "more clearly" under these restricted conditions, and there is little evidence to support the idea that the visual system assumes Lambertian shading. On the contrary, empirical and experimental evidence both seem to indicate that surface shape and depth may be communicated *more* effectively as the "complexity" of the shading model increases. In particular, the perception of shape from shading seems to be helped, if anything, rather than hindered by the representation of specular highlights.

Todd and Mingolla [1983] have found that the accuracy of observers' judgments of the curvature of cylindrical patches increases when specular highlights are incorporated into the shading model; in their experiments, curvature was significantly underestimated for all stimuli when shading was the only shape cue provided, but the severity of the underestimation was slightly less in the specular than in the diffuse cases. Mingolla and Todd [1986] have found that observers' judgments of local surface orientation are slightly better at points on elongated ellipsoidal surfaces where the specular component of illumination is relatively higher, although overall performance, measured in terms of orientation judgments across the entire surface, does not seem to be affected in a significant way by either the presence or absence of specular highlights.

Blake and Bülthoff [1991] have shown that human observers are able to disambiguate convex from concave surface shapes on the basis of the apparent location in depth of the specular highlight under conditions of stereo viewing. The surface depth indicated by stereo was flat for the stimuli in these experiments, and diffuse shading provided an ambiguous cue to surface shape. Figure 3.11 demonstrates why a specular highlight will seem to float in front of a concave surface and behind a convex one; it can appear either in front of or behind a hyperbolic surface, depending on the viewing geometry and surface orientation. Zisserman *et al.* [1989] show how shape information can be derived, computationally, from the patterns of motion, in response to viewpoint, of specular highlights on surfaces in two-dimensional images. Specular highlights will move relatively more slowly over a surface when the viewing direction changes if the surface is highly curved at the location where the specularity falls, and the direction of motion of a specular highlight will either follow the direction of motion of the observer, if the surface is convex, or oppose it, if the surface is concave. Although the separation in depth of specular highlights from the underlying surface may be useful for providing shape information, the fact that the specular highlights are not actually attached to the surface may make them potentially misleading as indicators of depth.



**convex: highlight appears behind** **concave: highlight appears in front**

**Figure 3.11: A diagram explaining why specular highlights are not, in general, perceived as lying on a curved surface but as floating in space either in front or in back of it.**

While cast shadows have been shown to be useful for disambiguating the position of an object relative to an underlying ground plane, the role of self-shadows in shape perception has not, as far as I can tell, been specifically investigated — possibly because the kinds of surfaces on which self-shadows occur with great frequency are generally more complex than the typical stimuli examined in psychophysical experiments. I speculate that there would turn out to be no particular perceptual advantage to *not* representing self-shadows when and where they are indicated. The real question of interest is how much of a difference it would make to take the trouble to represent them; this would have to be measured in the context of the goals of the particular presentation.

The role of secondary illumination, and in particular diffuse interreflections, is another topic that seems to have been largely ignored in discussions of how we perceive shape and depth from patterns of shading. Although diffuse interreflections might seem irrelevant to our perception of the local shape of an individual object, there are enough striking examples in the computer graphics literature of the differences in the appearances of complex scenes rendered with and without radiosity to at least raise questions about the prudence of dismissing this shading effect offhand. Figure 3.12 [Goral *et al.* 1984] illustrates one such example, for a very simple geometrical scene. In this case, the progressive darkening of surfaces towards the inner recess of the cube and along the diagonal lines where adjacent edges meet seems to be a very important cue, to both the concave nature of the configuration and the depth extent of the figure. Forsyth and Zisserman [1990] show that a significant amount of the irradiance measured on a surface may be due to mutual illumination, particularly when the surface is concave. Citing the seemingly intractable relationship between surface shape and irradiance in general, they propose that discontinuities in the distribution of irradiance may be more effective cues to shape than patterns of smoothly changing irradiance.

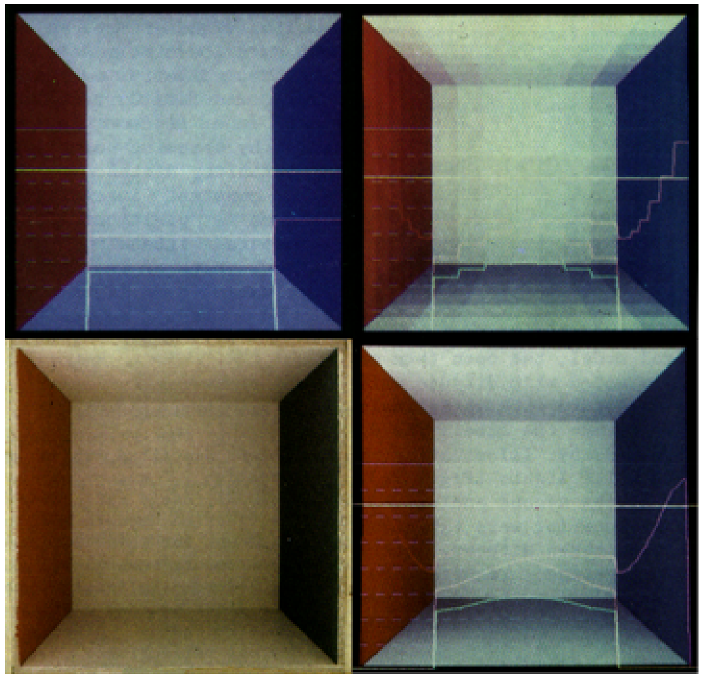


**Figure 3.12: An illustration, from [Goral *et al.* 1984], of the radiosity algorithm they proposed for modeling the interaction of light between diffuse surfaces. Upper left: flat shading. Upper right: after radiosity calculations using 49 patches. Lower right: after linear interpolation across the 49 patches. Lower left: a photograph of the simulated environment.**

Various techniques for using light to convey three-dimensional surfaces to their best advantage have been described by artists over the centuries, in particular the admonition to avoid lighting peoples' faces from below. Koenderink and van Doorn [1980] note that veridical shape perception may be easier in some light fields than others and that forms tend to appear particularly flat when the light field is isotropic (corresponding to illumination by a parallel light source along the line of sight). Johnston and Passmore [1994] specifically examined how observers' judgments of surface shape were affected by the direction of illumination and found that curvature discrimination tended to be more accurate under conditions of oblique lighting. They also measured slant discrimination at vertical light source angles of 25˚, 45˚ and 65˚ and found that observers' judgments of the direction (left vs. right) of slant were more accurate when the angle of illumination was closer to the line of sight. Despite the fact that shape judgments tend to vary in quality depending on the illumination direction, it is unclear to what extent knowledge of the direction of illumination might aid the veridical perception of shape from shading. Erens *et al.* [1993] report that subjects were able to disambiguate convex from concave surface patches in shape discrimination tasks when the direction of illumination was indicated by cast shadows. Mingolla and Todd [1986], however, found that observers' ability to correctly identify the direction of illumination from the pattern of shading across an ellipsoidal surface was not correlated with the accuracy with which they were able to make judgments about the local surface orientation and concluded that knowledge of the direction of illumination did not appear to be a prerequisite for this task.

Although gradients of shading appear to be play an essential role in indicating the existence of surface relief, there is considerable evidence that the human visual system uses shading as a secondary rather than primary source of information about shape — that shading, on its own, is a rather weak cue to surface shape.

Erens *et al.* [1993] present both theoretical and experimental evidence of substantial inherent ambiguities in the task of inferring shape from patterns of shading, describing how observers cannot (and have no basis for being able to) distinguish between convex, concave, and hyperbolic (saddle-shaped) surface patches when the direction of illumination is unknown, and that they cannot distinguish between elliptic and hyperbolic shapes on the basis of their shading even when the illumination direction is indicated by cast shadows.

This phenomenon of "depth inversion", or the tendency of convex surface features to appear to be concave and vice versa, has been studied extensively, and several different explanations for this illusion have been offered. Yonas *et al.* [1979] and Howard *et al.* [1990] have demonstrated that, in the absence of indications to the contrary, people tend to presume that surfaces are lighted from "above", where "above" means "from the top of the head" rather than "towards the ground". Reichel and Todd [1990] show that for surfaces that resemble height functions on a slanted plane, convex/concave reversals are perhaps better described by a preference by observers to perceive a surface as receding in depth towards the top rather than the bottom of the image, or a preference for "ground" as opposed to "ceiling" surfaces. This effect is illustrated in figure 3.13, in which a photograph of a crater, illuminated from above the surface but below the line of sight, is shown in both its original and upside-down orientations. Both Gregory [1970] and Ramachandran [1988] have demonstrated that depth inversion can also be attributed, in some cases, to a preference for perceiving a familiar object in its convex form, even when cues from perspective or motion or the direction of illumination contradict this interpretation. This "inversion" phenomenon has been noted for centuries. Both Wheatstone [1838] and Brewster [1826] remark on the phenomenon and cite reports of it dating from the mid-1700's, and von Helmholtz [1925] specifically discusses the mask illusion. I have verified this illusion for myself by viewing a child's plastic lion mask from the back side and found the depth inversion to be very robust, destroyed only when the second eye is opened or the mask is turned so that the snout becomes occluded. The contradictory cues from the reversed velocity gradients of points at different depths make the face appear to "swim" unsteadily and to follow you as you move your head so that you are aware that something is "not right" about the object, but these factors aren't sufficient to overcome the bias toward perceiving the known object in its familiar form.

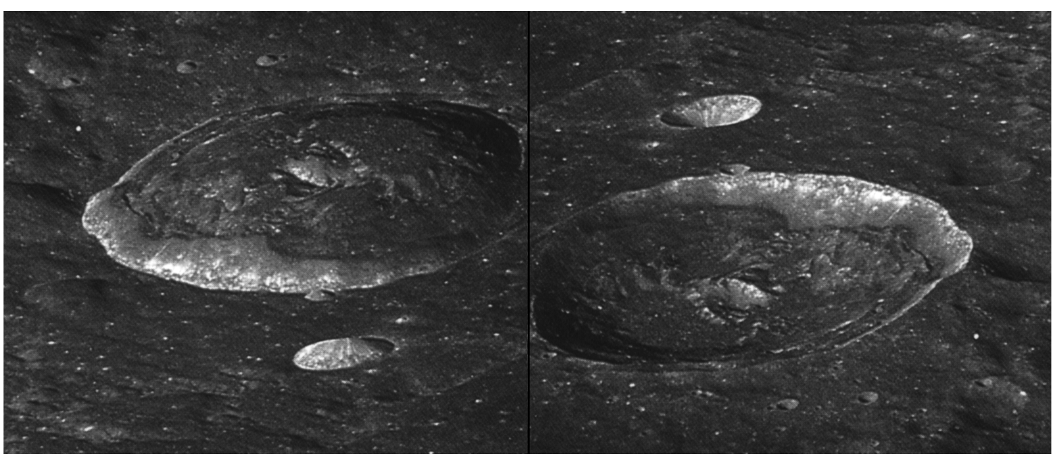


**Figure 3.13: A photograph illustrating the so-called "crater illusion", in which concavities are perceived as convexities when the image is rotated by 180°. This particular picture, excerpted from a NASA photograph printed in [Krisciunas and Yenne 1989], shows Hershel Crater (on the moon).**

Even when the shape category (such as convex or cylindrical) is known, psychophysical studies [Todd and Mingolla 1983, Mingolla and Todd 1986, Bülthoff and Mallot 1988] have consistently shown that observers have great difficulty deriving accurate absolute estimates of global surface curvature or local surface orientation from shading information. Todd and Mingolla [1983] found that observers underestimated the absolute curvature of cylindrical surfaces by 53% when Lambertian shading provided the only cue to shape. Bülthoff and Mallot [1988] found that observers were able to perceive very little depth (measured using a stereo probe) in images of ellipsoidal surfaces, oriented end-on, when shading was the only available shape cue. They also found that subjects were essentially unable to discriminate between ellipsoids of varying eccentricities on the basis of the shading information alone. Mingolla and Todd [1986] reported a consistent tendency for observers to underestimate the amount by which ellipsoidal surfaces of varying eccentricities were slanted out of the picture plane, with the result that objects appeared to be flatter than they really were.

Although we do not appear to be able to estimate local surface shape or absolute surface curvature very well solely on the basis of shading information, there is some evidence that people are fairly good at making *relative* judgments about the slants and curvatures of adjacent surface patches. Johnston and Passmore [1994] conducted experiments in which they varied first the curvature and then the slant of a circular patch on the surface of a spherical object. They found that subjects were able to indicate with considerable accuracy whether the patch was more curved or less curved than the surface upon which it sat in the first case, and whether it had been rotated to the right or to the left in the second. Johnston and Passmore also found that the pattern of results seemed to suggest that observers were perceiving curvature directly and did not estimate it as a local difference between the surface normals at two neighboring points.

De Vries *et al.* [1994] provide evidence that observers' judgments of qualitative surface shape, defined in terms of the *shape index* proposed by Koenderink (and illustrated in figure 3.14 [Koenderink 1990]) may be generally independent of curvature judgments. Under conditions of stereo viewing, the accuracy with which subjects in their experiments were able to determine the surface shape category of a stimulus was not affected by the amount of curvature of the surface or by the orientation of the surface relative to the line of sight.

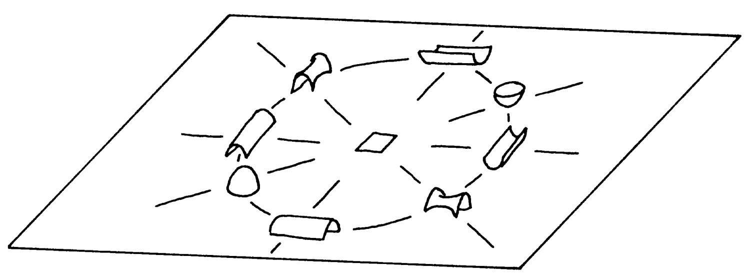


**Figure 3.14: A diagram, from [Koenderink 1990], representing the qualitative surface shape categorization defined by his shape index. This measure, which depends on the values of the principal curvatures, is a function over the interval [-1, 1] in which concave and convex umbilics map to the endpoints, a symmetric saddle maps to the midpoint, and concave and convex cylinders map to the points $-\frac{1}{2}$, $+\frac{1}{2}$.**

The best explanation for how we are able to obtain a vivid and stable perception of three-dimensional shape from patterns of shading in spite of the fact that we can't make accurate judgments of local surface orientation or even disambiguate the direction of the surface normal at arbitrary interior points may be that our understanding of surface shape should perhaps be thought of as an organization of space, a global rather than a local phenomenon, based on a qualitative rather than a quantitative estimate of depth along the line of sight.

Ramachandran [1988] provides numerous examples of the global rather than local nature of our perception of shape from shading. One of the most compelling is a stimulus in which a row of circles, shaded to simulate either convex bumps illuminated from the right or concave depressions illuminated from the left, is placed immediately above a second row of circles shaded in the reverse orientation. Although the perception of depth is inherently ambiguous in this display, one row of circles will always appear convex while the other appears concave; if, by force of will, one circle is made to reverse in depth, all of the other circles will simultaneously reverse. The display always appears to be illuminated by a single light, sometimes coming from an opposite direction than earlier.

