20

Shape from X: Psychophysics and Computation

Heinrich H. Bülthoff

The Many Routes to Shape

The human visual system derives a variety of information about the three-dimensional (3D) structure of the environment from different cues. This is illustrated in figure 20.1, where computer simulations of surface properties of a simple geometric form under different lighting conditions can lead to quite different 3D impressions. If an ellipsoid of rotation with Lambertian reflectance properties (like a table tennis ball) is simulated to be illuminated only by ambient light (equal amount of light from all directions), no inference of the 3D shape of the object can be made. The addition of a single point light source in the far distance (i.e., parallel illumination) allows our visual system to interpret the shading variations as a three-dimensional form; in other words, it computes shape from shading. The 3D impression of the ellipsoid becomes stronger when a highlight is added to the image by using a different shading model (Phong, 1975) for the computer graphic simulation. We get the strongest impression of the 3D shape of the object in the lower right of figure 20.1, where an additional source of information is available through simulation of surface texture. Note that not only the form of the object but also the perceived orientation of the object changes with the number of simulated depth cues. By observing figure 20.1 we can ask ourselves, what are the correct form and orientation of the object? Can we infer the correct 3D shape from 2D images? What are the best cues for shape? Which are better for orientation? We hope to answer some of these questions in the next few sections.

The outline of this chapter is as follows. First we motivate the need for cue integration in human and machine vision. In the next section we describe different representations of depth and how they can be assessed in psychophysical experiments. We discuss in detail two different techniques to measure shape-from-X, local and global shape probes, and how they are used to measure

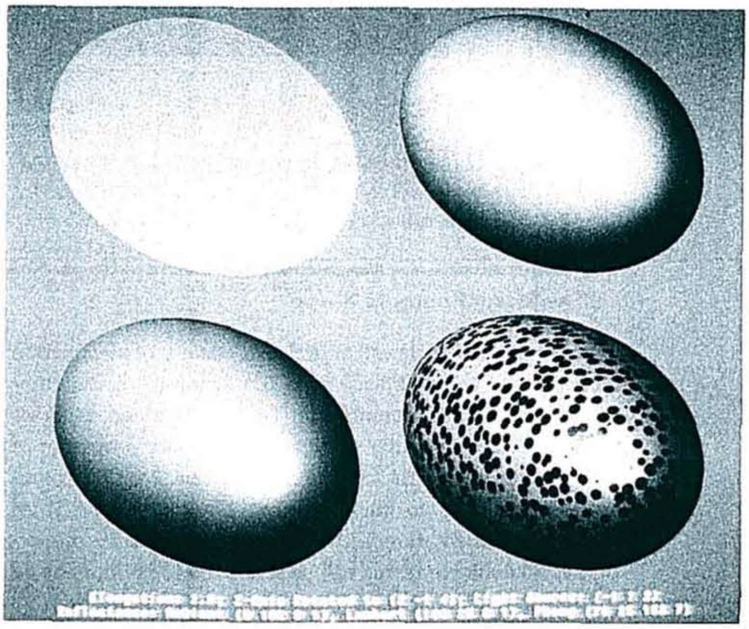


Fig. 20.1
Shape from X. The shape of the four objects looks quite different because the visual system derives different shape information from different shape cues. All four images were generated for the same 3D shape (ellipsoid of rotation) but with different simulated surface properties and under different lighting conditions.

shape from stereo, shading, and texture. In the following section we discuss an important and often neglected aspect of stereo vision: intensity-based stereo or shape from disparate shading. Bülthoff and Mallot (1988) showed that the human visual system can perceive depth in disparate images which have no discontinuities (zero crossings in the Laplacian of filtered images). This is a surprising finding because many theories in human and machine stereo vision are based on matching discontinuities in image intensities. An intensity-based stereo mechanism can be very useful for "direct" surface interpolation of surfaces with large smooth regions and should integrate with more robust measurements of edge-based stereo. In the section entitled Shape from Highlights we demonstrate an additional source of shape information that has been regarded previously as more of a nuisance than a useful cue to shape. Blake and Bülthoff (1990) showed that the human visual system can make use of the relative disparity of highlights in glossy images. For most machine vision algorithms these highlights are most undesirable because the disparity (or motion) of the highlight is different from the underlying structure and therefore can lead to false depth (or motion) measurements. The human visual system, on the other hand, can use this information in situations where only ambiguous information about surface shape is available, for example, in order to disambiguate the convex-concave ambiguity of shape-from-shading. This is a perfect example of the "disambiguation" type of cue integration. Other types of cue integration are discussed in Integration of Depth Modules. In the final section, a theoretical framework for cue integration is discussed briefly. A more detailed description of this framework can be found in Bülthoff and Yuille (1990).

The Need for Integration

The shape and depth cues simulated in figure 20.1 (and others) have been formalized in terms of computational theory and have been implemented as single modules in machine vision systems. Related studies from psychophysics and computational vision exist mainly for stereo (Julesz, 1971; Marr & Poggio, 1976, 1979; Mayhew & Frisby, 1981; Prazdny, 1985) and shading (Blake, Zisserman & Knowles, 1985; Ikeuchi & Horn, 1981; Mingolla & Todd, 1986; Pentland, 1984). There are also a number of studies on depth from texture (Aliomonos & Swain, 1985; Bajcsy & Lieberman, 1976; Cutting & Millard, 1984; Kender, 1979; Pentland, 1986; Witkin, 1981), line drawings (Barrow & Tenenbaum, 1981), surface contours (Stevens, 1981; Stevens & Brookes, 1987) and structure-from-motion (Koenderink, 1986; Longuet-Higgins & Prazdny, 1981; Landy, 1987; Ullman, 1979,1984), accommodation (Pentland, 1985), and occlusion (Haynes and Jain, 1987). Most implementations are quite successful for synthetic images but less reliable for natural images. On the contrary, the human visual system more easily extracts depth from the multiple 3D cues available in natural images compared to the isolated cues found in synthetic images (e.g., random dot stereograms). In order to study how the human visual system can integrate the information from multiple cues so successfully, we developed methods for quantitative measurement of perceived depth and shape with stimuli that are closer to natural images than those used in most psychophysical experiments. Using computer graphic techniques, we have precise control over the different shape and depth cues and we can use them in supportive or contradictory combinations to study the interaction between them and get a better idea how different cues are integrated into a stable representation of the 3D world. But before we discuss this, we will examine the question of what kinds of representations can be used by our visual system.

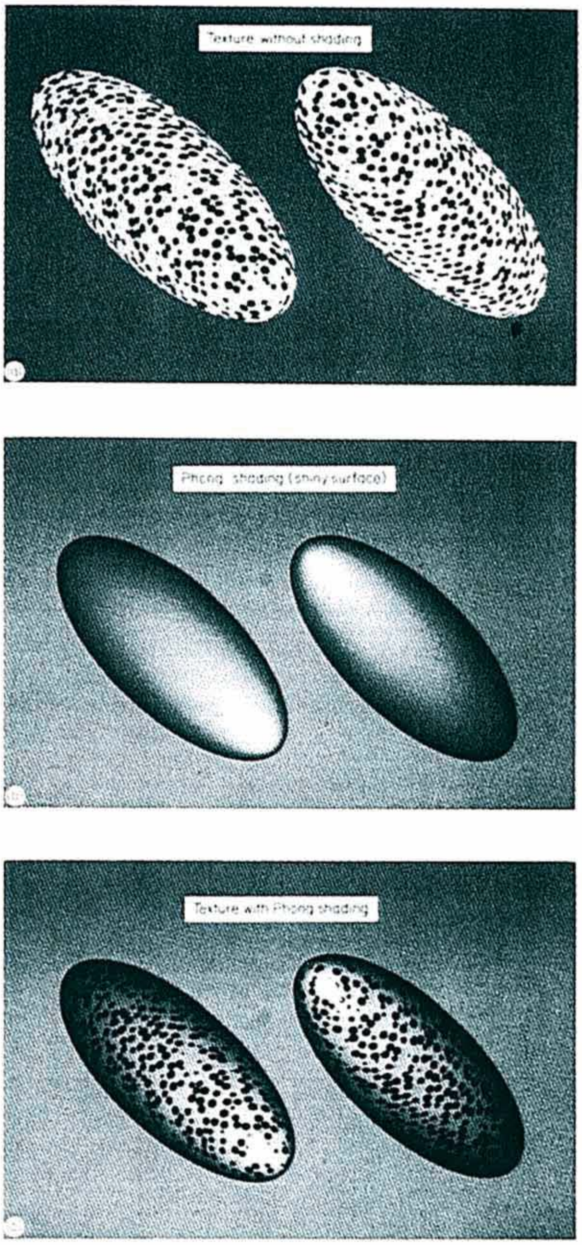


Fig. 20.2
3D descriptors. Different depth cues provide information about different 3D descriptors (e.g., range, shape, orientation). Try to estimate the angle between the long axes of the ellipsoids (for the correct answer, see text). (After Bülthoff & Mallot, 1990.)