**Figure 3.15: Photographs of bronze sculptures in which surface shape is shown through contours and shading, [left: Réunion des musées nationaux, right: Frayling *et al.* 1992]. Left: a sculpture by Houdon [1790] in which both global and local form, by extension, can be fairly readily understood from the patterns of (mostly specular) shading. Right: a sculpture by Boccioni [1913] in which, although the impression of relief is very vivid, the global shape of the figure and the relative shapes of the local surface patches are much harder to understand. (It is difficult, for example, to determine whether the area enclosed in the black rectangle is convex or concave, despite the boundary information and the possibility of comparison with other surface regions.)**

Koenderink and van Doorn [1980] introduce a theoretical foundation for interpreting the structure of the illuminance distribution in terms of characteristics that are invariant with respect to the direction of illumination. Koenderink and van Doorn [1993] elaborate on this approach, describing how on a generic Lambertian surface the illuminance distribution can be characterized by two types of critical points (where the gradient of the illuminance is zero): illuminance maxima corresponding to generic surface points where the surface normal is aligned with the direction of illumination, and illuminance maxima, minima or inflections corresponding to parabolic surface points where the direction of illumination is aligned within the plane of the cylinder axis (the direction along which the principal curvature vanishes). They describe how the second category of points can be distinguished from the first by the fact that they cling to the parabolic lines and trace a closed loop over the surface as the direction of illumination varies. While differential geometers often describe surfaces in terms of the Gaussian curvature, it remains to be seen what role perceiving the locations of parabolic lines has in human shape understanding. I will return to briefly discuss this issue in somewhat more detail in chapter four.

Koenderink and van Doorn [1995] note that observers tend to make fairly consistent judgments about the direction of tilt (the directional component of surface orientation that lies in the image plane) at selected points on images of shaded surfaces but that judgments of slant (the angle of orientation of the surface away from the line of sight) tend to exhibit much more variance, both between observers and in repeated measurements by a single observer, with the result that final judgments of surface orientation appear to be (inconsistently) scaled by a highly variable depth factor. Koenderink and van Doorn also found that while observers were able relatively easily to judge the depth order of nearby points on a depicted surface, they could not reliably judge the depth order of points in disparate parts of the image. They conclude that our understanding of surface relief is based on depth order relationships, a finding that was also proposed by [Todd and Reichel 1989].

De Haan *et al.* [1995] describe an experiment in which they systematically investigated the accuracy and stability with which a coherent surface definition could be obtained from a series of local judgments of surface slant and tilt in monocularly viewed images, where shading was the only readily available cue to shape and no information about the direction of illumination was given. They found that the consistency of the surface orientation judgments was not significantly affected either by the overall amount of surface slant or the amount of illumination contrast, which varied as a function of the inclination of the incident light (its angle with the perpendicular to the image plane) but was strongly affected by the azimuth of the light source (roughly corresponding to a point within the image plane). Most importantly, although the shading patterns they used did not unambiguously define a unique surface shape under these experimental conditions, de Haan *et al.* found, through an analysis of the curl of the perceived depth gradient, that the reported surface normals, while often varying significantly from the veridical direction, were nonetheless consistent with the perception of some continuous underlying surface. They report that a meaningful and stable surface shape was arrived at, despite all of the various inaccuracies in local orientation judgments. De Haan *et al.* interpret this fact to be an indication that the human visual system is able to resolve the ambiguities inherent in determining shape from shading in a consistent manner, most likely by relying on a set of assumptions about such things as the direction of illumination (overhead or from the right), the global direction of slant of the surface (with depth increasing in the upward direction), or the orientation of the surface with respect to the line of sight (perceiving it as more closely aligned with the image plane, having less total slant).

Effectively representing the shapes of transparent surfaces (without resorting to methods that rely on distortion of the background due to refraction of the light rays) is complicated by the fact that these surfaces in practice do not exhibit Lambertian shading properties. The only source of shape from shading information that we have in this case comes from the specular highlights. The paucity of the shape information available in a photograph of a transparent surface (which will be demonstrated in chapter four) provides an additional empirical example of the importance of the role of illuminance gradients in communicating shape and depth.

### 3.4: Binocular disparity (stereo)

When we look at a particular point in space with both of our eyes (which is the normal case in vision), the views perceived by each individually eye will be slightly different, because of their spatial separation in the head. The focal point will fall on corresponding locations in the retinal images of each eye, as will all other points that are equidistant from the viewpoint (defining a plane, called the horopter, of points in space that have zero retinal disparity in that view), but points that are closer or farther from the viewpoint than the fixation point will be mapped onto disparate locations in the two retinal images. Objects that are closer to the viewpoint (in front of the horopter) are seen in *crossed disparity*, while objects that are farther away are seen in *uncrossed disparity*. Our visual system is able to interpret the relative depths between two points in space from the amount of retinal disparity in their projections to each eye and to determine the depth order of these points from the sign (crossed or uncrossed) of this disparity. Exactly how the two flat views are united in the brain to yield a unified perception of three-dimensional space has been a topic of considerable investigation [Poggio and Poggio 1984, Julesz 1986], which I don't have the space here to describe. While about 5% of the population cannot perceive depth from binocular disparity [Coutant and Westheimer 1993], for the majority of the population it is one of the most phenomenologically important and perceptually acute sources of depth information.

Wheatstone [1838], although perhaps not the first to recognize that each eye sees a slightly different view of the world, appears to have been the first to recognize its significance for the perception of depth. He was also the first to demonstrate, with the invention of the stereoscope, that viewers could achieve a vivid three-dimensional perception of an object that was depicted by a pair of appropriately drafted flat drawings. (Of historical note, Wade [1983] discusses, at great length, the controversy between Brewster and Wheatstone over the right to claim precedence of the invention of the stereoscope and the recognition of the phenomenon of stereoscopic vision and explains the origins of some erroneous reports, e.g., by Kaufman [1974], that cast doubt upon the originality of Wheastone's ideas.)

Many of the earliest theories of binocular vision operated on the assumption that stereopsis was a higher-level visual process that depended on the pairwise matching of specific features recognized separately in each of the retinal images. Julesz, however, introduced the concept of the random dot stereogram, illustrated in figure 3.16 (and whose phenomenal characteristics are described in great detail in [Julesz 1964]), to demonstrate that an integrated

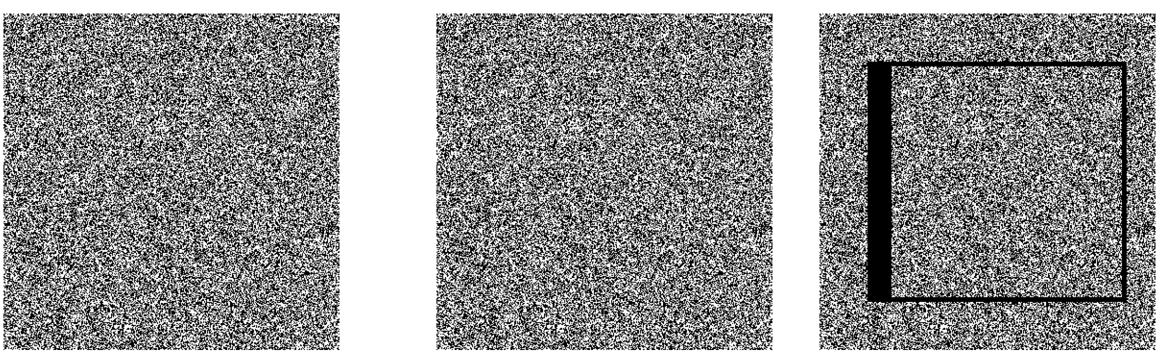


**Figure 3.16: A random dot stereogram, after [Julesz 1964], represented in uncrossed disparity. When the image on the left is fused with the image in the middle, a central square will appear to come forward in depth. The rightmost square illustrates the process used to generate this stereogram — the outlined part of the left eye image is displaced to the right in the view to the right eye and the "uncovered" pixels randomly filled in. Perceived depth will be proportional to the magnitude of the displacement.**

perception of three-dimensional form can be achieved even when no feature information is immediately accessible in either of the images individually. This should not be taken as an indication that edges are not useful for stereopsis but only as an illustration of the fact that stereo perception does not depend on a prior understanding of monocular form. Saye and Frisby [1975] have shown, however, that the perception of images in random dot stereograms is facilitated when monocularly conspicuous features are present, possibly because they provide more immediate feedback to guide vergence adjustments.

While it might seem that *some* discontinuity information, even if fragmented into pixel-sized regions, would be needed to recover disparity, Bülthoff and Mallot [1988] have shown that depth, albeit of diminished extent, can be perceived in stereoscopic displays of smoothly curved surfaces over which the intensity gradually varies but in which no edge information, in terms of zero-crossings of the Lapacian, is anywhere available. (This is of interest to us because we will be using stereo in chapter six to display smoothly curving surfaces and want to be sure that there is a basis for assuming that local depth information will be effectively conveyed.) Bülthoff and Mallot verified the additional contribution of the stereo information in these experiments by comparing the local surface orientation judgments obtained using a probe in the disparate case with the local surface orientation judgments obtained when identical views of the object were displayed to each eye. Unfortunately an inherent feature of this paradigm is that, because binocular vision of identical images is used for the nondisparate case rather than monocular vision of a single image, an additional, contradictory cue to flatness is introduced by this process. It is likely that greater depth would be perceived in the monocular case without this conflicting cue, and it would be interesting to compare the results of monocularly viewed stimuli to the results obtained with the binocularly viewed disparate and nondisparate images. Rogers and Collett [1989] did exactly this for images in which depth was indicated by motion parallax and found that when zero retinal disparity was binocularly indicated, perceived depth decreased dramatically from the levels judged under conditions of monocular viewing.

Although I will talk more in a later section about the relationships between the various depth cues, the fact that there is a relationship between convergence and binocular disparity needs to be stressed here. When generating stereo images, and in particular when measuring the perceived depth of surfaces depicted in them, one needs to take precautions to avoid unwittingly placing any of the kinesthetic cues of stereopsis, convergence or accommodation in conflict with each other. Buckley and Frisby [1993] found, for example, in an experiment investigating the interaction between stereo, texture and outline cues to depth, that qualitatively different results were achieved when actual physical objects were substituted for the computer-generated stimuli, and they attributed at least some of the cause of these differences to the inherent conflict, in the computer-simulated images, between the depths indicated by binocular disparity and the long-discounted cue of accommodation. Accommodation may be a meaningful cue to the flatness of the screen in many situations in which the viewing distance is relatively close. Also of interest, Wallach and Zuckerman [1963] have shown that the amount of depth perceived from a fixed amount of binocular disparity in a displayed pair of images will vary as the viewing distance (and the corresponding vergence angle) is changed. Johnston [1991] investigated how the veridicality of perceived depth in stereo images might be influenced by the viewing distance in experiments in which appropriate vergence parameters are maintained for all stimuli. She found that estimated depth, derived from judgments of the apparent circularity of cylindrical surfaces displayed in random dot stereograms, was most veridical at a viewing distance of about 1m, systematically underestimated when the viewing distance was greater and overestimated when it was less. Johnston concludes that incorrect egocentric estimates of viewing distance cause the binocular disparities to be incorrectly scaled and that perceived shape was distorted as a result.

## 3.5: Depth from motion

As we move about, our view of the world is constantly changing. The eye, in fact, only "sees" well when the view is continuously updated or refreshed, although we normally don't notice the tiny tremors or microsaccades that prevent images from being stabilized on our retina. While eye movements alone only change the visual field by a rigid rotation [Koenderink 1986] and do not provide any information about depth, the changes in the visual field (or "optic flow") resulting from translations of the viewpoint supply the visual system with a "cue" to depth that is nearly as powerful as stereo vision.

### 3.5.1: Motion parallax

The idea that changes in retinal stimulation due to observer or object movement are capable of inducing a vivid impression of depth in three dimensions was formally put forward nearly a century ago [von Helmholtz 1925] and was undoubtedly appreciated by many people well before then. The significance of this observation and in particular the notion that motion parallax information alone was sufficient for conveying a robust impression depth, independent of all other cues, was stressed by Rogers and Graham [1979] who demonstrated, using a paradigm similar to the Julesz random dot stereogram, that observers could perceive form from motion in situations where no stationary cues to depth were available. In these experiments, Rogers and Graham mounted the observer and the oscilloscope on separate movable platforms and continuously updated the image in correspondence with the changes in the locations of either the represented object or the viewpoint so that the appearance of the dot patterns was always consistent with a stationary three-dimensional surface attached to the face of the display device. Although in the absence of movement the dots appeared to represent nothing more than a two-dimensional pattern on the screen, when the motion cue was added observers consistently reported a strong impression of viewing a three-dimensional surface and were able to describe the qualitative surface shape without error. Rogers and Graham measured the accuracy with which subjects were able to judge surface depth in these displays, under conditions of active and passive motion, by asking subjects to match the depth, indicated by disparity, on a second device, of a stereo probe to the perceived depth of the surface represented on the oscilloscope. In addition to verifying that the amount of perceived depth increased with the extent of the motion (as was expected), they found that observers consistently perceived more depth in the situations where they moved their head than in the situations where the experimenter moved the display device.

Motion parallax, or the sensation of depth that our visual system derives from the changing information in the optic array, appears to function in a very similar manner to stereopsis; the two sources of depth information are, in a sense, geometrically comparable, and a brief description of the phenomenon helps make this more clear. When we fixate on a point in space and then move our head (for example by a translation to the right), objects that are farther away from us than the fixation point will move across the visual field in the same direction as our head motion while objects that are closer to us than the fixation point will move in the opposing direction, and the speed of an object's motion will be proportional to its distance in depth from the fixation point [Braunstein 1962]. In terms of the retinal image, while stereo depth perception has its foundation in the correspondence of depth distances to binocular disparities, depth perception from motion parallax comes about as a result of the displacements over time of the locations of projected points on the retina of a single eye. The similarities and differences between these two methods for specifying depth have been examined by a number of researchers. Rogers and Graham [1982] demonstrated that the spatial frequency thresholds for perceiving depth in random dot patterns are remarkably similar for depths specified by motion parallax and depths specified by binocular disparity. They remarked that the similarity of the sensitivities of these two methods seemed to support the hypothesis that there was a close relationship between them. Nawrot and Blake [1991] cite and supplement (through cross-adaptation experiments) evidence that our visual system may process stereo and motion parallax information, if not in a combined fashion, at least through the same neural mechanisms. Richards [1975] also cites studies that indicate some dependencies between these two cues but says that the processing of depth from motion parallax information appears to follow and possibly depend on the indication of depth by stereopsis. Braunstein *et al.* [1986], however, found that six subjects who could perceive no stereoscopic depth in static displays were able to make correct judgments of depth order in 90% of the experimental trials after motion cues to depth were added. More recently, Tittle *et al.* [1995] have shown that observer's judgments of three-dimensional shape due to stereo and motion parallax cues appear to be systematically distorted, and in different ways. While the direction and extent to which observers in their experiments misjudged shape from stereo was dependent on the viewing distance, the veridicality of observers' judgments of shape from motion decreased with increasing slant of the object away from the frontoparallel plane, indicating that there are some limits to the similarities between these two sources of depth information. Although some researchers have described stereo and motion parallax as providing "redundant" information [Sekuler and Blake 1994], I believe it would be a mistake to interpret this as meaning that both cannot be more useful together than either one is separately.