How to Represent the Third Dimension?

Raw data, such as a range map from depth and shape cues, can be thought of as a trivial, or *zero-order representation* of the spatial structure of a scene. *Higher-order descriptors* can be derived from image data that make interesting spatial properties of the viewed scene explicit. The question of what constitutes a useful 3D descriptor can be answered in the light of the action that it should subserve. For example, a *pointwise depth map* can be useful for precise manipulation of objects while *surface curvature* (without exact range data) might be useful for the recognition of complex 3D shapes such as faces.

Which cues are relevant to one particular 3D descriptor? Occlusion contributes more readily to depth ordering than to surface curvature. Shading contributes more qualitatively to curvature than quantitatively to a depth map,

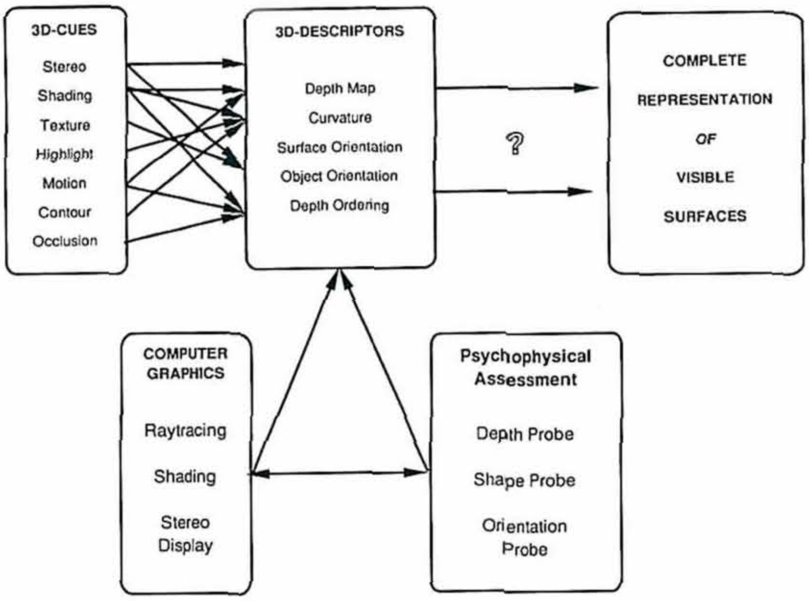


Fig. 20.3
Shape representation. 3D structure perceived from 2D image can be represented at different levels of abstraction. The depth cues themselves constitute multiple *zero-order* representations. Higher-order representations, i.e., *3D descriptors*, can be derived from interaction and integration of several of these zero-order representations. Different psychophysical experiments (much as computer vision tasks) involve various combinations of 3D descriptors. It is not clear whether a unique 3D representation exists that serves as a common data basis for all types of behavior dealing with the spatial structure of the environment. (Redrawn from Bülthoff & Mallot, 1990.)

and texture more to object orientation than to object form. This is illustrated in figure 20.2, where three pairs of ellipsoids are shown whose long axes of elongation are orthogonal to each other. The orthogonal orientation is best seen in part C, where texture and specular shading provide sufficient 3D information. If texture is used without shading (part A), the orientation of the objects can usually be perceived correctly, while the objects themselves appear flat. On the contrary, if shading is the only cue (part B), the objects appear nicely curved but it is difficult to see them orthogonal to each other.

How to Assess Properties of Multiple Representations?

Since the perception of three-dimensional scenes relies on so many different depth cues, which can lead to various descriptions of that scene in terms of distance, surface orientation, surface curvature and shape, we measured some of these 3D descriptors (depth map, curvature, and object orientation) for different depth cues (stereo, shading, highlight, texture, see figure 20.3).

The relation of shading (with and without highlights), stereo, and texture in the 3D perception of smooth and polyhedral surfaces was studied with computer graphic psychophysics (see appendix A). For polyhedral and textured objects, stereo disparities were associated with localized features, that is, the intensity changes at the facet or texel boundaries, while for the smooth surfaces only shading disparities occurred. For most of our experiments we used ellipsoids of revolution (viewed end-on) for the following reasons:

- As is shown later (Images Without Zero-Crossings), images of Lambertian shaded smooth ellipsoids with moderate eccentricities do not contain Laplacian zero-crossings when illuminated centrally with parallel light. This allows us to study intensity-based stereo mechanisms.
- The surfaces are closed and are naturally outlined by a planar occluding contour. This contour was placed in the zero disparity plane and did not allow the subjects to derive depth from binocular disparities.
- Convex objects such as ellipsoids do not cast shadows

or generate reflections on their own surface. Therefore, shading (attached shadows) could be studied without interference from cast shadows.

- End-on views of ellipsoids can be thought of as a model example of depth interpolation of a surface patch between sparse edge data.

Local and Global Depth Probes

Depending on the type of representation we wanted to measure quantitatively, we used two different types of probes (see appendix B). The depth probe can measure locally perceived depth, but has some disadvantages with depth cues which are better viewed monocularly (e.g., shading and texture). The global shape probe is more appropriate for the latter case but cannot be used to derive a precise depth map for all points in the image:

- *Local depth.* We mapped perceived depth with a small probe or cursor that was interactively adjusted to match the depth of the perceived surface (further details in appendix B). The depth of this probe was defined by edge-based stereo disparity and all other cue combinations were compared to the percept generated by edge-based stereo. All images were viewed binocularly with the depth cursor superimposed and hence had a zero disparity cue in them. Each adjustment of the probe gives a graded measurement of distance, or local depth, that is, this experiment corresponds to the 3D descriptor *depth map* in the scheme of figure 20.3. Note that the binocularly viewed local probe can interfere with monocular cues like shading and texture. Therefore, a more global shape probe was used to extend the range of possible shape measurements.

- *Global shape.* The global shapes of two objects with different combinations of depth cues were compared directly (further details in appendix B). Since all images showed end-on views of ellipsoids with different elongation, this measurement corresponds to *curvature* or *form* as a 3D descriptor.

- *Global orientation.* Object orientation can be measured in a matching task where long ellipsoids of different orientation are compared. While surface orientation is apparently hard to determine for human observers (Mingolla & Todd, 1986; Todd & Mingolla, 1983), the orientation of entire objects (e.g., orientation of *generalized cylinders*) can be measured easily in a matching task.

Shape From Stereo and Shading (Local Measurements)[1]

In the first series of experiments, 165 measurements were performed, each consisting of 45 adjustments of the depth probe to the perceived surface. Results were consistent in all three subjects and were pooled, since the differences were noticeable only in the standard deviation. The 16 plots of figure 20.4 show the averaged results of all subjects for the four types of experiments and four different elongations of Lambertian shaded ellipsoids.

The perceived elongation in the images with consistent cue combinations depends on the amount of information available. In figure 20.5 a measure of perceived elongation is derived from the depth map shown in figure 20.4 by a principle component analysis (see appendix C) and plotted as a function of displayed elongation. As can be seen from figure 20.5, the perceived elongation is almost correct when shading, intensity-based and edge-based disparity information are available (D^+E^+). This is not too surprising because this condition involves basically a disparity-to-disparity match (the probe is a disparity cued probe). This disparity match should work perfectly as long as the probe is not too distant from the grid intersections (edges) of the polygonal ellipsoid. In the case of smooth-shaded disparate images (D^+E^-), the edges are missing and depth perception is reduced. When shading is the only cue (D^-E^-), perceived elongation is much smaller and almost independent of the displayed elongation. Phong shading (highlights) instead of Lambertian shading did not change perceived depth significantly (dashed lines). A much stronger influence on the type of shading can be measured with the shape probe (see below).