### 3.5.2: Kinetic depth effect

Much of the research in depth perception from motion has been based on experiments in which the observer is stationary and the object rotates in depth (the scaling and perspective effects of translations in depth have been discussed earlier in this section, in the context of pictorial depth cues). In this situation the cue is usually referred to as the "kinetic depth effect". Wallach and O'Connell [1953] performed a series of experiments to attempt to define the conditions under which observers would perceive a three-dimensional object rigidly rotating in depth when the stimulus consisted only of a two-dimensional silhouette (parallel shadow projection) changing in shape on a flat plane, and found that this perception seemed to depend on the property that both the lengths and orientations of the two-dimensional shadow projections changed. Further investigations, described in [Braunstein 1976], revealed that if all cues to flatness were eliminated from the experimental setup, there appeared to be no particular restrictions on what kinds of motion in the plane could be interpreted as motion in depth. When cues to flatness were present, the characteristics of motion most conducive to a perception of depth were simultaneous changes in both length and orientation, as indicated by Wallach and O'Connell, and indications of perspective (converging horizontal disparities, compressed vertical disparities). Braunstein also reports that the propensity of subjects to perceive depth in computer-generated kinetic displays, where the surfaces are represented by a collection of discrete elements (such as lines or dots), does not seem to be strongly related to whether or not the observers perceive a rigidly rotating object. Ramachandran *et al.* [1988] found that visual system will readily abandon a "rigidity" assumption in favor of an interpretation in which the depth at any point is indicated by the velocities of the discrete elements. Most of the experiments investigating kinetic depth effects are based on stimuli that represent objects rotating about a vertical axis in the display screen. Green [1961] specifically examined the effect of changing the axis of rotation and found that, for surfaces indicated by dots, while the perception of a coherent object was strongest when rotation was depicted about this vertical axis, and worst for rotation about an oblique axis or tumbling about an axis whose orientation continuously changed, there was not much difference in the amount of depth perceived under these different conditions. While Wallach and O'Connell were able to demonstrate a perception of depth-from-motion based on a parallel projection from 3D to 2D, Gibson and Gibson [1957], in similar investigations, stressed the importance of using a perspective transformation. Braunstein [1966] systematically examined the effects of using a parallel rather than a perspective projection to represent the surface that was rotating in depth. He found that when a parallel projection was used, observers were unable to disambiguate the direction of rotation and unable to perceive depth from rigid translations of the target across the field of view, although they were easily able to achieve both of these impressions when a perspective projection was used. He also found that subjects reported a stronger impression of depth when stimuli were generated using a perspective rather than a parallel projection and that the strength of this depth impression increased even further when the perspective was "forced".

### 3.5.3: Optic flow

Gibson [1950a] proposed a theory of visual perception from motion based on "motion perspective" (or "optic flow"), a continuous gradient of the deformation of the optic array due to movement by the observer. Koenderink and van Doorn [1976] described how the motion parallax field (caused by movement of observer relative to a surface) can be understood in terms of four elementary fields representing different types of affine transformations: curl (rigid rotation component), divergence (isotropic expansion and compression component), and deformation (pure shear component), along with simple translation. Koenderink [1986] later expanded on this analysis and discussed how the visual system might use optic flow information in the recognition of three-dimensional shape.

### 3.5.4: Active versus passive viewing

A number of researchers [Rogers and Graham 1979, Lipscomb 1981, van Damme 1994, Durgin *et al.* 1995] have offered evidence that the perception of depth from motion is facilitated more by active movements of the observer than by passively viewed movements of the scene. How much better is the spatial understanding that we get from active, as opposed to passive, movements, or from head movements as opposed to object movements, when the object is under interactive control? As far as I can tell, many of the experiments defined passive motion as motion not controlled by the observer. In these cases a number of other potentially complicating factors can arise such as uncertainty about the direction or extent of the object rotation, etc. that should really be dealt with separately if one is interested primarily in determining the advantages of vestibular information. There are other tradeoffs between head motion and interactive object manipulation to consider as well that will depend on the scale of the object and its distance from the observer. If one is regarding a sculpture in a museum, walking around it to get a better view makes perfect sense; if one is choosing a figurine from a craft shop, picking it up and turning it around is probably the most effective way to inspect it. The question is, by how much, in what ways, and under what conditions can we expect to enhance the veridical perception of three-dimensional space by allowing unrestricted observer movement within a fully represented virtual environment? With the recent technological advances in the development of low latency head-tracking devices and high-speed, high quality graphics engines, it is becoming both more important and more feasible to perform these kinds of studies.

### 3.5.5: Stereokinetic depth

The stereokinetic depth effect, usually regarded as a "depth illusion", was first recognized by Musatti (1924), who demonstrated that certain two-dimensional patterns, when rotated in the image plane, are capable of evoking a compelling impression of depth. Figure 3.17 provides an example of the type of pattern that is typically used to demonstrate stereokinetic depth. Under optimal conditions, the rotational motion of the circles isn't seen, only their shifting relative positions. Proffitt *et al.* [1992] show that it is this translational motion that stimulates the perception of depth, and demonstrate that a stereokinetic depth impression can be evoked from stimuli in which explicit evidence of an absence of rotation is provided. They assert that the kinetic depth effect can be decomposed into two transformations: one indicating the depth relationships between points on a surface (depth order) and the other indicating the orientations of the surfaces relative to the observer (depth magnitude), and characterize the stereokinetic depth effect as conveying information of the first kind but not the second. Despite the inherent ambiguity of both depth magnitude and direction in a stereokinetic display, Proffitt *et al.* reported highly consistent inter-observer judgments of the amount of apparent depth represented.

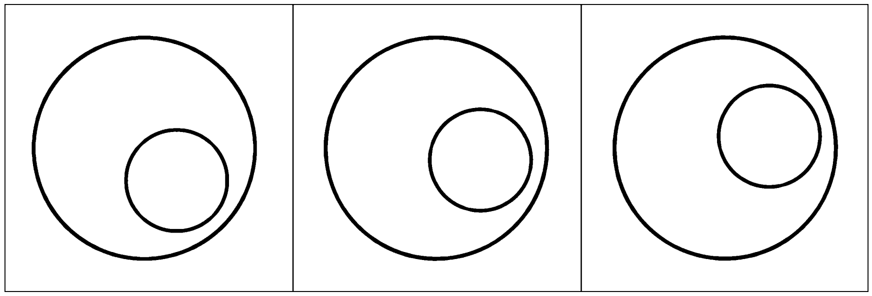


**Figure 3.17: An sequence of frames illustrating the type of stimulus than can evoke a stereokinetic impression of depth.**

### 3.5.6: Flicker

Although it is difficult to imagine how flicker could be of practical use as a device for enhancing appreciation of the relative depths of objects in a three-dimensional scene, it bears at least mentioning that Wong and Weisstein [1984] reported finding that when selected regions of a random dot field were flickered, the nonflickering regions appeared to segregate from them and come forward in depth and that the optimal temporal frequency for inducing this separation was 6Hz [Wong and Weisstein 1985]. I am unable to think of any experiential manifestation of this particular "depth cue". Although one can easily imagine that during normal vision and locomotion it would be relatively more likely for various areas of the "background" to be alternately visible and then occluded than areas of the foreground, the vastly greater flicker frequency at which this depth segregation was experimentally observed seems to preclude any connection between these two phenomena.

## 3.6: The integration of depth cues

The whole issue of whether depth cues should be regarded as "learned" or "inherent", whether visual perception in general would be better described as a prediction based on past experience [Ittelson and Kilpatrick 1952] or as a direct, unmediated response to retinal stimulation [Gibson 1950b], seems to have been a topic of some debate in years past, and there are important implications in each of these models. For every example of an illusive perception that is not destroyed when its illusory nature is understood, there is another example of a perception that can be alternated "at will". Nakayama and Shimojo [1992] have recently proposed a theory of visual perception ("inverse ecological optics") that integrates elements of both schools of thought, emphasizing the importance of the role of learning, but in a bottom-up sense.

Nakayama and Shimojo [1992] demonstrate that our interpretation of the three-dimensional structure of the scene represented in any single two-dimensional image (or binocular pair of images) appears to operate from a built-in assumption that the viewpoint is generic. We should not take this assumption of a generic viewpoint to mean, however, that the shapes of objects are necessarily more accurately perceived when they are viewed from an oblique angle. Perrett and Harries [1988] found that when viewing unfamiliar, irregular objects with the goal of remembering their shape, observers tend to show a striking amount of similarity in their preferred "characteristic views" and that these views seem to be defined on the basis of gross contour dimensions rather than specific contour singularities. Mingolla and Todd [1986] found that observers seemed to have a distinct bias toward perceiving objects in images displayed on a computer monitor to be more axially aligned with the display screen than they actually were and that they seemed to make errors in their judgments of surface slant and tilt because of this predisposition to assume a "characteristic view".

While it can be argued, for our purposes, that it is not as necessary to know how or why the various depth cues function as it is to understand their relative potential for communicating shape and depth in computer-generated images, the importance of having a basic understanding of both the underlying principles and psychophysical evidence for each of the depth cues cannot be discounted. In nearly all respects, coming up with a set of principles that define "optimal" modeling or rendering techniques for representing shape and depth so that they can be most accurately perceived turns out to be more complex than it seems at first glance, a conclusion that was reached by Wanger *et al.* [1992] in a similar investigation.

The question of how the visual system integrates the depth information provided by all of the various available sources to arrive at a single, stable perception of three-dimensionality is far from having been satisfactorily answered. Most of what is currently believed about the relative importance of the various cues to depth has been teased out of the results of experiments based on cue-conflict paradigms, in which stimuli are artificially manipulated so that the depth indicated by one cue contradicts the depth indicated by another, and the assumption is made that the dominating perception under these conditions will be determined by the relative "strengths" of the opposing cues. In the natural world, however, there are typically many different *equally valid* sources of depth information available, and recent research [Young *et al.* 1993] suggests that it might be wrong to assume that the strategies used to resolve strongly conflicting depth cues would be necessarily indicative of the methods by which complementary depth information is combined. Bruno and Cutting [1988] conducted experiments in which depth information was either provided or not indicated by each of four different pictorial depth cues: relative familiar size, height in the image plane, occlusion, and motion parallax. Although the depth information specified by the cues was in many cases "redundant", subjects always perceived more depth in the displays in which more of the cues were present. Bruno and Cutting concluded that the information from the separate sources was being combined in an additive fashion, and as far as I can tell, most subsequent research seems to confirm this mode of behavior under non-conflicting conditions. Johnston *et al.* [1993] review several different possible mechanisms by which the visual system might derive a single estimate of depth from the information provided by all of the available sources. The first of these is called *veto* and corresponds to a situation in which strongly conflicting information is by the different cues. When vetoing occurs, the "stronger" cue simply overrides the weak cue, and the perception of depth from both together is equivalent to the perception of depth that would have come from the stronger cue alone. In the experiments by Bülthoff and Mallot [1988], for example, the depth information provided by the zero retinal disparity cue seemed to veto the depth information (possibly) provided by shading, resulting in the perception of a flat surface under those conditions. The second and third possible mechanisms for cue combination, whose definition they attribute to Clark and Yuille [1990], are termed *weak fusion* and *strong fusion*. In the weak fusion model, depth information is processed separately by each "module" and the independently arrived at estimates are then combined in a weighted linear fashion, while in the strong fusion model the depth cues interact in non-linear ways with, for example, one cue providing disambiguating information that enables depth information to then be derived from another. Young *et al.* [1993] have more recently proposed, and offered evidence consistent with, a "modified weak fusion" model in which the independent depth estimates arrived at on the basis of each of the separate sources of depth information are weighted, before being additively combined by an amount that reflects their "apparent reliability" in the particular situation.

An important concern, for our purposes, is knowing the extent to which we might communicate depth information more effectively by incorporating as many different depth cues as possible into the generation and display of our data. Does the representation of additional depth cues always help, or is the information provided by the various cues essentially redundant, so that we can hope to gain little extra by representing them all? Some of the depth cues described earlier in this section (such as relative familiar size, occlusion and height in the image plane) are actually direct reflections of the geometry of the scene and in that sense are difficult to "get wrong" (unless, of course, we use a parallel rather than a perspective projection). Others involve modeling aspects (such as introducing straight lines to emphasize linear perspective cues to depth, using overhead, oblique illumination to most effectively convey surface relief, modeling aerial perspective by introducing "depth cueing" in which the luminance at any point is modified according to depth, and so on). Probably the most directly relevant study for this dissertation on the issue of cue combination was done by Johnston *et al.* [1993], who examined whether the perceived elongation in depth of a stereoscopically viewed surface could be improved if, instead of being covered by a dot texture whose constant gradients were indicative of a fronto-parallel plane, it was covered with a veridical texture that more accurately reflected the surface geometry. Although the strength of the stereo cue was sufficient to provide a good indication of surface depth when the texture cue was "absent", Johnston *et al.* found that perceived depth increased in all cases when the more geometrically appropriate texture was used (even when the depth indicated by stereo already represented an overestimation of the actual depth). The implications for this dissertation are that even in the presence of other cues to depth it is reasonable to hope that by adding texture to a transparent surface we can further improve the ability of observers to accurately judge its shape and depth. It is probably fair to say, based on the above indications about how our visual system seems to process depth information, that we cannot "go wrong" by representing shape and depth in as many ways as reasonably possible. At some point, the accuracy of the perceived depth will be such that there will be little room for improvement, but it does not seem that we are in any present danger of reaching that point.

While the issues of the perception of shape and depth are fascinating, I now move to the exploration of techniques for more effectively communicating the three-dimensional shape and relative depth of a layered transparent surface. Further information about the various depth cues can be found in the references cited above, or in the following publications not directly cited: Bruce and Green [1990], Parker *et al.* [1992], Bülthoff [1991].

## 3.7: Shape and depth from texture

As a large, planar surface recedes in depth, the projected size (under perspective projection) of any unit area of the surface will decrease isotropically with distance and anisotropically with the inclination of the plane away from perpendicular to the viewing direction. If the surface is covered with markings (optical texture), it follows that the projected shapes and sizes of these surface markings will vary as a function of their distance from the viewpoint and the orientation of the plane.

Gibson [1950a, 1950b] was the first to stress the importance of the role of these *texture gradients* in human visual perception. He regarded the "terrain extending to the horizon" as the "most fundamental surface" and consequently focused his attention on the gradients of texture element size and spacing that seemed to characterize our perception of this ground plane. Gibson hypothesized that distance was understood from the density of surface texture (using a definition of density that incorporated both element size and inter-element spacing) and that the amount of surface slant was understood from the rate of change of texture density, with the direction of the surface slant specified by the direction of this density gradient.

### 3.7.1: Texture properties characterizing a large planar surface receding in depth

Gibson [1950a] conducted the first observer experiments aimed at exploring the role of texture gradients in our perception of surface slant. Gibson's experiments used slides depicting wallpaper patterns photographed from different vertical camera angles (ranging from 10˚–45˚), which were back projected onto a large translucent screen and viewed monocularly through a pair of holes arranged in depth so that such that the circumference of the closer hole (defining the perimeter of the view) would be out of focus and the solid angle subtended by the view would match the camera parameters. Subjects were asked to adjust the slant of a tactually presented board to match the apparent slant of the depicted surface. Two different wallpaper patterns, shown in figure 3.18 [Gibson 1950a], were used in these experiments, and Gibson notes that the particular textures were chosen, with some care, to represent as nearly as possible the subjective epitome of "regularity" and "irregularity". Gibson found that the judged slant increased, in a fairly consistent manner, with increases in the actual slant, but that observers consistently underestimated the amount of the actual slant, particularly in the case of the irregular texture pattern.

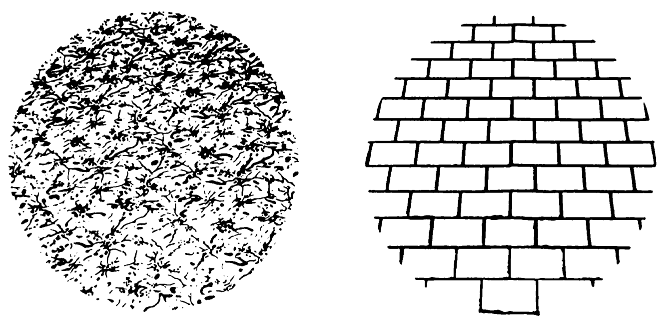


**Figure 3.18: Stimuli used by Gibson [1950a] in experiments investigating the effects of texture gradients on the perception of surface slant.**

Gibson was able to show convincingly through these experiments that observers could use the information contained in patterns of projected surface texture to interpret the slant of a surface. The particular characteristics of the patterns of optical texture that were primarily responsible for this effect were not clearly identified, however, and much of the subsequent research in shape and depth perception from patterns of surface texture conducted in the years since these first experiments has been directed toward exploring this question.

Throughout this section, I will be describing, as briefly as possible, what I feel to be the most important aspects of the various experimental designs in addition to reporting the authors' conclusions because I believe that having an understanding of certain details of the experimental method is crucial for knowing how to interpret the experimental results. People have found, for example, that in studies of slant perception where the edges of the ground plane were made explicit, the effect on slant perception of these edges completely overwhelms any effects due to the particular characteristics of the texture on the interior of the slant plane [cf. Braunstein 1976]. Also, although most studies found a significant underestimation of actual surface slant, the amount of this regression towards the frontal plane was discovered to be dependent on, in addition to surface texture, various experimental factors such as the field of view, and the size and distance of the surface [Braunstein 1976].

Noting that Gibson had reported obtaining better measurements of judged slant for stimuli depicting a "regular" as opposed to "irregular" texture pattern, Flock and Moscatelli [1964] conducted a carefully controlled experiment to help determine which particular aspects of texture "regularity" might have been most responsible for this increased accuracy. They took six 6-ft x 3-ft sheets of white cardboard and covered each of them with 288 small black shapes, either square or "irregular" in shape, of either similar or differing sizes, and spaced either at the center of each 3" x 3" region or at a random location within this grid cell, as shown in figure 3.19. The planes were each variously arranged at 9 different slants ranging from -40° to +40°. Subjects viewed the stimuli monocularly through an aperture that prevented them from seeing the edges of the planes, and were asked to adjust the slant of a (monocularly viewed) palmboard to match the perceived slant of the textured plane. Flock and Moscatelli found that judged slant came closest to veridical when all three of the texture parameters were regular, but that the slant was still underestimated in this case (they computed a mean regression coefficient of .78 for this stimulus; 1.0 represents maximal accuracy). Slant judgments were least accurate for the stimuli in which all three of the parameters were irregular (in which case the mean regression coefficient was 0.53) but still significantly better than in the control case where the plane was untextured (the mean regression coefficient in this case was 0.05). They found intermediate levels of accuracy in judged slant for the intermediate stimuli, in which some of the parameters were regular and others irregular, although only the differences between the two extremes were at the level of statistical significance.

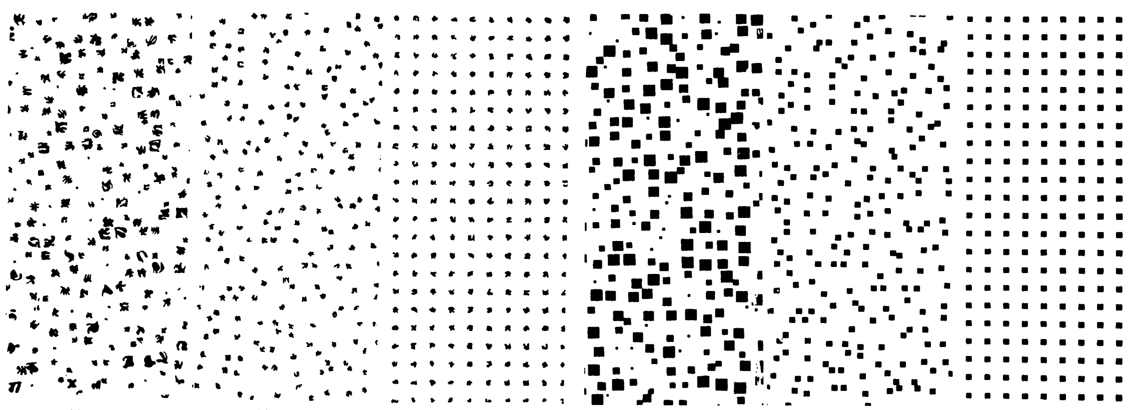



**Figure 3.19: Stimuli used by Flock and Moscatelli [1964] to measure how various types of pattern regularity in surface texture influence the accuracy of slant judgments.**

Attneave and Olson [1966] were among the first to demonstrate the dominant role of linear convergence cues in determining the effectiveness with which a texture pattern evoked a perception of a receding surface. They had observers make a forced-choice judgment, based on subjective preference, about the direction (left/right) of apparent recession of the plane indicated by various line patterns drawn in ink on 9-in x 12-in cards, viewed (presumably binocularly and without actual slant) from a distance of 5 ft. Stimuli in which the transverse set of grid lines radially converged, regardless of the latitudinal spacings of the verticals, were consistently and effortlessly judged to represent a plane receding in the direction indicated by the linear convergence, as was the stimulus in which randomly distributed line segments of equal length were similarly oriented in a radial pattern, as if converging toward a vanishing point off to the left or right of the image. The other stimuli, in which line segments of gradually decreasing length were uniformly oriented in either a horizontal or vertical direction, were judged to be much less effective, despite the clear gradient of line segment length.

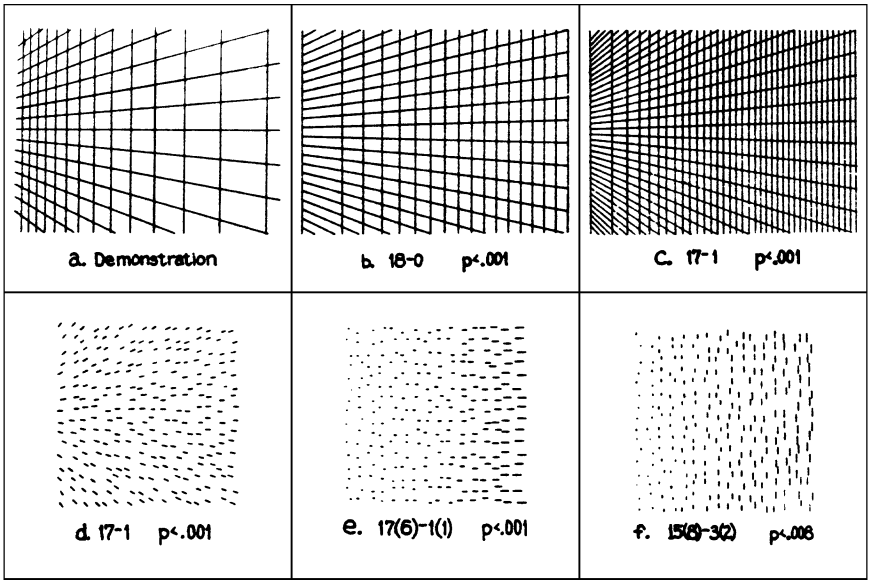



**Figure 3.20: Stimuli used by Attneave and Olson [1966] to demonstrate the effectiveness of linear convergence as a cue to the direction of recession of a planar surface. The numbers of leftward and rightward slanting judgments recorded for each of these stimuli are specified below the corresponding patterns, with the forced choices (where subjects balked at giving a response because they thought the pattern looked essentially flat) indicated in parentheses**

Braunstein and Payne [1969] confirmed the dominant role of perspective convergence over vertical compression, sometimes referred to as "foreshortening" [Gillam 1970], in influencing judgments of the relative slant of planar surfaces rotated about a horizontal axis. Subjects viewed pairs of projected images, representing surfaces slanted at five different angles from 0˚ – 65˚, from a distance of 25.4cm through apertures that both restricted the field of view of each eye and allowed each eye to see only one of the two images of the pair. Sample stimuli are illustrated in figure 3.21. Subjects were asked to make a 2-alternative, forced-choice decision about which of the two surfaces appeared to be more slanted in depth. Braunstein and Payne also found no significant differences in the responses when dot instead of line patterns were used, as long as the dots were placed at the locations where the grid lines would have intersected. However, subjects' ability to make accurate judgments about relative slant decreased considerably when, prior to the rotation in depth, the dots were placed at random locations on the plane instead of being horizontally and vertically aligned. Their estimate of the total amount of slant portrayed decreased considerably in this case as well. In the random dot images, linear convergence cues were not directly available, and depth could only be inferred from the relative "density" of element spacing (with elements being rather more tightly spaced with proximity toward the vanishing point of the perspective projection).

Phillips [1970] also found that gradients of texture "density", defined in terms of inter-element spacing only and not element size, were fairly poor indicators of the degree of surface slant. Subjects in his experiments judged the relative slant of computer-simulated surfaces textured with randomly distributed circles in which the inter-element spacing was manipulated separately from the element size and compression. Phillips concluded, from the results of these experiments, that subjects were basing their judgments of surface slant almost entirely on the relative sizes and shapes of the individual ellipses, paying relatively little attention to the pattern of distribution of the ellipses across the plane.

The last of the surface slant experiments I will describe were conducted by Cutting and Millard [1984]. These experiments were preference studies, designed to gain insight into the

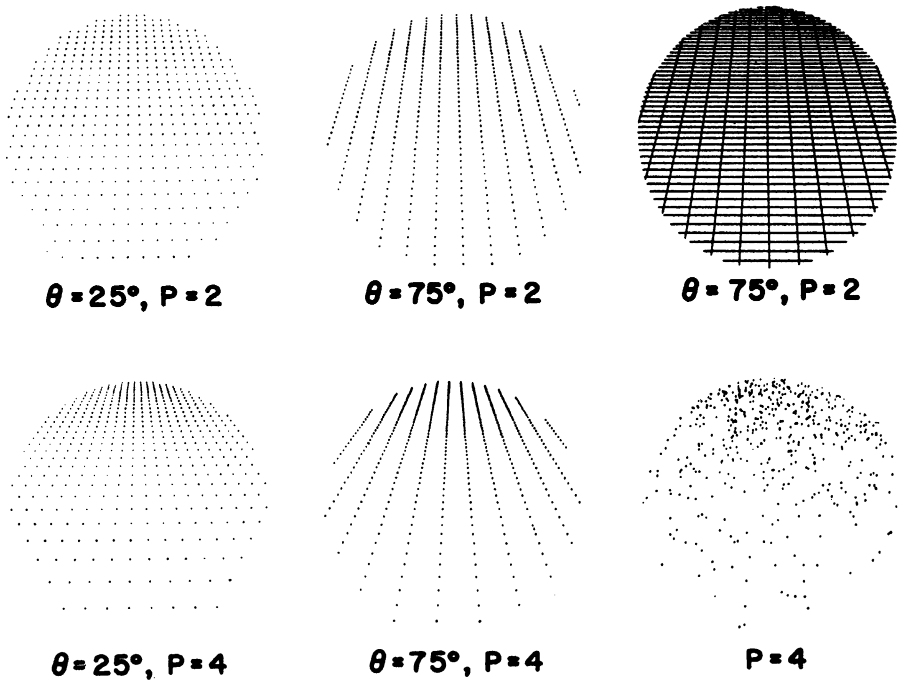



**Figure 3.21: Stimuli used by Braunstein and Payne [1969] in experiments comparing the relative effects on slant judgments of "form ratio" (the relative ratio of vertical height to horizontal width) and "perspective projection" gradients. Steeper gradients of form ratio are manifested in faster vertical compression; steeper gradients of perspective are manifested in a sharper angle of linear convergence. These studies confirmed a dominant effect of linear convergence cues in slant perception.**

kinds of texture characteristics most important for evoking a subjective, pictorial impression of a receding flat plane. Images were displayed in pairs (side-by-side, separated by a single vertical line) on a computer monitor in a moderately lit room and viewed, apparently binocularly, from what seems to have been a distance of about 1.7m. For the first set of stimuli, shown in figure 3.22a, observers were asked to make preference judgments about which looked more like a flat surface receding into the distance. Observers were also shown the stimuli of figure 3.22b and asked to make similar judgments about which looked more like a curved surface receding into the distance. In addition to making a 2-alternative forced choice response for each pair of images, subjects were asked to rate, an a scale from 1–9, the amount by which they felt the preferred stimulus better conveyed the desired impression. Responses were also obtained for sets of stimuli in which irregularly shaped octagons were used instead of circles. Analyzing the results, Cutting and Millard concluded that the "perspective" (size) gradient played the largest role in evoking an impression of a flat surface receding in depth, with a smaller but still significant role played by the "density" (spacing) gradient, and that the "compression" (shape) gradient was the only gradient capable of evoking an impression of a curved surface receding in depth. The original shape (circular or irregularly octagonal) of the elements had no measurable effect on the responses in either case. Although I am not convinced, from looking at the stimuli, that the so-called "density" gradient did not also communicate linear convergence in the case of the flat surfaces (because of the overall trapezoidal shape of the element distribution due to the progressive clustering, with depth, about the midline), the main contribution of these experiments is their clear demonstration that the texture characteristics most useful for conveying information about a receding plane can be significantly different from texture characteristics most useful for conveying information about a curved surface.

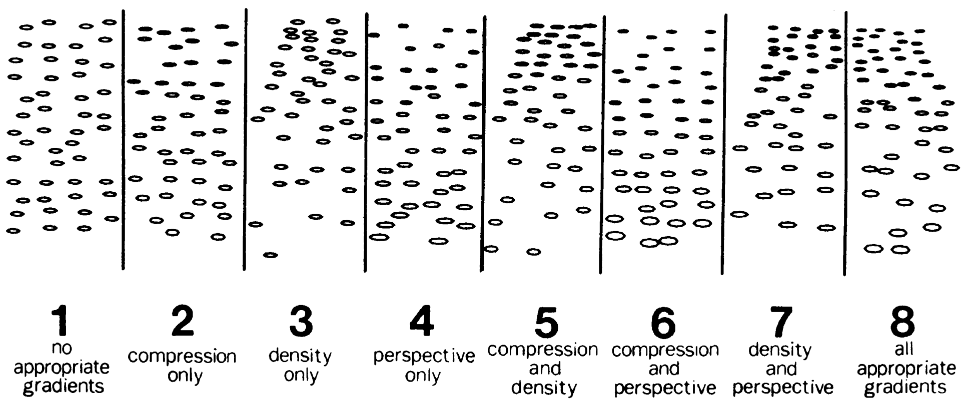



**Figure 3.22a: Representation of some of the stimuli used by Cutting and Millard [1984] in experiments designed to reveal the relative contributions of "perspective" (manifested in terms of element size), "density" (manifested in terms of inter-element spacing, with a tendency for elements to bunch up more towards the center and away from the sides as depth distance increases), and "compression" (manifested as deformation of element shape, in this case from circular to elliptical) cues in evoking a subjective impression of a flat surface receding in depth.**

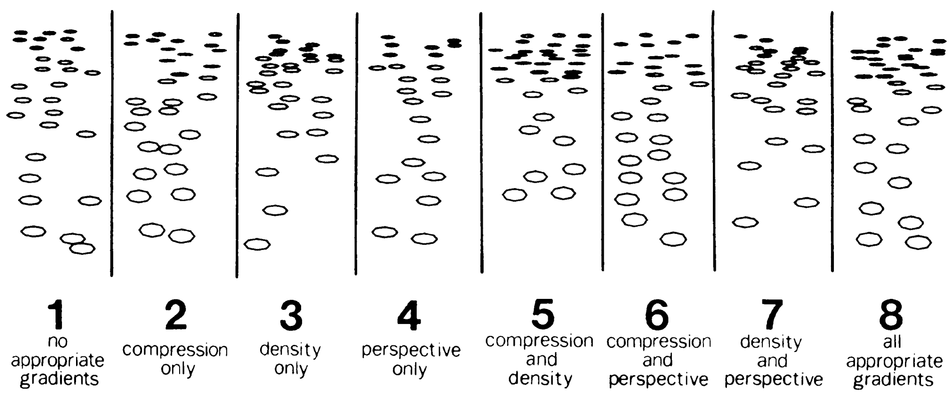



**Figure 3.22b: Representation of some of the stimuli used by Cutting and Millard [1984] to examine the relative weights of "perspective" (size), "density" (spacing) and "compression" (shape) cues in evoking a subjective impression of a smoothly curving, convex surface receding in depth.**

Although absolute slant is consistently found to be underestimated, judgments of relative slant have been shown to be fairly accurate. Stevens and Brookes [1987] demonstrated through a series of experiments that observers were able with great precision to define the apparent slant, indicated by the amount of foreshortening in the slant direction, of a circular probe to match the apparent slant of a smoothly curving developable surface, represented by contour lines, in either orthographic or perspective projection.