In experiment D^-E^+, two identical images (zero disparity) of polyhedral ellipsoids (edges) were shown. Although shading alone provided some depth information as shown in experiment D^-E^-, the fact that edges occurred at zero disparity was decisive. The perceived depth did not vary with the elongation suggested by the shad-

[1] In collaboration with Hanspeter Mallot, Ruhr Universität Bochum.

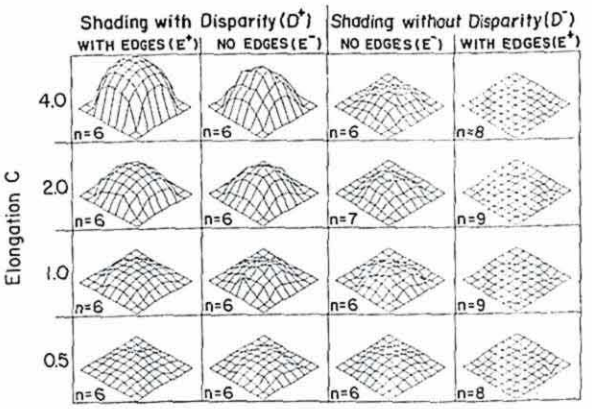


Fig. 20.4
Perceived surfaces. Each plot shows the average of six to nine sessions from three subjects. *Perceived depth decreases with the following sequence* of cue-combinations: disparity, edges, and shading ($D^+ E^+$); disparity and shading but no edges ($D^+ E^-$); shading only ($D^- E^-$); contradictory disparity and shading ($D^- E^+$). The elongation of the displayed objects *is denoted by c (depth not drawn to scale)*. (Redrawn from Bülthoff & Mallot, 1988.)

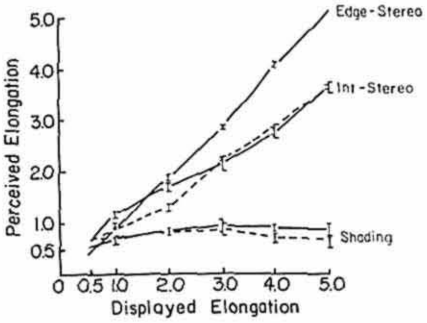


Fig. 20.5
Perceived elongation. Depth perception decreases with fewer cues available. The significant separation between the middle and lower curves (smooth shading with and without disparity) *illustrates the influence of disparity information* even in the absence of edges. Solid lines: Lambertian shading; dashed lines: Phong shading. (Redrawn from Bülthoff & Mallot, 1988.)

ing (and perspective) information and took slightly negative values which, however, were not significantly different from zero. This veto type of relationship between stereo and shading is probably due to the depth probe technique, which enforces disparity-to-disparity matching. A different type of integration between stereo and shading or texture (see, for example, Buckley, Frisby & Mayhew, 1989) might go unnoticed with this technique and therefore a more global shape probe was used in other experiments.

Depth can still be perceived when no disparate edges are present. This is not surprising, since shading information was still available. A comparison of the results (figure 20.5) for smooth-shaded images with and without disparity information, however, establishes a significant contribution of shading disparities (intensity-based stereo). The curves for D^+E^- and D^-E^- are significantly separated for all elongations except 0.5.

Shape from Shading and Texture (Global Measurements)

As discussed earlier, global shape cues like shape from shading and texture cannot be assessed with the local depth probe without interference with the shape-from-stereo module. Therefore we measured all cues, which are better viewed monocularly to eliminate zero disparity cues, with our global shape comparison technique (appendix B). All of our images with single monocular cues lead to large errors in perceived shape. With *shading* and *texture* curvature is underestimated (figure 20.6A, B), with a highlight it is overestimated (figure 20.6C). Note, that the reference ("given elongation") was displayed in stereo and that the elongation of the shaded or textured ellipsoid was chosen by the subject ("chosen elongation"). Underestimation of elongation corresponds therefore to chosen values above the dashed line and overestimation to values below the line.

Shape from Shading

One remarkable result of the comparison technique is that the shape-from-shading performance is much better with this technique than with the local depth probe technique. The adjusted shading scales with the displayed elongation of the stereoscopically displayed reference ellipsoid and does not level off as in the case of the depth probe measurements. There is still a strong underestimation of the elongation of the shaded ellipsoid for a given stereoscopically displayed reference ellipsoid, but in conjunction with a texture cue (figure 20.6D) the slope of the shape-from-texture-and-shading curve is close to 1.0 (veridical).

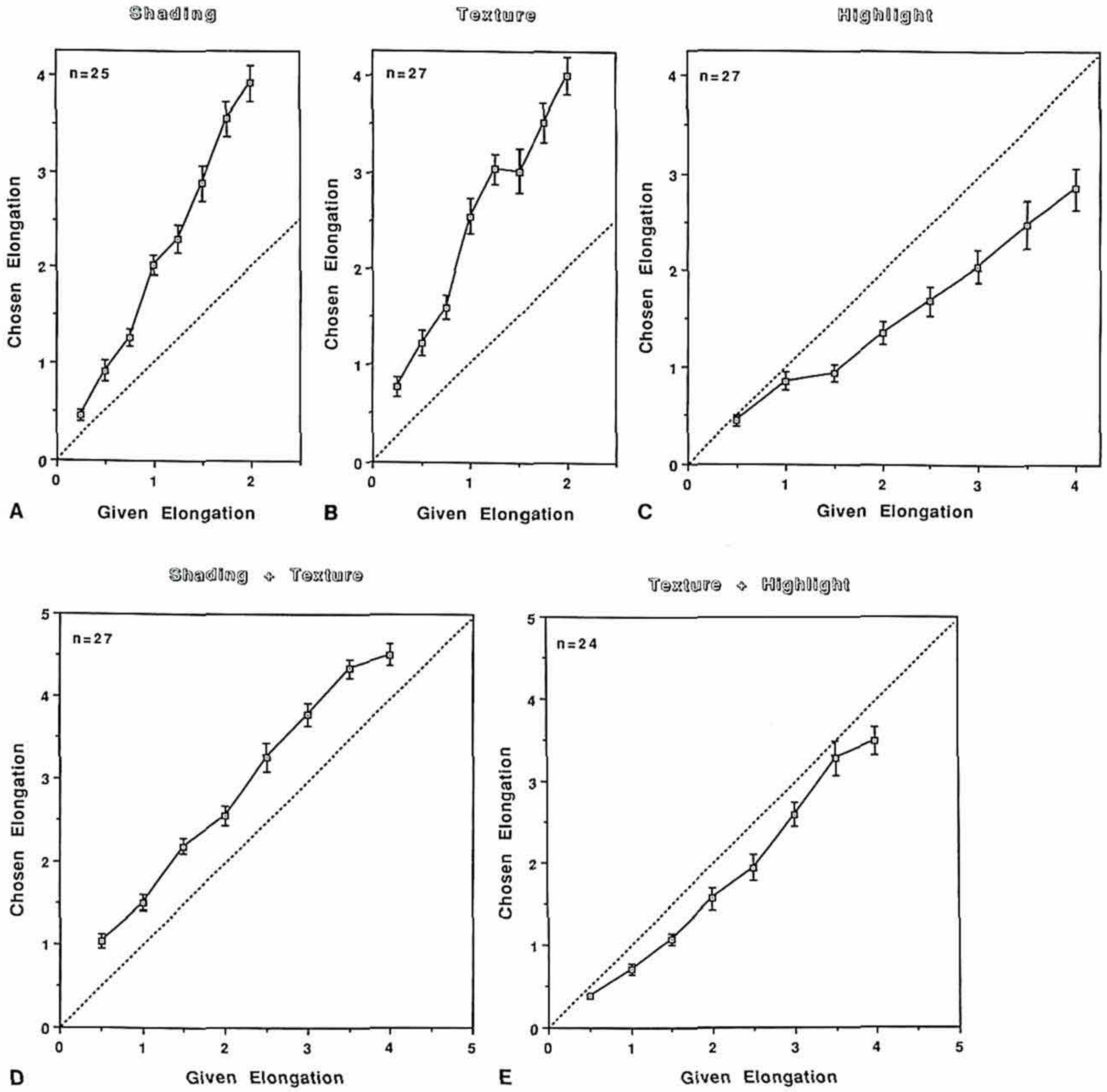


Fig. 20.6
Global shape. (A, B) Shape-from-shading and shape-from-texture lead to an underestimation of shape (slope > 1). (C) If a highlight is added to the shading (Phong shading model) the shape is overestimated in the adjustment task. (D) If shading and texture are presented simultaneously the shape is adjusted almost correctly (slope = 1) with a bias to adjust a larger elongation than necessary. (E) If a highlight is added the slope stays the same but the bias changes towards an overestimation of shape. (Redrawn from Bülthoff & Mallot, 1990.)