In summary, we have seen that texture gradients are useful for indicating the slant of a planar surface, although absolute slant judgments tend to be underestimated. More accurate judgments of slant can be obtained for surfaces covered with one type of texture rather than another, which indicates that if we want to communicate slant information, it makes a difference which particular texture pattern we choose. Attempts to differentiate between the various texture characteristics have focused on estimating the relative contributions of the variables of element size, inter-element spacing, element shape, and element orientation (insofar as it indicates linear convergence), which I will discuss below in roughly descending order of importance. The dominance of linear convergence cues (either explicit or implicit) was demonstrated by Attneave and Olson [1966] and Braunstein and Payne [1969], and hinted at in several of the other studies, including the one by Flock and Moscatelli [1964] in which linear convergence cues are provided by the parallel edges of the square elements in textures 4-6 and by the regular distribution of elements in texture 3, as can be seen in figure 3.19. Although the gradients of element size and inter-element spacing were regarded as together indicative of texture "density" in Gibson's pioneering work, later studies differentiated between the two and indicated a much more important role for element size than inter-element spacing; random dot textures, in which spacing is the only variable changing with depth, were shown to be particular ineffective as indicators of surface slant. Element shape, or the amount of compression in the direction of slant, was also found to be a poor cue to surface slant in the studies by Cutting and Millard [1984], Attneave and Olson [1966] and Braunstein and Payne [1969]. Phillips [1970] found that shape and size together were more effective cues than spacing but did not assess the individual roles of shape vs. size. All of these studies assumed a perspective projection, and Cutting and Millard [1984] argue strongly that perspective is a cue not to be lightly dispensed with.

### 3.7.2: Defining the dimensions of texture characteristics

There are many different ways of characterizing the properties that distinguish various texture patterns, and although this may seem like a fairly intuitive or obvious process, some of the work that has been done to attempt to define the ways in which people differentiate textures bears brief mentioning.

In an attempt to identify the "higher-level" features that people use to characterize different texture patterns (to probe the dimensions and parameters of "texture space"), Rao and Lohse [1993] asked subjects to classify 30 pictures from Brodatz's album on texture [Brodatz 1966]. Using a combination of statistical techniques (hierarchical cluster analysis and nonparametric multidimensional scaling) they found that two dimensions, which they subjectively characterize as representing "periodicity vs. irregularity" and "directionality vs. nondirectionality" described 90% of the variability in the classifications of the two-dimensional texture patterns, and that a third dimension, which they interpret as "structural complexity", accounted for an additional 6% of the variability.

Similarly, Ware and Knight [1992], who were concerned with multi-dimensional data display and saw the possibility of using texture as an additional channel (orthogonal to the three color channels) for communicating quantitative information, identified orientation, size and contrast as the "dimensions" of flat texture most useful in this respect.

### 3.7.3: Texture properties characterizing cylinders and ellipses of various eccentricities

Stevens [1981a] informally reported that when a variety of texture patterns from Brodatz's album were pasted onto cylinders and viewed monocularly under conditions that minimized shading cues, a compelling impression of curvature was obtained only in the cases of the most regular, synthetic patterns (specifically, wire mesh and rattan), despite the obvious foreshortening cues available in all of the texture patterns.

Todd and Akerstrom [1987] were among the first to thoroughly investigate the effects of texture pattern characteristics on the ability of observers to accurately judge the shapes of curved surfaces. They used a cue-conflict paradigm in which they separately manipulated individual aspects of a particular texture pattern (the implementation is described in [Todd and Mingolla 1984]) and measured how eliminating certain gradients affected the ability of observers to judge the relative eccentricities of ellipsoidal surfaces viewed end-on. (Five different eccentricities were used.) An unfortunate repercussion of this type of selective manipulation of the surface texture is that the results in many cases are not "ecologically valid" — they suggest a surface appearance that could have come about only as a result of a highly unlikely coincidence of circumstances — and we cannot rule out the possibility that our response to improbable stimuli might be weighted more by the degree of improbability of the occurrence than by the importance to shape perception of the particular texture gradient eliminated from the pattern. These studies nonetheless contribute a significant portion of what is known about how various properties of surface texture affect our perception of surface shape. Figure 3.23 shows some of the sample stimuli.

One effect clearly shown in these experiments is that subjects perceive more depth in images rendered with a polar (perspective) rather than parallel (orthographic) projection. The effects of projection method can be seen in the images of figure 3.23a–b. The depth difference is not huge, but is significant and consistent. No significant differences in depth judgments were found when surfaces were textured with rectangular elements of varying proportions and areas, as shown in figure 3.23c, rather than square elements of equal size, as long as the element sizes and elongations were randomly perturbed on an individual basis before projection, injecting no bias into the pattern of irregularity. Todd and Akerstrom [1987] note that these results tend to support the theory that observers' judgments of overall shape-from-texture are based on a global perception of the texture gradients rather than on highly local measurements of the lengths or orientations of individual elements, which would be have been disrupted by the noise. When the texture elements were scaled, to eliminate any differences in their projected sizes (and, it appears, inter-element spacing, although this was not explicitly mentioned), as shown in figure 3.23e, judged depth decreased by 37.8%. When the shapes of the projected texture elements were constrained to be square, as shown in figure 3.23f, or the orientations of the texture elements randomly reassigned after projection, as shown in figure 3.23d, the impression of depth was essentially destroyed and subjects reported seeing a flat surface, despite the clear gradients of size and spacing in the image with the square elements. I suggest, however, that the likelihood of

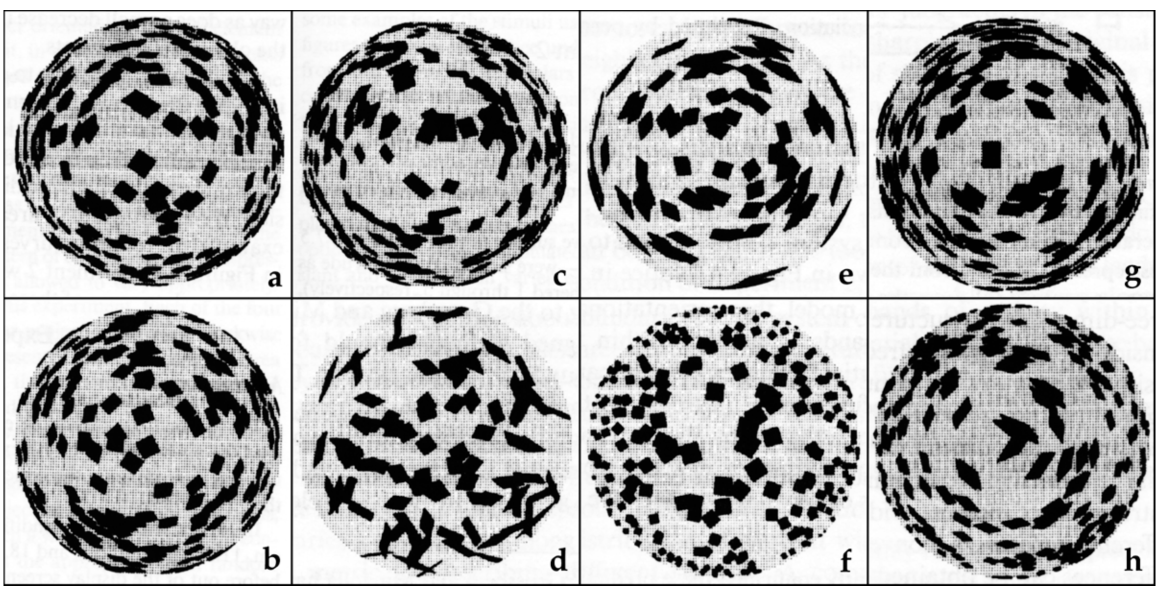



**Figure 3.23: Stimuli, representing textured, elongated ellipsoids viewed end-on, used by Todd and Akerstrom [1987] in experiments designed to reveal the characteristics of texture important for communicating shape. From upper left: a) polar projection, regular texture; b) parallel projection, regular texture; c) irregularly-shaped texture elements; d) randomly reoriented texture elements; e) constant area; f) constant compression; g) elongated elements, veridical texture gradients; h) elongated elements, constant compression.**

a real surface texture producing either of these effects is extremely slim. It would have required an extraordinary coincidence, in the former case, to have dozens of parallelograms precisely scaled and oriented on an ellipsoidal surface so that each one would project to a square, and it can be argued that the images in the latter case lend themselves more naturally to an interpretation of a collection of parallelograms thrown onto a flat dinner plate than anything else. Interestingly, when appropriately oriented (constantly proportioned) parallelograms of equivalent size were substituted for the squares in the case of “constant compression”, as shown in figure 3.22h, the impression of an ellipsoidal surface returned, highlighting the significance of element *orientation* in shape perception, even without the compression cue. Judged depth was consistently greater when slightly elongated elements were used instead of square elements under otherwise normal conditions, as shown in figure 3.23g.

In the final section of their paper, Todd and Akerstrom [1987] presented the image shown in figure 3.24 to suggest potential limitations of the usefulness of discrete element texture as a descriptor of surface shape when it is applied to more complex stimuli. Two specific features of the square element texture seem to most greatly limit its usefulness in this particular situation. The first is that the texture elements are located in the tangent plane to the surface rather than directly on the surface itself. This means that their edges will always be straight and will not curve along with the surface as the contour lines do. (The difference between being on the surface and being in the tangent plane is not great when the underlying surface is well-approximated by a planar patch over the area covered by the texture element, but the aptness of this implementation method will decrease as the scale of the significant surface curvature increases relative to the scale of the texture elements.) The second limiting feature may be that the orientation of the square texture elements is randomly defined. Much of the information communicated by the contour lines in the image on the right comes directly from the fact that these lines have a consistent orientation and are evenly spaced in three dimensions. In a later section I will discuss some work by Todd and Reichel [1990] that follows up on this issue, and in chapter five will discuss my own investigations into the various pros and cons of using different texture pattern characteristics to communicate surface shape.

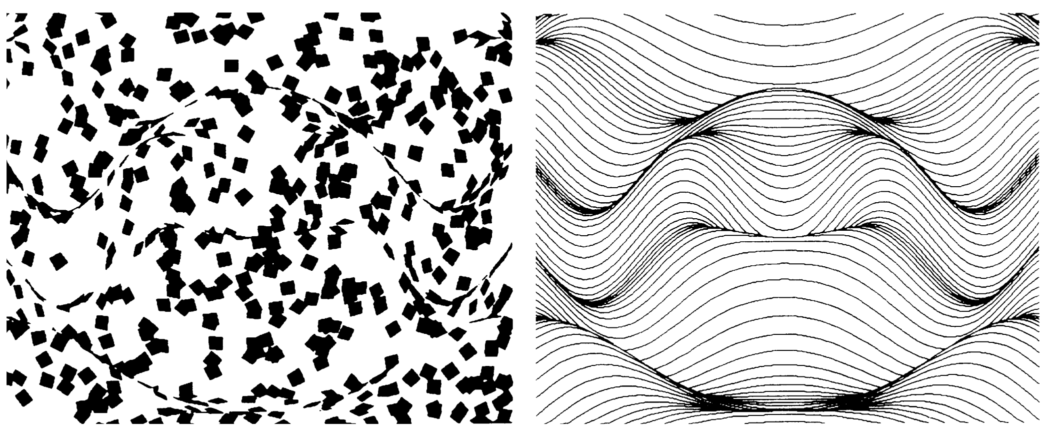


**Figure 3.24: A demonstration, by Todd and Akerstrom [1987], of the results obtained when this same isolated element texture is applied to a more complicated surface, contrasted with the results obtained when a ”contour line” texture (after [Stevens 1981b]) is used.**

Recognizing the potential limitations of relying on a cue-conflict paradigm to estimate the relative contributions of the various gradients of texture most important for communicating shape, Blake *et al.* [1993] conducted experiments in which they compared human performance to the performance of an ideal observer model for making shape judgments from orthographically projected images of parabolic cylinders and found that the accuracy of subjects’ shape judgments was greater than what would be expected if they were using element density (inter-element spacing) information alone. They also show, both theoretically and experimentally (using a cue conflict method for parabolic cylinders covered with line element textures, under orthographic projection) that the "compression" cue (manifested as differences in element elongation) dominates the density cue for communicating eccentricity in depth. When circular texture elements were substituted for line segment elements, response accuracy increased but only modestly. Blake *et al.* [1993] conclude from this that subjects are not inferring local surface orientation from the shapes of individual ellipses but are responding to the circular texture and line element texture in much the same way. Johnston and Passmore [1994] and Koenderink and van Doorn [1995] provide further evidence supporting the notion that shape and curvature perception are based on a global appreciation of surface properties and not built up, point-wise, from local estimates of surface orientation.

Cumming *et al.* [1993] conducted a series of experiments aimed at determining the characteristics of texture that most influenced surface shape judgments under stereoscopic viewing conditions. They used a volumetric texturing technique to apply texture elements to the surface of a horizontally-oriented convex hemi-cylinder, represented in a perspective projection and viewed, in stereo, from a distance of 2 meters. For each texture condition, in which element size, spacing and vertical compression were independently manipulated, the two-dimensional image represented by texture was held constant while the binocular disparity was varied. An estimate of the perceived depth indicated by texture was obtained by having observers make a series of two-alternative forced-choice decisions, under different disparity conditions, about whether the depth of the represented cylinder appeared to be greater or less than half of its height (judging, in effect, whether the surface appeared to represent an elongated or flattened circular cylinder). Examples of the stimuli are shown in figure 3.25. Cumming *et al.* found that perceived depth was least for stimuli in which element compression was held constant, and that the presence or absence of appropriate gradients of element size and/or spacing had little effect on observers' depth judgments. They note that this seems to contradict the results obtained by Todd and Akerstrom [1987], who found a significant effect of element size, in combination with compression, for monocularly viewed stimuli in which square texture elements were applied to elongated ellipsoidal surfaces. Another contradiction, which they unfortunately do not explicitly discuss, is that while Todd and Akerstrom [1987] found that perceived eccentricity increased when elongated texture elements were used, Cumming *et al.* [1993] found exactly the opposite result. (In both studies the direction of elongation was orthogonal to the direction of surface slant.) When horizontally-oriented ellipsoids of randomly varying eccentricities were substituted for the circles in the volume texture, perceived depth decreased with increasing average elongation. Because subjects' depth judgments returned to their previously levels when the elongated ellipsoids were randomly oriented before projection, Cumming *et al.* [1993] conclude that shape-from-texture is ineffective when the surface texture is anisotropic. Another possibility that I see, however, is that the failure of shape-from-texture in this case might be due less to the anisotropic nature of the texture than to the fact that the direction of anisotropy is coincidentally aligned with the cylinder axis and the cylinder axis is parallel to the image plane. Changing either the direction of the texture across the surface or the direction of view could make the compression gradients more easily visible and might improve the amount of depth perceived. (The differences in element compression with respect to position appear, to me, to be much more subtle in the image of figure 3.25-center right than in the image of figure 3.25-lower right, despite the authors' assurances that the overall gradient of compression is identical in all cases. While a clearly perceptible difference in texture characteristics between the "middle and "edge" regions is obviously not sufficient to evoke a veridical perception of eccentricity, it is hard to imagine how such a perception might be facilitated when a visible distinction along these lines cannot be easily made.)

Even with all of the experimental evidence collected so far, several questions about the perception of shape from texture remain open. First, is correct element compression important

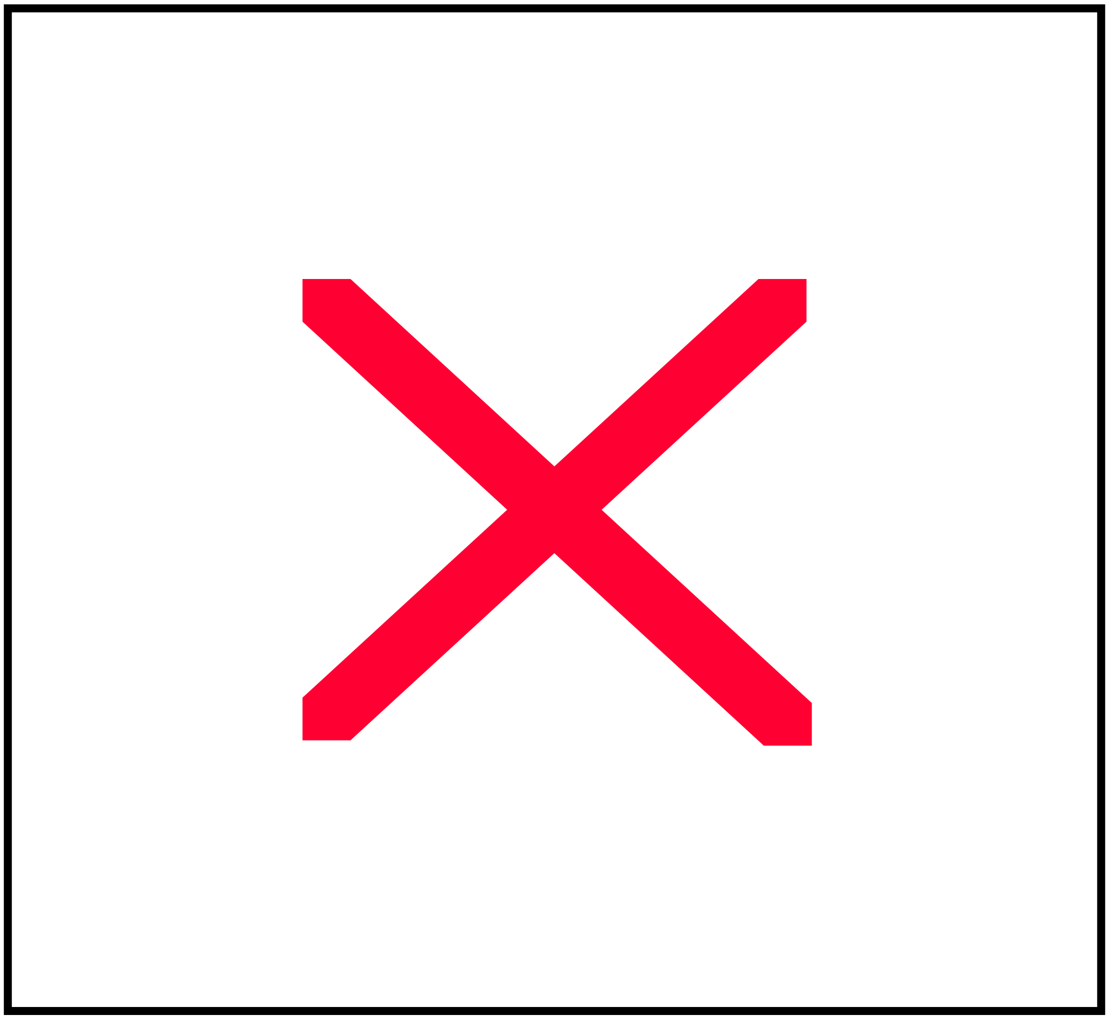


**Figure 3.25: Stimuli used by Cumming *et al.* [1993] in experiments examining the characteristics of texture most important for communicating shape and depth in stereoscopically viewed images. Note the relatively flatter appearance of the images in the middle row as compared with the images in the bottom row.**

for communicating surface shape because of the information contained in the compression gradients (the increasingly squashed or stretched element shapes) or because of the information communicated by the orientation of the elongated texture elements? Conflicting results have been obtained, and at this point the evidence does not seem clear. Second, is it really fair to attribute the abysmal performance on shape-from-texture judgments in the "constant compression" solely to the absence of compression gradients, without considering the fact that the fortuitous combination of circumstances required to produce views such as these from anything other than a flat surface is so unlikely? Nakayama and Shimojo [1992] have provided evidence that our visual system assumes genericity of viewpoint. Perhaps the particular stimuli showing surfaces that lack compression gradients are interpreted as flat because any other interpretation would require the assumption of an "accidental viewpoint" — if these really were curved surfaces, any slight change in viewpoint would result in at least one of the circles being seen as an ellipse or one of the squares being seen as a parallelogram. Third, Cumming *et al.* [1993] hypothesize that at least some of the differences between their results and the results of

Todd and Akerstrom [1987] are due to the different impact of certain texture characteristics under the conditions of stereo viewing. [Rogers and Collett 1989] have shown that stereo and motion cues to depth are not duplicative, i.e., that more depth is perceived in displays containing both rather than either alone. It would be interesting to investigate the effects of different texture cues on judgments of shape in experiments in which both stereo and interactive object manipulation capabilities are provided. In particular, this approach would allow any viewpoint-dependent effects to be clearly identified. A related point is that all of the experiments conducted so far in the perception of shape from texture have relied on judgments of the relative elongations of very simple surfaces of known, symmetric geometry. It would be interesting to look at the ability of observers to make shape and curvature judgments (using Koenderink's [1990] richer definitions of these terms) for a broader range of surface stimuli. In particular, this approach would allow shape-dependent effects to be identified and might help shed more light on the question of why elongated elements seem to aid the perception of depth in some cases and hinder it in others.

### 3.7.4: Insights from computational models for extracting shape from texture

While the sheer volume of work in computer vision on extracting shape from texture (or other surface markings) defies easy or concise summation (and I will not go into the details of the various computational models for extracting shape from texture), some insights of relevance to the topic of this dissertation can be gained from that work. Most computational models for extracting shape information from patterns of surface texture operate either from an assumption of homogeneous texture element distribution (in which case shape and depth are indicated by differences in inter-element spacing after projection) or from an assumption of isotropic element orientation (in which case shape, but not depth, is indicated by the orientation anisotropies in the projected texture). In many cases, the algorithms are designed to be applied to planar surfaces and generalize to curved surfaces only through the assumption that any particular local surface area can be approximated by a planar patch. In many cases, the algorithms assume orthographic rather than perspective projection, ostensibly on the grounds that the differences between the two are minimal as long as the depth extent of the surface is small relative to the distance of the surface from the viewpoint, although a more practical explanation might be that incorporating perspective would severely complicate the algorithmic complexity. One of the earliest and most often followed computational approaches for determining shape from texture was proposed by Witkin [1981]. As shown in figure 3.26, Witkin observed that the distribution of the orientations of elongated texture elements will be biased toward a direction aligned with the tilt axis (where the tilt axis is defines as the direction, in the image plane, about which the surface plane is slanted), and that this orientation bias will be stronger as the amount of slant increases (where slant is defined as the angle between the surface normal and the viewing direction). With particular insight, Witkin emphasized that the local orientation of a surface could be successfully computed from the pattern of foreshortening of surface texture even in situations where element orientation was not isotropic, as long as it was possible to distinguish the inherent anisotropies of

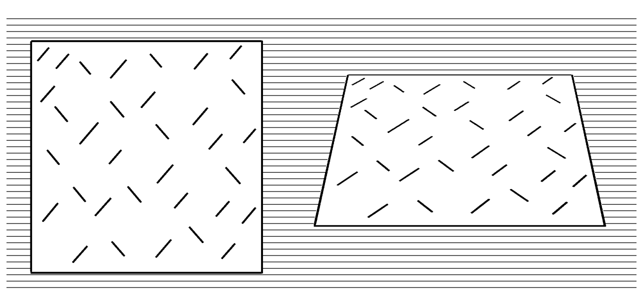


**Figure 3.26: An illustration of the type of texture anisotropy generated by surface slant, as described by Witkin [1981].**

the texture pattern from the effects of compression due to the 3D-to-2D projection. Stone [1993] has enlarged upon this theme, noting that human observers seem to presume "homotropy" of texture element orientation rather than either isotropy or homogeneity of element distribution — that we confuse flat and curved surfaces most when the orientations of texture elements are not invariant with respect to position across the plane. Figure 3.27 illustrates the difference between homotropic and non-homotropic texture patterns. It is difficult not to interpret the patterns of shifting element orientation as cues to non-planarity. Stone also provides what appears to be a clear and thorough summary of the computations shape-from-texture literature; other references to the literature in this area can be found appended to the references for this chapter.

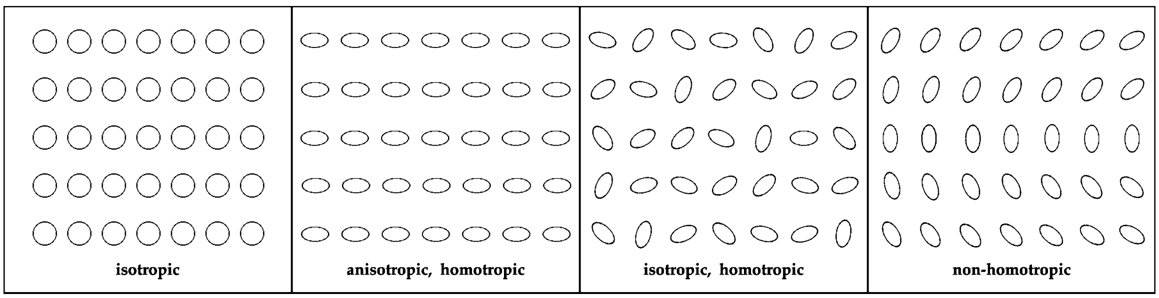



**Figure 3.27: An illustration of the difference between homotropic and non-homotropic texture, after Stone [1993].**

### 3.7.5: Representing a curved surface with "contour" lines

Stevens [1981b] emphasizes the particular usefulness for describing shape of surface texture patterns in which elements are oriented along the principal directions. His claim that we can safely assume that many if not most of the lines that we observe on surfaces will in fact be oriented in these directions is difficult to accept; nevertheless his analysis of the kinds of shape information that can be communicated by these line patterns is of particular significance when one is considering the issues of texture design — how to define a texture that will show shape best.

Stevens asserts that if we assume that two intersecting lines of "texture" on a surface (defined as a height function on a plane receding in depth towards the top of the projected image) are orthogonal in 3-space and each locally oriented in the directions of greatest and least normal curvature over the surface, and if we furthermore assume a generic viewpoint (precluding the possibility of, for example, looking at a hyperbolic surface along an asymptotic direction), we can infer the sign of the Gaussian curvature of the local surface patch from the signs of the curvatures in the projection onto the image plane of each of the lines at their intersection point. Figure 3.28, based on a diagram in [Stevens 1981b], attempts to illustrate both the potential and the limitations of this concept. Not pointed out by Stevens is the fact that unless one presumes a surface normal direction pointing outward towards the viewing direction and upwards toward the top of the image, it is not possible to distinguish either between convex and concave surface patches or between elliptic and hyperbolic patches.

While it would seem that the restrictiveness and lack of plausible generality of the these above-mentioned assumptions would mean that, under practical circumstances, it would be difficult for us to actually infer any useful information about shape from the curvatures of lines across surfaces, Todd and Reichel [1990] present compelling empirical evidence, reproduced in figure 3.29, of the effectiveness with which surface contours can be used to convey a robust impression of three-dimensional shape under a wide variety of slightly irregular circumstances. One of the things they demonstrated was that the compelling nature of the three-dimensional

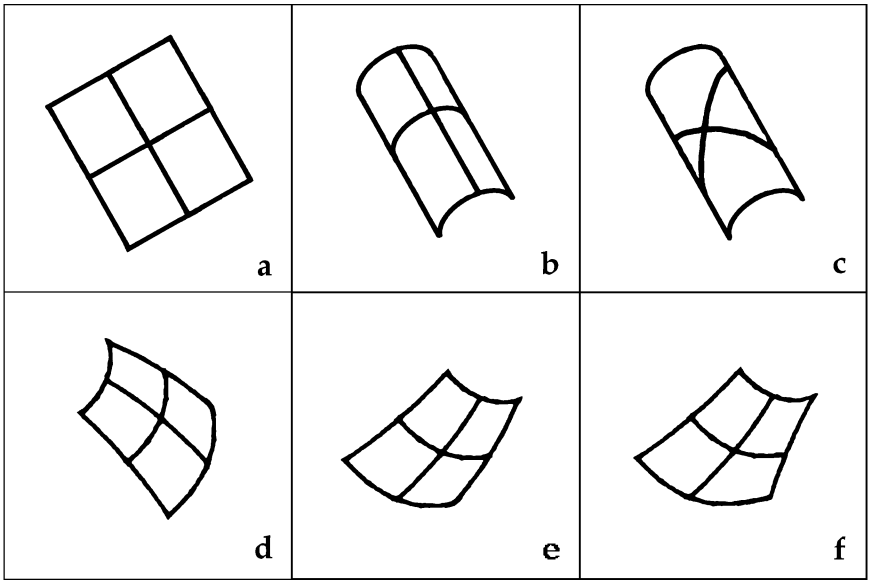



**Figure 3.28: A diagram representing how one might imagine that local surface shape could be inferred from the signs of the curvatures, in the projected image, of surface "contour" lines oriented in each of the principal directions: a) a planar patch, representing a degenerate case in which $k_1 = k_2 = 0$ (normal curvature is zero in all directions on this surface); b) a cylindrical patch, where $k_1 > 0$ and $k_2 = 0$; c) a misleading picture showing two orthogonal directions on a cylindrical patch that are not aligned with the principal directions (in this case the normal curvature is positive in each of the orthogonal directions); d) an elliptical patch, where $k_1$, $k_2$ >0, and which appears, in this orientation, to be convex; e) the same elliptical patch rotated by 90° in the clockwise direction (here one may prefer a concave interpretation); f) a hyperbolic patch, where SIGN( $k_1$) ≠ SIGN( $k_2$). Notice that the apparent curvatures of the two principal direction lines in this patch are identical to the apparent curvatures of the principal direction lines in the patch to the left of it. The hyperbolic nature of this surface patch is revealed *not* by the curvatures of the principal direction lines but rather by the fact the curvatures of the left and right edges differ in sign, indicating the presence of an inflection in the projected curvature of one of the principal direction lines.**

shape perception is not seriously diminished if instead of being continuous the lines are broken up into short segments centered at randomly-spaced locations across the surface, as in figure 3.29c (although the impression of depth is lost when the line segments lose their capacity to convey orientation information, as in figure 3.29d). They also demonstrated that if, instead of being equally spaced and oriented parallel to the axis of rotation, as in figure 3.29a, the contour lines are randomly perturbed in orientations by up to 45°, as in figure 3.29e, their capacity to portray shape is not significantly weakened, although the impression of depth is lost if the line orientation is completely random, as in figure 3.29f. Similarly, when the line segments instead of flowing horizontally are rotated on the surface to flow in a consistent, oblique direction, as in figure 3.29b, or random noise is added to the directions of each of the lines as they proceed across the surface, as in figure 3.29g, the textures retain their ability to communicate shape and depth, although the impression of depth may be lost if the lines are replaced by sinusoids and texture compression is not fully represented, as shown in figure 3.29h of the lower block. One may note that the understanding of depth obtained from these images is diminished if they are viewed at an angle other than the one in which they appear in figure 3.29. Reichel and Todd [1990] attribute this effect to the presence of occlusion contours that restrict the depth order of the surfaces on either side of the occlusion boundary, making the depth relationships appear to be piecewise inconsistent, as shown in figure 3.30.