Shape from Texture

The performance of shape-from-texture and shape-from-shading is very similar. Both curves have almost the same offset and slope. This is not so surprising because the computational problem of shape-from-texture and shape-from-shading is very similar (Aliomonos & Swain, 1985). This similarity in the computational structure could be the reason for the strong cooperativity and almost veridical perception if both cues are present (figure 20.6D).

Shape from Highlights

A highlight on the shaded surface also seems to have a much larger influence with this technique and leads to an overestimation of curvature (figure 20.6C). This overestimation can be seen also if both texture and highlights are used (figure 20.6E). Again, in this case the cooperativity between modules shows up in the much more veridical perception of shape than with single modules. But compared to texture with shading (figure 20.6D), the texture and highlight curve (figure 20.6E) signals an overestimation of shape.

Shape from Disparate Shading (Intensity-Based Stereo)

As mentioned earlier, a very surprising finding is the strength of depth perception (70 percent) obtained from disparate shading under various illuminant conditions and reflectance functions. In computational theory, most studies have focused on edge-based stereo algorithms (for review, see Poggio & Poggio, 1984). This is due to the overall superiority of edge-based stereo, which is confirmed by the finding that edge-based stereo gives a better depth estimate than disparate shading (Blake et al., 1985). However, in the absence of edges and for surface interpolation, graylevel disparities appear to be more important than is usually appreciated.

Images Without Zero-crossings

One of the most important constraints in early vision for recovering surface properties is that the physical processes underlying image formation are typically smooth. The smoothness property is captured well by standard regularization (Poggio, Torre & Koch, 1985) and exploited in its algorithms. On the other hand, *changes of image intensity* often convey information about physical edges in the scene. The locations of sharp changes in image intensity very often correspond to depth discontinuities in the scene. Many stereo algorithms use dominant changes in image intensity as features to compute disparity between corresponding image points. In order to localize these sharp changes in image intensity, zero-crossings in Laplacian-filtered images are commonly used (Marr & Hildreth, 1980).

The disadvantage of these feature-based stereo algorithms is that only sparse depth data (at image features) can be computed. This forces an additional stage in which sophisticated algorithms (Blake & Zisserman, 1987; Grimson, 1982) interpolate the surface between data points. In order to test for the ability of human stereo vision to get denser depth data by using features other than edges, or even a completely featureless mechanism, we computed images without sharp changes in image intensity. We show that for an orthographically projected image of a sphere with Lambertian reflection function and parallel illumination, zero-crossings in the Laplacian are missing. Consider a hemisphere given in cylindrical coordinates by the parametric equation

$$z = \sqrt{1 - r^2}. \tag{1}$$

In the special case of a sphere, the surface normal simply equals the radius, that is,

$$\mathbf{n} = (r\cos\varphi, r\sin\varphi, \sqrt{1 - r^2}). \tag{2}$$

For the illuminant direction $\mathbf{l} = (0, 0, 1)$ and the Lambertian reflectance function, we obtain the luminance profile

$$I(r) = I_0(\mathbf{l}\cdot\mathbf{n}) = I_0\sqrt{1 - r^2}, \tag{3}$$

where I_0 is a suitable constant, i.e., the image luminance is again a hemisphere. For the Laplacian of I, we obtain

$$\nabla^2 I(r) = I''(r) - \frac{1}{r}I'(r) = -I_0\frac{r^2}{(1 - r^2)^{3/2}}. \tag{4}$$

This is a nonpositive function of r, with $\nabla^2 I(0) = 0$; i.e., the Laplacian of I has no zero-crossings.

Unfortunately, this result does not hold for ellipsoids with $c \neq 1$. A similar computation for an ellipsoid with elongation c yields

$$I_c(r) = I_0\frac{\sqrt{1 - r^2}}{\sqrt{1 - (1 - c^2)r^2}}, \tag{5}$$

which reduces to equation 3 for $c = 1$. The luminance-profiles for elongations $c \geqslant 2$ are no longer convex. That is to say that the second derivatives of these profiles in fact have zero-crossings, and a similar result holds for the Laplacians. However, when filtering with the Laplacian of a Gaussian or with the difference of two Gaussians (DOG) is considered, it turns out that these zero-crossings are insignificant for the elongations used here. Pixel-based convolutions failed to show the "edges" unequivocally, and even a Gaussian integration algorithm run on the complete function rather than on the sampled array produced no zero-crossings beyond the single-precision truncation error. We therefore conclude that the slight zero-crossings in the unfiltered Laplacian of our luminance profiles do not correspond to significant edges. For oblique illumination we found numerically that the self shadow boundary corresponds to a level rather than a zero-crossing in the DOG-filtered image.

Independently of our own work, images of ellipsoids may be useful in the study of the psychophysical relevance of Laplacian zero-crossings.

Local or Global Mechanisms?

Are there features other than zero-crossings that can account for the shape-from-disparate-shading performance found in our experiments? Possible candidates include the intensity peak as proposed by Mayhew and Frisby (1981) and level-crossings in the DOG-filtered image which, according to Hildreth (1983), might account for Mayhew and Frisby's data as well.

In order to distinguish between a localized (feature-based) and a distributed mechanism for shape-from-disparate-shading we tested the effect of a small disparate token displayed in front of a nondisparate background with the depth probe (figure 20.7). Our data show that for large elongations, a single stereo feature (ring) is not sufficient to produce the same percept as full disparate shading (compare part A of figure 20.7 with parts B to D). For small elongations (0.5 to 2.0; not shown in figure 20.7) the differences were not pronounced. We therefore conjecture that disparate shading does not rely on feature matching and thus can be used for surface interpolation when edges are sparse. This view is well in line with the finding that edge information, whenever present, overrides shape-from-disparate-shading (figure 20.8).

Note, however, that we do not propose the naive idea of pointwise intensity matching as a mechanism for shape-from-disparate-shading because of its sensitivity to noise in both the data and in neural processing. Even in the absence of image noise, intensity-based algorithms (e.g., Gennert, 1987) can lead to severe matching errors when run on our stimuli (see Psychophysical Support for the Bayesian Framework). A window-based correlation mechanism like the one used for optical flow computation (Bülthoff, Little & Poggio, 1989) might be more appropriate for shape-from-disparate-shading. This type of algorithm has been successfully used for stereo (D. Weinshall, personal communication). For a comparison see also the SWITCHER algorithm described in chapter 21. In the next section we will look at one additional cue (highlights) that is used by the visual system in cases where shape-from-shading or shape-from-texture does not provide unambiguous shape information.

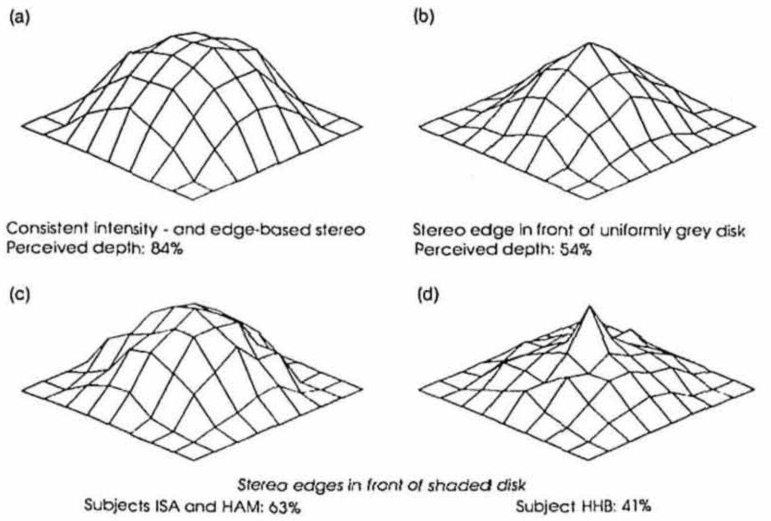
Shape from Highlights[2]

Many images of artificial and natural scenes contain "highlights" generated by mirrorlike reflections from glossy surfaces. Computational models of visual processes have tended to regard these highlights as *obscuring* underlying scene structure. Mathematical modeling shows that, on the contrary, highlights are rich in local geometric information. This section will demonstrate that the brain can apply that information. Stereoscopically viewed highlights or "specularities" can serve as cues for 3D local surface geometry. The human visual system seems to employ a physical model of the interaction of light with curved surfaces—a model firmly based on ray optics and differential geometry. We develop such a model in the next section.

The Computational Model

The basic principle of the "specular stereo" model is quite simple (figure 20.9). According to ray optics, the image of a light source—a specularity—appears behind a glossy, convex surface and (generally) in front of a concave one,

[2] In collaboration with Andrew Blake, Oxford University.

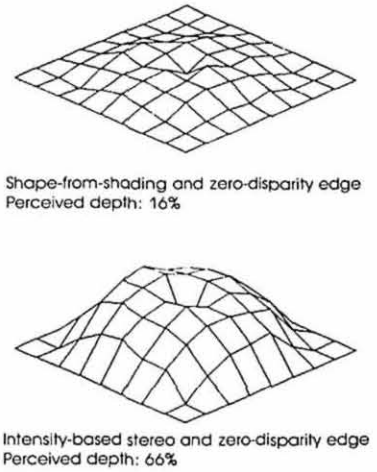


Consistent intensity - and edge-based stereo
Perceived depth: 84%

Stereo edge in front of uniformly grey disk
Perceived depth: 54%

Stereo edges in front of shaded disk
Subjects ISA and HAM: 63%
Subject HHB: 41%

Fig. 20.7
Surface interpolation. (a) Shape-from-disparate-shading plus disparate edge information leads to an almost correct percept ($n = 6$). (b) A *single edge token in front of a uniformly gray disk yields a cone-like* subjective surface ($n = 6$). (c, d) Shape-from-shading plus disparate edge information leads to an ambiguous perception ($n = 3 + 3$). Some subjects fused the edge-token and the surround into one coherent surface (c) while others saw the edge-token floating in front of a rather flat surface (d). Only data for elongation 4.0 are shown. (Redrawn from Bülthoff & Mallot, 1988.)



Shape-from-shading and zero-disparity edge
Perceived depth: 16%

Intensity-based stereo and zero-disparity edge
Perceived depth: 66%

Fig. 20.8
Veto. (A) A zero-disparity edge token vetoes shape-from-shading ($n = 7$) and (B) shape-from-disparate-shading ($n = 6$). Only data for elongation 4.0 are shown. (Redrawn from Bülthoff & Mallot, 1988.)

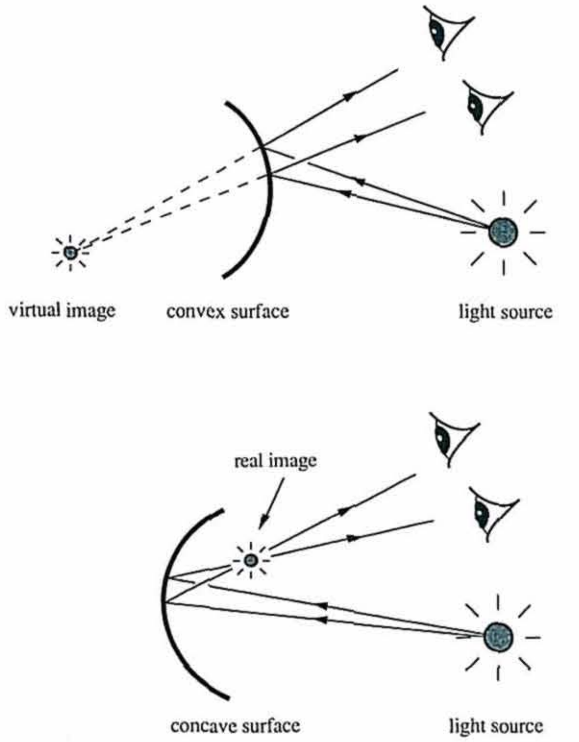


Fig. 20.9
Specular stereo—the basic principle. Specularities appear behind a convex mirror but in front of a concave one. (Redrawn from Blake & Bülthoff, 1990.)

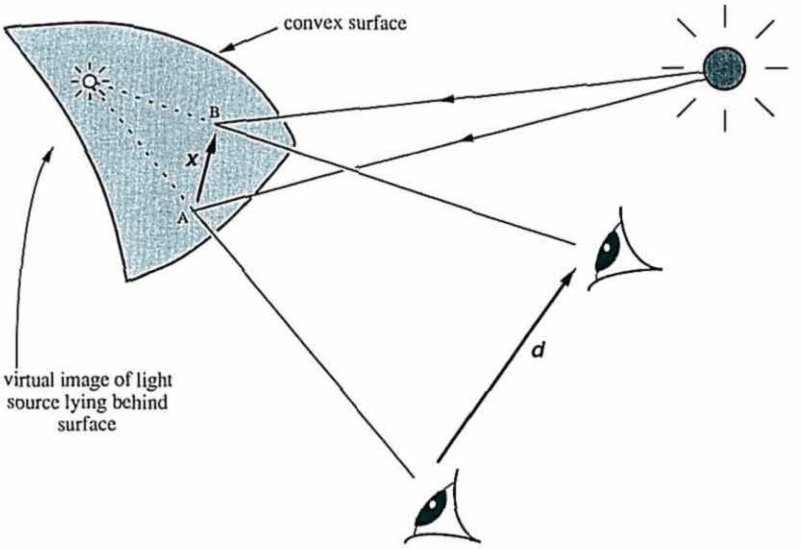


Fig. 20.10
Geometry of specular stereo. Ray optics establishes direct relationship between surface shape and measured disparity. Stereoscopic relative disparity **δ** is a projection of the displacement vector **x** of the specularity on the reflecting surface. Displacement **x**, in turn, is related linearly to baseline vector **d**, the coefficients of the relation *being a function solely of surface geometry.* If the visual system knows the physics of specular reflection and the light source position, then the relative disparity of a specularity would be consistent only with certain values of local surface curvature.

provided both viewer and source are sufficiently distant from the surface. But this simple idea must be expanded. How, for example, does a specularity appear in a surface that is hyperbolic? Whether it appears behind or in front depends on the orientation of the surface. Even on an elliptic surface, the apparent depth of the image varies if the surface is rotated about the line of sight.

In fact we are forced to hesitate at the notion of apparent "depth." What we actually observe is determined by *horizontal and vertical relative disparities.* Stereoscopic disparity is a vector quantity, conventionally taken (Mayhew & Longuet-Higgins, 1982) to have a horizontal and a vertical component equal to the differences in x, y coordinates of corresponding image points in left and right planar projections. Horizontal disparity is the component of the disparity vector parallel to the stereoscopic baseline (**d** in figure 20.10) and vertical disparity is the orthogonal component. *Relative* disparity of a specularity is (roughly) the difference between its disparity and that of a nearby point on the surface. Surface features (scratches, for example) obey the "epipolar" constraint (Arnold & Binford, 1980; Mayhew & Frisby, 1981). Once the epipolar lines are known—a nontrivial problem of camera calibration in computer vision (see chapter 21)—vertical *disparity of one surface point relative to another is zero.* Specularities, however, are not surface features (that is, they are not stuck to a surface) so they do not obey the epipolar constraint. They frequently have nonzero vertical relative disparities. Both horizontal and vertical relative disparities of a specularity vary as the surface is rotated about the line of sight. Now, the actual depth of a *surface* feature is approximately proportional to horizontal disparity, but perceived depth could be affected by the introduction of vertical disparity (Koenderink & van Doorn, 1976). Only in cases where vertical disparity is negligible (e.g., on a spherical surface with slant less than about 30°) can we confidently talk about the depth of a specularity.