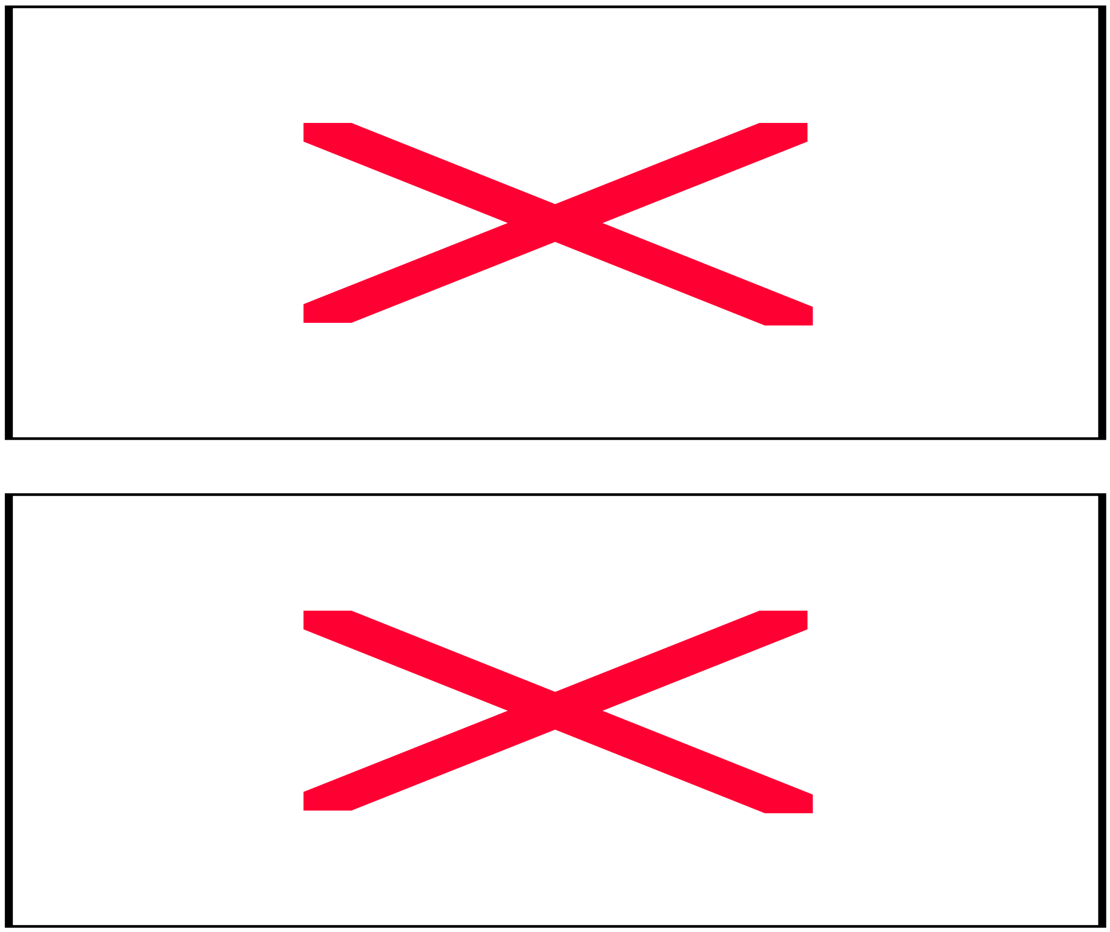


**Figure 3.29: Images generated by Todd and Reichel [1990] to demonstrate the wide variety of conditions under which surface "contours" (texture lines) can, and cannot, convey a compelling impression of three-dimensional shape. In the lower group of pictures Todd and Reichel have selectively projected only portions of the texture, to avoid what they refer to as "shading cues", the dark spots that result from extreme texture compression where the view direction almost grazes the surface.**

While Todd and Reichel do not attempt to draw any general conclusions about the characteristics of surface lines that make them useful for showing shape (it would be premature to make any such statements on the basis of just a few example images), it is clear that surface lines can usefully show shape in more general situations than those described by Stevens.

The surfaces used in all of the above examples are not truly three-dimensional but are perhaps better described as $2\frac{1}{2}$D, representing height functions on a plane. We have already seen that the particular texture characteristics useful for conveying an impression of a flat plane receding far into depth did not to generalize to the case of curved surfaces with limited depth extents. The apparent demonstrated dependence of some of these effects on assumptions about "floor" as opposed to "ceiling" surfaces seems to indicate that similar caution must be used in making assumptions about the direct applicability to surfaces in 3D of the texture properties shown to be effective in the $2\frac{1}{2}$D domain.

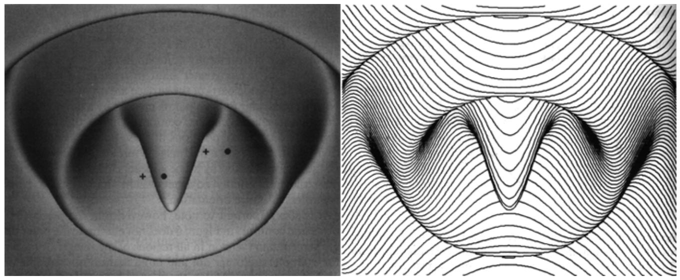


**Figure 3.30: Images generated by Reichel and Todd [1990] to demonstrate how occlusion contours can affect the global consistency of locally inferred relative depth order judgments. When this surface is interpreted as receding in depth towards the top of the image rather than towards the bottom, the perceived local depth order between the upper pair of points will be inconsistent with the perceived local depth order between the lower pair of points.**

The usefulness of surface contours for conveying shape seems to come from their being arranged consistently across the surface in a flow. There is actually very little that one can assume about shape from the projected curvature of a single surface curve. Cipolla and Zisserman [1992] investigated the information about surface shape that could be communicated by arbitrary surface curves tracked through a series of different viewpoints. They show how, by following the inflections of a surface curve, knowing little more than the general direction of observer motion (left-right or right-left), it is possible to recover the sign of the normal curvature along that curve and disambiguate some aspects of the surface shape. The projected curvature, in two dimensions, of a point on an arbitrary line on a surface in 3D can be zero under only two conditions: either the surface is locally flat at that point (which occurs non-generically) or the viewing direction is within the osculating plane, defined as the plane spanned by the normal and tangent directions of the curve at the point (which is a generic event). Based on this understanding, Cipolla and Zisserman observe that it is possible in general to track the location of the inflection point of a space curve along the curve as the viewpoint changes, and they prove that the sign of the normal curvature can be derived from this information if the general direction of changing viewpoint is known. They show, in effect, that when the viewpoint is moving from left to right and the inflection moves along the curve so that a point at which curvature was zero in one view is seen to have positive (negative) curvature in the next view, the normal curvature at the point can be assumed to be convex (concave). This information can be used to disambiguate the shape of the surface in a similar way as the information from moving specularities: the surface cannot be convex if the normal curvature is known to be concave in some direction; it cannot be concave if the reverse is true.

Although this treatment of shape and depth perception has by no means been exhaustive, I hope that it has provided some insights that may help explain why the shape and depth of transparent surfaces are so difficult to perceive, support the idea that a partial solution may lie in the use of opaque surface texture, and motivate possible approaches for defining a sparse, opaque texture that can convey surface shape in a particularly representative way.

### References for chapter three


Louvre: guide to the collections, (Photographic credits: Réunion des musées nationaux), Hérissey, 1991.

Fred Attneave and Richard K. Olson. "Inferences About Visual Mechanisms from Monocular Depth Effects", *Psychonomic Science*, vol. 4, pp. 133-134, 1966.

Andrew Blake, Heinrich H. Bülthoff. and David Sheinberg. "Shape from Texture: Ideal Observers and Human Psychophysics", *Vision Research*, vol. 33, no. 12, pp. 1723-1737, 1993.

Andrew Blake and Heinrich Bülthoff. "Shape from Specularities: computation and psychophysics", *Philosophical Transactions of the Royal Society of London*, B, vol. 331, pp. 237-252, 1991.

Umberto Boccioni. *Unique Forms of Continuity in Space*, 1913.

Myron L. Braunstein, George J. Andersen, Michael W. Rouse and James S. Tittle. "Recovering Viewer-Centered Depth from Disparity, Occlusion, and Velocity Gradients", *Perception & Psychophysics*, vol. 40, no. 4, pp. 216-224, 1986.

Myron Braunstein. Depth Perception Through Motion, Academic Press, 1976.

Myron L. Braunstein and John W. Payne. "Perspective and Form Ratio as Determinants of Relative Slant Judgments", *Journal of Experimental Psychology*, vol. 81, no. 3, pp. 584-590, 1969.

Myron L. Braunstein. "Sensitivity of the Observer to Transformations of the Visual Field", *Journal of Experimental Psychology*, vol. 72, no. 5, pp. 683-689, 1966.

Myron L. Braunstein. "Depth Perception in Rotating Dot Patterns: Effects of Numerosity and Perspective", *Journal of Experimental Psychology*, vol. 64, no. 4, pp. 415-420, 1962.

David Brewster. "On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena", *Edinburgh Journal of Science*, vol. 4, pp. 99-108, 1826. (published anonymously)

Phil Brodatz. Textures; a photographic album for artists and designers, Dover, 1966.

Vicki Bruce and Patrick R. Green. Visual Perception: Physiology, Psychology and Ecology, second edition, Lawrence Erlbaum Associates Ltd., 1990.

Nicola Bruno and James E. Cutting. "Minimodularity and the Perception of Layout", *Journal of Experimental Psychology: General,* vol. 117, no. 2, pp. 161-170, 1988.

David Buckley and John P. Frisby. "Interaction of Stereo, Texture and Outline Cues in the Shape Perception of Three-Dimensional Ridges", *Vision Research*, vol. 33, no. 7, pp. 919-933, 1993.

Heinrich H. Bülthoff. "Shape from X: Psychophysics and Computation", in Computational Models of Visual Processing, Michael S. Landy and J. Anthony Movshon, eds., MIT Press, 1991.

Heinrich H. Bülthoff and Hanspeter A. Mallot. "Integration of depth Modules: Stereo and Shading", *Journal of the Optical Society of America, A*, vol. 5, no. 10, pp. 1749-1758, October 1988.

Patrick Cavanagh and Yvan G. Leclerc. "Shape from Shadows", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15, no. 1, pp. 3-27, 1989.

Roberto Cipolla and Andrew Zisserman. "Qualitative Surface Shape from Deformation of Image Curves", *International Journal of Computer Vision*, vol. 8, no. 1, pp. 53-69, 1992.

James J. Clark and Alan L. Yuille. Data Fusion for Sensory Information Processing Systems, Kluwer Academic Publishers, 1990.

Clodion (Claude Michel). *Nymphs Bathing with Leda and the Swan*, 1782.

Ben E. Coutant and Gerald Westheimer. "Population Distribution of Stereoscopic Ability", *Opthalmic and Physiological Optics*, vol. 13, no. 1, pp. 3-7, January 1993.

Bruce G. Cumming, Elizabeth B. Johnston and Andrew J. Parker, "Effects of Different Texture Cues on Curved Surfaces Viewed Stereoscopically", *Vision Research*, vol. 33, no. 5/6, pp. 827-838, 1993.

James E. Cutting and Robert T. Millard. "Three Gradients and the Perception of Flat and Curved Surfaces", *Journal of Experimental Psychology: General*, vol. 113, no. 2, pp. 198-216, 1984.

Wim van Damme. "Active Vision: exploration of three-dimensional structure", Doctoral Dissertation, University of Utrecht, September 1994.

Frank H. Durgin, Dennis R. Proffitt, Thomas J. Olsen and Karen S. Reinke. "Comparing Depth from Motion with Depth from Binocular Disparity", *Journal of Experimental Psychology: Human Perception and Performance,* vol. 21, no. 3, pp. 679-699, 1995.

Hiroyuki Egusa. "Effects of Brightness, Hue, and Saturation on Perceived Depth Between Adjacent Regions in the Visual Field", *Perception*, vol. 12, pp. 167-175, 1983.

James Thomas Enright. "Art and the Oculomotor System: perspective illustrations evoke vergence changes", *Perception*, vol. 16, pp. 731-746, 1987.

Roderik G. F. Erens, Astrid M. L. Kappers and Jan J. Koenderink. "Perception of Local Shape from Shading", *Perception & Psychophysics*, vol. 54, no. 2, pp. 145-156, 1993.

S. Kay Fisher and Kenneth J. Ciuffreda. "Accommodation and Apparent Distance", *Perception*, vol. 17, pp. 609-621, 1988.

Howard R. Flock and Anthony Moscatelli. "Variables of Surface Texture and Accuracy of Space Perceptions", *Perceptual and Motor Skills*, vol. 19, pp. 327-334, 1964.

David Forsyth and Andrew Zisserman. "Shape from Shading in the Light of Mutual Illumination", *Image and Vision Computing*, vol. 8, no. 1, pp. 42-49, February 1990.

Christopher Frayling, Helen Frayling and Ron van der Meer. <u>The Art Pack. A unique, three-dimensional tour through the creation of art over the centuries: what artists do, how they do it, and the masterpieces they have given us</u>, Alfred A. Knopf, 1992.

James J. Gibson and Eleanor J. Gibson, "Continuous Perspective Transformations and the Perception of Rigid Motion", *Journal of Experimental Psychology*, vol. 54, no. 2, pp. 129-138, 1957.

James J. Gibson. <u>The Perception of the Visual World</u>, Houghton Mifflin, 1950(a).

James J. Gibson. "The Perception of Visual Surfaces", *American Journal of Psychology*, vol. 63, pp. 367-384, 1950(b).

Barbara Gillam. "Judgments of Slant of the Basis of Foreshortening", *Scandinavian Journal of Psychology*, vol. 11, pp. 31-34, 1970.

Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg and Bennett Battaile. "Modeling the Interaction of Light Between Diffuse Surfaces", *Computer Graphics*, vol. 18, no. 3, pp. 213-222, July 1984.

Vernon W. Grant. "Accommodation and Convergence in Visual Space Perception", *Journal of Experimental Psychology*, vol. 31, no. 2, pp. 89-104, August 1942.

Bert F. Green. "Figure Coherence in the Kinetic Depth Effect", *Journal of Experimental Psychology*, vol. 62, no. 3, pp. 272-282, 1961.

Richard L. Gregory. <u>The Intelligent Eye</u>, McGraw-Hill, 1970.

Erik de Haan, Roderik G. F. Erens and André J. Noest. "Shape from Shaded Random Surfaces", *Vision Research*, vol. 35, no. 21, pp. 2985-3001, 1995.