Ray optics establishes a direct relationship between surface shape and measured disparity (Blake, 1985; Blake & Brelstaff, 1988; Zisserman, Giblin & Blake, 1989). To a good approximation, the relative disparity vector **δ** depends linearly on the stereo baseline **d**, and the coefficients *of the linear relation are solely a function of surface* geometry (figure 20.10). Suppose this model were fully utilized by the visual system, and light source position were known, then the relative disparity of a specularity would be consistent only with certain values of local surface curvature. Even if nothing is known about source position, relative disparity still constrains curvature: No convex surface can generate a negative horizontal relative disparity; a concave surface hardly ever generates a positive one. The experiments described in this section aim to test whether the human visual system exploits such constraints.

The idea that human vision employs physical constraints is, of course, not new—it has been argued vigorously by Marr (1982) and is exemplified by surface continuity and epipolar constraints in theories of stereo vision (Julesz, 1971; Marr & Poggio, 1979; Mayhew & Frisby, 1981). Continuity constraints also underly certain theories of motion perception (Bülthoff et al., 1989; Hildreth, 1984; Yuille & Grzywacz, 1988) that also have some psychophysical support. While continuity is a mathematically precise notion, its application to the physical world is intrinsically imprecise—it is scale dependent. However, the epipolar constraint *is* precise, and expressible in terms of the equations of projective geometry. But it is "internal"—a consequence of the physics of the eye itself rather than of the external world. In the case of the analysis of specularities, however, it seems that the visual system may have summarized an algebraic theory that describes the physics of surfaces in the world. The theory is both "external" and precise. The next two sections describe two experiments aimed to test whether the human visual system exploits such constraints.

Surface Quality from Highlights

An adjustment task was devised in which the subject interactively changes both horizontal and vertical disparities of a highlight. Images of glossy, textured, curved surfaces are generated with a computer graphics workstation (Symbolics, Inc.) and displayed on a high-resolution color monitor with a stereo viewing system (see appendix A). The texture is of sufficient density to furnish strong cues for curvature from edge-based stereo. Simulation of surface gloss causes a specularity to appear superimposed on the texture, as in figure 20.11A. As discussed earlier, edge-based stereo cues can override cues such as monocular or disparate shading. We might therefore expect also that specularity cues should be overridden; that is precisely what happens. When the specular relative disparities are veridical the whole surface appears glossy as in figure 20.11A and not just in the vicinity of the specularity (Beck, 1972). However, when horizontal relative disparity is nonveridical the surface ceases to look glossy. For example, if the specularity is in front of the surface with large convergent (−) relative disparity, surface quality is reported to be matte and opaque, with a puff of cloud in front of the surface (figure 20.11B). The cloud patch is not perceived as a specularity and therefore there is no reason for the surface to look glossy. For excessively divergent (+) relative disparity, subjects usually report that the surface looks transparent, with a source of light behind it (like a frosted glass light bulb). Again, an incorrect position (relative disparity) of the specularity discounts the bright patch as a specularity and the visual system finds a different interpretation for the way in which the patch was generated. The interpretation of surface property changes from opaque to transparent. When the relative disparity is zero the simulated specularity looks like a powdery patch on the surface and the surface does not look glossy. Note, however, that in nonstereo images (like any photograph) surfaces can look glossy even with zero relative disparity. In this case a cue conflict does not really exist because all surfaces are flat and relative disparity does not have any meaning in these images.

In an informal two-alternative, forced-choice (2AFC) experiment, 11 out of 12 naive observers who were asked which of two presented surfaces was the "polished" surface chose the surface shown in figure 20.11A, in agreement with the prediction of the model.

In an adjustment task naive subjects were asked to achieve the most realistic looking glossy surface. They repeatedly pressed buttons which (unknown to the naive subjects) caused the relative disparity of a specularity to vary. They were simply told that pressing the two buttons would make the surface appear more or less shiny. Either vertical disparity was held constant (at the value determined by the ray-optic model) while horizontal disparity was varied or vice versa. Steps in specular disparity for each button press were sufficiently small (2 pixels or about 1.5 min arc) that most of the subjects did not perceive the specularity to be moving in depth. Four test surfaces were used in the adjustment task—a convex sphere, two convex ellipsoids and one concave ellipsoid.

Results for the convex sphere (figure 20.11C) show that, on average, subjects' adjusted values were not significantly different from veridical for horizontal

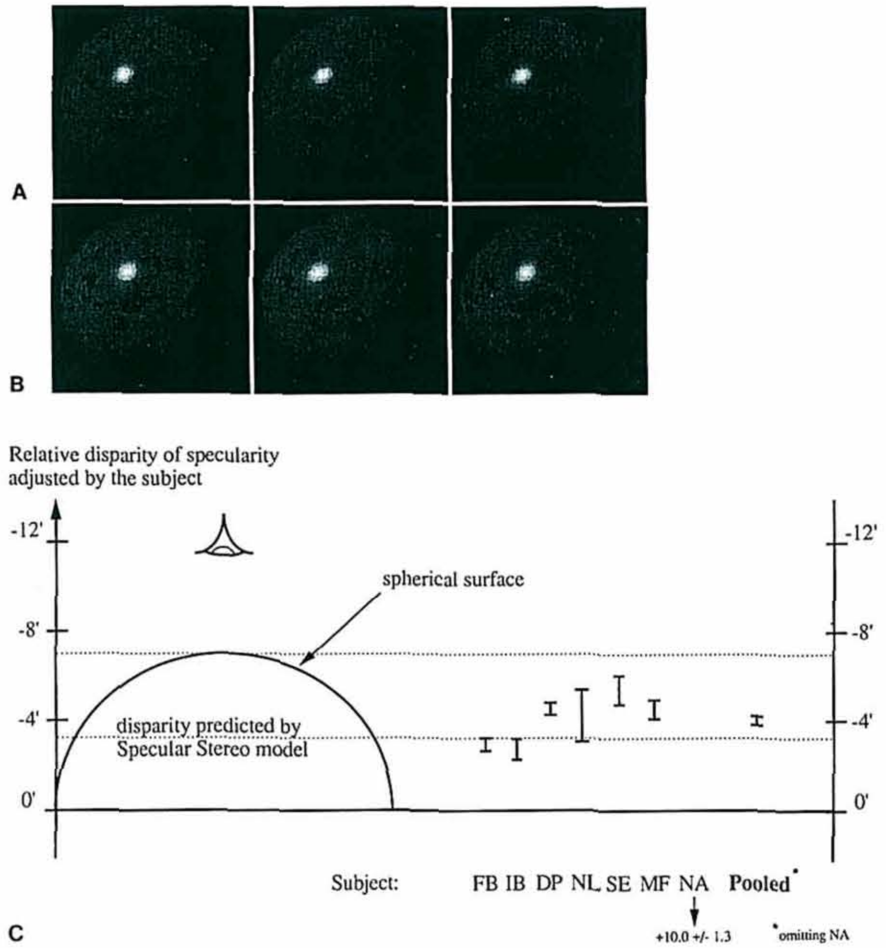


Fig. 20.11

Surface property. The perception of surface properties can change by moving a specular highlight relative to the surface. The surface of the sphere (A) (stereo view) looks metallic because the highlight is in the correct position behind the surface. If the highlight is in front of the surface (B) the surface looks more dull and not metallic (mirrorlike) at all. The human visual system seems to exploit the laws of reflection in the 3D interpretation of 2D images. In the psychophysical adjustment task most subjects put the specular highlight slightly (not significantly) displaced from the correct position for the sphere (C).

($P < 0.001$, F = 2). Note that the *sign* of the horizontal relative disparity after adjustment is always correct. This corresponds to robust discrimination between convex and concave surfaces as mentioned earlier. It is difficult to get significant vertical disparity effects for this surface because the veridical vertical disparity is close to zero (0.5 min arc). Four naive subjects adjusted the circumpolar disparity close to zero, but it is conceivable that there is some regression toward zero. Therefore, we tested a situation in which the correct vertical disparity of a specularity was quite different from zero. This is the case for the oblique-oriented ellipsoid. Five naive subjects and the two authors made adjustments whose signs were as predicted by the model. The visual system apparently has some dedicated competence for analysis of specularities and apparently "knows" enough about the physics of specularity to predict the sign of the vertical disparity correctly. Similar results are obtained for the two convex ellipsoids; the average adjusted disparities are close to veridical. Poorer agreement is obtained in the case of the concave ellipsoid, and the sign of the relative horizontal disparity after adjustment is inconsistent. Subjects reported that, for this surface, the adjustment task was relatively difficult to perform.

The conclusion of this experiment is that the human visual system models the physics of specular reflection well enough to predict relative disparity effects. Agreement with predictions is good qualitatively (sign is preserved), and there even is a degree of quantitative agreement. In particular, in the case of a convex sphere for which we can associate horizontal disparity with depth, the visual system "expects"—correctly—that a specularity lies behind not on the surface.

Surface Curvature from Highlights

The second experiment is complementary to the first. Can the visual system accommodate to variations in specular relative disparities by changing its hypothesis about surface curvature, rather than its hypothesis of glossiness?

We devised the stimulus of figure 20.12A—a stereo, textured variant of an ambiguous (reversible) shaded surface. The texture elements all have zero disparity, consistent with a frontoparallel surface. Nonetheless, monocular shading/texture cues are not entirely overridden, so that subjects can usually see both convex and concave (like a dog bowl) interpretations. A superimposed specularity (figure 20.12B), with either convergent or divergent relative disparity (±5′) biases the interpretation. As the physics predicts, convergent relative disparity biases subjects' interpretation toward concave and divergent toward convex (figure 20.12C). The effect develops gradually with repeated exposures. Naive subjects made a forced choice (2AFC) between a convex or a concave interpretation. Time sequences (figure 20.12C) show that while initially subjects may be locked into one or the other interpretation, after around 20 exposures they reliably pick the interpretation that is consistent with the sign of horizontal relative disparity. Note that the change in position of the specularity is contrary to that of the surface—when the specularity is furthest away (divergent horizontal disparity) the center of the surface is nearest to the viewer (convex) and vice versa. Any explanation in terms of a pulling effect exerted by the specularity on the surface is thereby excluded.

How Important are Specularities?

It could be argued that specularity is a marginal visual phenomenon since specularities are relatively sparse in images compared with texture edges and other features. Moreover, it is associated more with artifacts, relatively recent on an evolutionary timescale, than with "natural" objects. Is it really likely, as we claim, that we have developed mechanisms to analyze specularities? In reply, it is worth noting first that specularities do commonly occur on (hairless) faces and that facial recognition is, presumably, important for survival. More significant though, it is not necessary to claim that the ability to deal with specular motion and stereo developed via evolution. The processing of specularities, therefore, could simply be an extended usage of the parallax mechanism, *learned* in a modern environment filled with specular artifacts.

Cognitive vs. Early Vision

Naive observers, asked where a specularity appears to be in relation to the surface that generated it, usually reply that it appears to lie *on* the surface. What we tried to show with the first experiment is that the early visual system "knows" better, choosing configurations that are broadly consistent with the physics of specular reflection. The second experiment demonstrates that the early visual system can use the information about the 3D position of the specularity to make some inference about the curvature of the underlying surface. One reason that the more

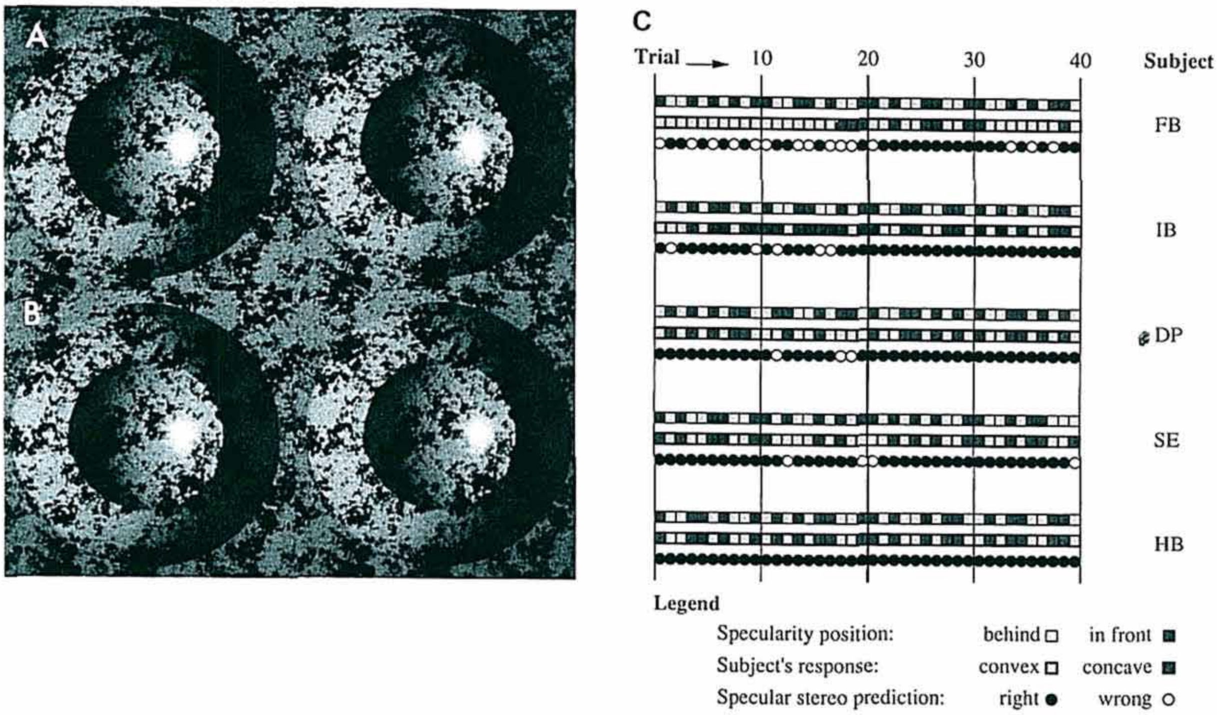


Fig. 20.12
Surface curvature. The perception of surface curvature can change with the position of a specular highlight. In order to demonstrate that the human visual system knows the physics of light reflection we used an image of a surface whose three-dimensional interpretation can flip easily between two states (convex/concave). If a highlight is added to the image the 3D interpretation of the inner part of the surface is biased more towards convex. A stereo pair was made with zero disparity for the textured surface, and then a specularity superimposed either in front of (A) or behind (B) the textured surface (uncrossed view), flipping randomly between the two, with 5 or 10 sec exposures separated by a random-dot masking frame. Subjects made a two-alternative forced choice (2AFC) between convex and concave. After a short training period (20 exposures without feedback) they made more choices that conform to the predictions of the model (C). A control experiment with a white disk of about the size of the specularity that did not look like a highlight at all, did not show any consistent effect between subjects on the perceived curvature. It might be difficult to experience the curvature effect if the images are not displayed on a CRT monitor because of the limited contrast range in the print. In order to get the best effect it is essential that the highlight look like a real reflection of the light source. (Redrawn from Blake & Bülthoff, 1990.)

cognitive level ignores this position information might be that it is better to ignore the virtual images of light sources around us; otherwise, we would perceive them as obstacles and we would be very busy trying to avoid all those specularities around us.

Integration of Depth Modules

Before we get to the final section on a computational model of cue integration, we summarize the interactions of different depth cues (as derived from our depth probe experiments) in figure 20.13. In some experiments we presented conflicting information from stereo and shading cues. Whenever visible, edge-based disparities were decisive for the perceived depth (see figures 20.4, D^-E^+, 20.7, and 20.8). Edge-based stereo thus overrides both shape-from-shading and shape-from-disparate-shading in our experiments. It is possible, however, that this veto relationship occurs only in the locally derived depth map (disparity matching) because the global percept of the polyhedral ellipsoid in experiment (D^-E^+) is not flat, but rather convex. Stevens and Brookes (1988) also reported that with conflicting monocular and stereo information

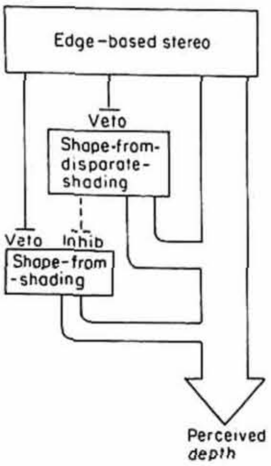


Fig. 20.13
Integration of depth cues. The size of the boxes and interaction channels reflects the contribution of the different depth cues to the overall perceived depth (accumulation). In contradictory cases, shape from both disparate and nondisparate shading is vetoed by edge-based stereo. An inhibitory influence of shape-from-disparate-shading on shape-from-shading is discussed in the text. (Redrawn from Bülthoff & Mallot, 1988.)

the 3D percept was dominated by the monocular information and not by stereo. Their task involved comparing the relative depth of two points on a planar surface that had contradictory monocular and stereo information and, in addition, surface orientation had to be estimated —which is a difficult task (see Todd & Mingolla, 1983). A conflicting cue combination of shape-from-shading and shape-from-disparate-shading was presented in the experiment with smooth-shaded nondisparate images (D^-E^-). In this case, shape-from-shading is not vetoed by the lack of shading disparities but leads to a reduced depth perception of about 25 percent. An inhibitory interaction between the two cues may account for this poor shape-from-shading performance and the ceiling effect in figure 20.5.