Hermann Ludwig Ferdinand von Helmholtz. <u>Treatise on Physiological Optics, vol. 3: The Perceptions of Vision,</u> (translated from the Third German Edition) James P. C. Southall, ed., The Optical Society of America, 1925.

Claes von Hofsten. "The Role of Convergence in Visual Space Perception", *Vision Research*, vol. 16, pp. 193-198, 1976.

Berthold K. P. Horn and Michael J. Brooks. <u>Shape from Shading</u>, MIT Press, 1989.

Jean-Antoine Houdon. <u>*Diana the Huntress*</u>. 1790.

Ian P. Howard, Sten S. Bergström, and Masao Ohmi. "Shape from Shading in Different Frames of Reference", *Perception*, vol. 19, pp. 523-530, 1990.

William H. Ittelson and Franklin P. Kilpatrick. "Experiments in Perception", *Scientific American*, vol. 185, pp. 50-55, 1952.

William H. Ittelson. "Size as a Cue to Distance: Static Localization", *American Journal of Psychology*, vol. 64, pp. 54-67, 1951(a)

William H. Ittelson. "Size as a Cue to Distance: Radial Motion", *American Journal of Psychology*, vol. 64, pp. 188-202, 1951(b).

Alan Johnston and Peter J. Passmore. "Shape from Shading. I: Surface curvature and orientation", *Perception*, vol. 23, pp. 169-189, 1994.

Elizabeth B. Johnston, Bruce G. Cumming and Andrew J. Parker. "Integration of Depth Modules: Stereopsis and Texture", *Vision Research*, vol. 33, no. 5/6, pp. 813-826, 1993.

Elizabeth B. Johnston. "Systematic Distortions of Shape from Stereopsis", *Vision Research*, vol. 31, no. 7/8, pp. 1351-1360, 1991.

Bela Julesz. "Stereoscopic Vision", *Vision Research*, vol. 26, no. 9, pp. 1601-1612, 1986.

Bela Julesz. “Binocular Depth Perception without Familiarity Cues”, *Science*, vol. 145, pp. 356-362, 1964.

Lloyd Kaufman. Sight and Mind: an introduction to visual perception, Oxford University Press, 1974.

Daniel Kersten and Gordon E. Legge. “Convergence Accommodation”, *Journal of the Optical Society of America*, vol. 73, no. 3, pp. 332-338, March 1983.

Jan J. Koenderink and Andrea J. van Doorn. “Relief: pictorial and otherwise”, *Image and Vision Computing*, vol. 13, no. 5, pp. 321-334, June 1995.

Jan J. Koenderink and Andrea A. van Doorn. “Illuminance Critical Points on Generic Smooth Surfaces”, *Journal of the Optical Society of America, A*, vol. 10, no. 5, pp. 844-854, May 1993.

Jan J. Koenderink. Solid Shape, MIT Press, 1990.

Jan J. Koenderink. “Optic Flow”, *Vision Research*, vol. 26, no. 1, pp. 161-180, 1986.

Jan J. Koenderink and Andrea J. van Doorn. “Photometric Invariants Related to Solid Shape”, *Optica Acta*, vol. 27, no. 7, pp. 981-996, July 1980.

J. J. Koenderink and A. J. van Doorn. “Local Structure of Movement Parallax of the Plane”, *Journal of the Optical Society of America*, vol. 66, no. 7, pp. 717-723, 1976.

Melvin K. Komoda and Hiroshi Ono. “Oculomotor Adjustments and Size-Distance Perception”, *Perception & Psychophysics*, vol. 15, no. 2, pp. 353-360, 1974.

Kevin Krisciunas and Bill Yenne. The Pictorial Atlas of the Universe, Mallard Press, 1989.

Victor Klymenko and Naomi Weisstein. “Spatial Frequency Differences Can Determine Figure-Ground Organization”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 12, no. 3, pp. 324-330, 1986.

H. Lehmann. “Formen landschaftlicher Raumerfahrung im Spiegel der bildenden Kunst”, *Mitteilungen d. Fränk. Geograph. Ges. Erlangen*, vol. 13/14, 1966/67.

James Lipscomb. Three-dimensional Cues for a Molecular Computer Graphics System, Doctoral dissertation, University of North Carolina, 1981.

Margaret S. Livingstone and David H. Hubel. “Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement, and Depth”, *The Journal of Neuroscience*, vol. 7, no. 11, pp. 3416-3468, 1987.

Jonathan A. Marshall, Christina A. Burbeck, Dan Ariely, Jannick P. Rolland and Kevin E. Martin. “Occlusion Edge Blur: a cue to relative visual depth”, to appear in *Journal of the Optical Society of America, A*, 1996.

Harry McGurk and Gustav Jahoda. “The Development of Pictorial Depth Perception: the role of figural elevation”, *British Journal of Psychology*, vol. 65, no. 3, pp. 367-376, August 1974.

Wolfgang Metzger. Gesetze des Sehens. W. Kramer & Co., 1975.

Ennio Mingolla and James T. Todd. “Perception of Solid Shape from Shading”, *Biological Cybernetics*, vol. 53, pp. 137-151, 1986.

Ken Nakayama and Shinsuke Shimojo. “Experiencing and Perceiving Visual Surfaces”, *Science*, vol. 257, pp. 1357-1363, 4 September 1992.

Ken Nakayama, Shinsuke Shimojo and Gerald H. Silverman. “Stereoscopic Depth: its relation to image segmentation, grouping, and the recognition of occluded objects”, *Perception*, vol. 18, pp. 55-68, 1989.

Mark Nawrot and Randolph Blake. “The Interplay Between Stereopsis and Structure from Motion”, *Perception & Psychophysics*, vol. 49, no. 3, pp. 230-244, 1991.

Robert P. O’Shea, Shane G. Blackburn and Hiroshi Ono. “Contrast as a Depth Cue”, *Vision Research*, vol. 34, no. 12, pp. 1595-1604, 1994.

Andrew J. Parker, Chris Christou, Bruce G. Cumming, Elizabeth B. Johnston, Michael J. Hawken and Andrew Zisserman. “The Analysis of 3D Shape: Psychophysical Principles and Neural Mechanisms”, Understanding Vision, Glyn W. Humphreys, ed., Blackwell, pp. 143-179, 1992.

Alex P. Pentland. “A New Sense for Depth of Field”, *IJCAI ‘85: Proceedings of the 9th International Joint Conference on Artificial Intelligence*, vol. 2, pp. 988-994, 1985.

David I. Perrett and Mark H. Harries. “Characteristic Views and the Visual Inspection of Simple Faceted and Smooth Objects: ‘Tetrahedra and Potatoes’”, *Perception*, vol. 17, pp. 703-720, 1988.

Richard J. Phillips. “Stationary Visual Texture and the Estimation of Slant Angle”, *Quarterly Journal of Experimental Psychology*, vol. 22, pp. 389-397, 1970.

M. H. Pirenne. "Vision and Art", in Handbook of Perception: Volume V, Seeing, Edward C. Carterette and Morton P. Friedman, eds., Academic Press, pp. 433-490, 1975.

Gian F. Poggio and Tomaso Poggio. "The Analysis of Stereopsis", *Annual Review of Neuroscience*, vol. 7, pp. 379-412, 1984.

V. S. Ramachandran, S. Cobb and D. Rogers-Ramachandran. "Perception of 3-D Structure from Motion: The role of velocity gradients and segmentation boundaries", *Perception & Psychophysics*, vol. 44, no. 4, pp. 390-393, 1988.

Vilayanur S. Ramachandran. "Perceiving Shape from Shading", *Scientific American*, vol. 259, no. 2, pp. 76-83, August 1988.

A. Ravishankar Rao and Gerald L. Lohse. "Identifying High Level Features of Texture Perception", *CVGIP: Graphical Models and Image Processing,* vol. 55, no. 3, pp. 218-233, May 1993.

Francene D. Reichel and James T. Todd. "Perceived Depth Inversion of Smoothly Curved Surfaces Due to Image Orientation", *Journal of Experimental Psychology: Human Perception and Performance,* vol. 16, no. 3, pp. 653-664, 1990.

Whitman Richards. "Visual Space Perception", in Handbook of Perception: Volume V, Seeing, Edward C. Carterette and Morton P. Friedman, eds., Academic Press, pp. 351-368, 1975.

Brian Rogers and Thomas S. Collett. "The Appearance of Surfaces Specified by Motion Parallax and Binocular Disparity", *The Quarterly Journal of Experimental Psychology*, vol. 41A, no. 4, pp. 697-717, 1989.

Brian Rogers and Maureen Graham. "Similarities Between Motion Parallax and Stereopsis in Human Depth Perception", *Vision Research*, vol. 22, pp. 261-270, 1982.

Brian Rogers and Maureen Graham. "Motion Parallax as an Independent Cue for Depth Perception", *Perception*, vol. 8, pp. 125-134, 1979.

Ann Saye and John P. Frisby. "The Role of Monocularly Conspicuous Features in Facilitating Stereopsis from Random-Dot Stereograms", *Perception*, vol. 4, pp. 159-171, 1975.

Robert Sekuler and Randolph Blake. Perception, third edition, McGraw-Hill, 1994.

George Sperling, Michael S. Landy, Barbara A. Dosher and Mark E. Perkins. "Kinetic Depth Effect and Identification of Shape", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15, no. 4, pp. 826-840, 1989.

Kent A. Stevens. and Allen Brookes. "Probing Depth in Monocular Images", *Biological Cybernetics*, vol. 56, pp. 355-366, 1987.

Kent A. Stevens. "Inferring Shape from Contours Across Surfaces", in From Pixels to Predicates: recent advances in computational and robotic vision, Alex P. Pentland, ed., pp. 93-110, Ablex, 1986.

Kent A. Stevens. "The Information Content of Texture Gradients", *Biological Cybernetics*, vol. 42, pp. 95-105, 1981(a).

Kent A. Stevens. "The Visual Interpretation of Surface Contours", *Artificial Intelligence*, vol. 17, pp. 47-73, 1981(b).

James V. Stone. "Shape From Local and Global Analysis of Texture", *Philosophical Transactions of the Royal Society of London, B*, vol. 339, pp. 53-65, 1993.

James S. Tittle, James T. Todd, Victor J. Perotti and J. Farley Norman. "Systematic Distortion of Perceived Three-Dimensional Structure From Motion and Binocular Stereopsis", *Journal of Experimental Psychology: Human Perception and Performance,* vol. 21, no. 3, pp. 663-678, 1995.

James T. Todd and Francene D. Reichel. "Visual Perception of Smoothly Curved Surfaces from Double-Projected Contour Patterns", *Journal of Experimental Psychology: Human Perception and Performance,* vol. 16, no. 3, pp. 665-674, 1990.

James T. Todd and Francene D. Reichel. "Ordinal Structure in the Visual Perception and Cognition of Smoothly Curved Surfaces", *Psychological Review*, vol. 96, no. 4, pp. 643-657, 1989.

James T. Todd and Robin A. Akerstrom. "Perception of Three-Dimensional Form from Patterns of Optical Texture", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 13, no. 2, 242-255, 1987.

James T. Todd and Ennio Mingolla. "Simulation of Curved Surfaces from Patterns of Optical Texture", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10, no. 5, pp. 734-739, 1984.

James T. Todd and Ennio Mingolla. "Perception of Surface Curvature and Direction of Illumination from Patterns of Shading", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 9, no. 4, pp. 583-595, 1983.

Sjoerd C. de Vries, Astrid M. L. Kappers and Jan J. Koenderink. "Influence of Surface Attitude and Curvature Scaling on Discrimination of Binocularly Presented Curved Surfaces", *Vision Research*, vol. 34, no. 18, pp. 2409-2423, 1994.

Hans Wallach and Carl Zuckerman. "The Constancy of Stereoscopic Depth", *American Journal of Psychology*, vol. 76, pp. 404-412, 1963.

Hans Wallach, Alexander Weisz and Pauline Austin Adams. "Circles and Derived Figures in Rotation", *American Journal of Psychology*, vol. 69, pp. 48-59, 1956.

Hans Wallach and D. N. O'Connell. "The Kinetic Depth Effect", *Journal of Experimental Psychology*, vol. 45, no. 4, pp. 205-217, April 1953.

Leonard R. Wanger, James A. Ferwerda and Donald P. Greenberg. "Perceiving Spatial Relationships in Computer-Generated Images", *IEEE Computer Graphics and Applications*, vol. 12, no. 3, May 1992.

Nicholas J. Wade, ed. Brewster and Wheatstone on Vision, Academic Press, 1983.

Colin Ware and William Knight. "Orderable Dimensions of Visual Texture for Data Display: Orientation, Size and Contrast", *Proceedings of CHI '92, the ACM Conference on Human Factors in Computing Systems,* pp. 203-209, May 1992.

Charles. Wheatstone. "Contributions to the Physiology of Vision. — Part the First On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision", *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371-394, 1838.

Andrew P. Witkin. "Recovering Surface Shape and Orientation from Texture", *Artificial Intelligence*, vol. 17, pp. 17-45, 1981.

Eva Wong and Naomi Weisstein. "A New Visual Illusion: Flickering fields are localized in a depth plane behind nonflickering fields", *Perception*, vol. 14, no. 1, pp. 13-17, 1985.

Eva Wong and Naomi Weisstein. "Flicker Induces Depth: Spatial and temporal factors in the perceptual segregation of flickering and nonflickering regions in depth", *Perception & Psychophysics*, vol. 35, no. 3, pp. 229-236, 1984.

Albert Yonas. "Attached and Cast Shadows", in Perception and Pictorial Representation, Calvin F. Nodine and Dennis F. Fisher, eds., Praeger, 1979.

Albert Yonas, Michael Kuskowski and Susan Sternfels. "The Role of Frames of Reference in the Development of Responsiveness to Shading Information", *Child Development*, vol. 50, no. 2, pp. 495-500, June 1979.

Albert Yonas, Lynn T. Goldsmith and Janet L. Hallstrom. "Development of Sensitivity to Information Provided by Cast Shadows in Pictures", *Perception*, vol. 7, no. 3, pp. 333-341, 1978.

Mark J. Young, Michael S. Landy and Laurence T. Maloney. "A Perturbation Analysis of Depth Perception from Combinations of Texture and Motion Cues", *Vision Research*, vol. 33, no. 18, pp. 2685-2696, 1993.

Andrew Zisserman, Peter Giblin and Andrew Blake. "The Information Available to a Moving Observer from Specularities", *Image and Vision Computing*, vol. 7, no. 1, pp. 38-42, February 1989.

**References describing computational methods for extracting surface shape from patterns of texture:**

John (Yiannis) Aloimonos and Michael Swain. "Shape from Patterns: Regularization", *International Journal of Computer Vision*, vol. 2, pp. 171-187, 1988.

Andrew Blake and Constantinos Marinos. "Shape from Texture: Estimation, Isotropy and Moments", *Artificial Intelligence*, vol. 45, pp. 323-380, 1990.

Jonas Gårding. "Shape from Texture for Smooth Curved Surfaces in Perspective Projection", *Journal of Mathematical Imaging and Vision*, vol. 2, pp. 327-350, 1992.

Ken-ichi Kanatani and Tsai-Chia Chou. "Shape from Texture: General Principle", *Artificial Intelligence*, vol. 38, pp. 1-48, 1989.

Ko Sakai and Leif Finkel. "Characterization of the Spatial-Frequency Spectrum in the Perception of Shape from Texture", *Journal of the Optical Society of America, A*, vol. 12, no. 6, pp. 1208-1224, June 1995.