Another summary of our data that includes both depth probe and shape comparison techniques is shown in figure 20.14. This representation is based on the idea (sketched in figure 20.3) that the integration of different *3D cues* can lead to the perception of different *3D descriptors* (range, shape, orientation). The contribution of single monocular cues is different for the 3D descriptors. Object orientation is best recovered from texture cues (Bülthoff & Mallot, 1988) while surface curvature (shape) can be inferred more easily from shading. With binocular shading (Lambertian or Phong shading) range perception is rather strong (70 percent). It is even stronger for the perception of shape (100 percent). The addition of a highlight to a shaded surface has no effect in the range-matching task, while a strong effect was found in the shape comparison task. Highlights always led to an overestimation of shape, while dull surfaces (Lambertian shading) were judged too flat.

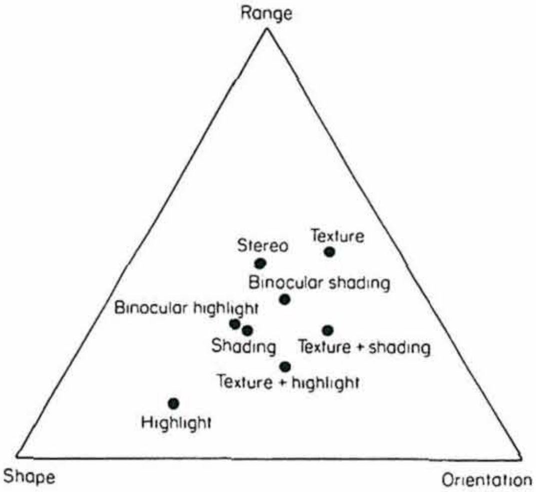


Fig. 20.14
Depth triangle. This representation of our depth probe and shape comparison data shows the relative importance of depth cues (stereo, shading, texture) for different 3D descriptors (range, shape, orientation); see also figure 20.3. Shading has a stronger influence on the perceived shape, while texture seems to be more important for orientation (compare with figure 20.2). Stereo is of equal importance for all 3D descriptors because the shape, orientation and distance to an object (range) can be easily derived from a complete depth map. (Redrawn from Bülthoff & Mallot, 1990.)

A Bayesian Framework for Cue Integration[3]

In this section a theoretical formulation for cue integration is introduced. This formulation is based on the Baye-

[3] In collaboration with Alan Yuille, Harvard University, and Davi Geiger, MIT.

sian approach to vision, in particular in terms of coupled Markov random fields. This formalism is rich enough to contain most of the elements used in standard stereo theories with the additional advantage that it allows integration of cues from different matching primitives. These primitives can be weighted according to their robustness. For example, depth estimates obtained by matching intensity are unreliable, since small fluctuations in intensity (due to illumination or detector noise) might lead to large fluctuations in depth, hence they are less reliable than estimates from matching edges. The formalism can also be extended to incorporate information from other depth modules (e.g., shading and texture) and provides a model for sensor fusion (Clark & Yuille, 1990).

Unlike previous theories of stereo that first solved the correspondence problem and then constructed a surface by interpolation (Grimson, 1982), this framework proposes combining the two stages. The correspondence problem is solved to give the disparity field which best satisfies the a priori constraints.

The model involves the interaction of several processes and is introduced here in three stages at different levels of complexity.

At the first level, features (such as edges) are matched using a binary matching field V_{ia} determining which features correspond. In addition, smoothness is imposed on the disparity field $d(\mathbf{x})$, which represents the depth of the surface from the fixation plane. In this case the correspondence problem, determining the V_{ia}, is solved to give the smoothest possible disparity field. The theory is related to work by Yuille and Grzywacz (1988) on motion measurement and correspondence and, in particular, to work on long-range motion. It can be shown that the cooperative stereo algorithms of Dev (1975) and of Marr and Poggio (1976) are closely related to this theory (Bülthoff & Yuille, 1990; Yuille, Geiger & Bülthoff, 1989).

At the second level, line process fields $l(\mathbf{x})$ (which represents depth discontinuities) (Geman & Geman, 1984) are added to break the surfaces where the disparity gradient becomes too high. For a different approach making use of the disparity gradient constraint, see chapter 21.

The third level introduces additional terms corresponding to matching image intensities. Such terms are used in the theories of Gennert (1987) and Barnard (1986) which, however, do not have line process fields or matching fields. A psychophysical justification for intensity matching is given in the section Shape from Disparate Shading. Thus the full theory is expressed in terms of energy functions relating the disparity field $d(\mathbf{x})$, the matching field V_{ia}, and the line process field $l(\mathbf{x})$.

By using standard techniques from statistical physics, particularly the mean field approximation, one can eliminate certain fields and obtain effective energies for the remaining fields (see Geiger & Girosi, 1989; Geiger & Yuille, 1989). As discussed in Yuille (1989), this can be interpreted as computing marginal probability distributions and allows us to show that several existing stereo theories are closely related to versions of the proposed framework. The three levels of this framework are presented in more detail in appendix D.

The Bayesian Formulation

Given an energy function model one can define a corresponding statistical theory. If the energy $E(d, V, C)$ depends on three fields: d (the disparity field), V the matching field, and C (the discontinuities), then (using the Gibbs distribution; see Parisi, 1988) the probability of a particular state of the system is defined by

$$P(d, V, C \mid g) = \frac{e^{-\beta E(d, V, C)}}{Z}, \tag{6}$$

where g is the data, β is the inverse of the temperature parameter, and Z is the partition function (a normalization constant).

Using the Gibbs distribution one can interpret the results in terms of Bayes' formula

$$P(d, V, C \mid g) = \frac{P(g \mid d, V, C) P(d, V, C)}{P(g)}, \tag{7}$$

where $P(g \mid d, V, C)$ is the probability of the data g given a scene d, V, C; $P(d, V, C)$ is the a priori probability of the scene; and $P(g)$ is the a priori probability of the data. Note that $P(g)$ appears in the above formula as a normalization constant, so its value can be determined if $P(g \mid d, V, C)$ and $P(d, V, C)$ are assumed known.

This implies that every state of the system has a finite probability of occurring. The more likely ones are those with low energy. This statistical approach is attractive because the β parameter gives us a measure of the uncertainty of the model (some refer to the temperature parameter $T = 1/\beta$). At zero temperature ($\beta \rightarrow \infty$) there is no uncertainty. In this case the only state of the system that has nonzero probability, hence probability 1, is the state that globally minimizes $E(d, V, C)$. In some nongeneric situations there could be more than one global minimum of $E(d, V, C)$.

Minimizing the energy function will correspond to finding the most probable state, independent of the value of β. The mean field solution,

$$\bar{d} = \sum_{d,V,C} dP(d, V, C|g), \tag{8}$$

is more general and reduces to the most probable solution as $T \rightarrow 0$. It corresponds to defining the solution to be the mean fields, the averages of the f and l fields over the probability distribution. This allows one to obtain different solutions depending on the uncertainty. A justification for using the mean field as a measure of the fields resides in the fact that it represents the minimum variance Bayes estimator (Gelb, 1974). More precisely, the variance of the field d is given by

$$Var(d:\bar{d}) = \sum_{d,V,C} (d - \bar{d})^2 P(d, V, C|g), \tag{9}$$

where $\bar{d}$ is the center of the variance and the $\sum_{d,V,C}$ represents the sum over all possible configurations of d, V, C. Minimizing $Var(d:\bar{d})$ with respect to all possible values of $\bar{d}$ we obtain

$$\frac{\partial}{\partial \bar{d}} Var(d:\bar{d}) = 0 \rightarrow \bar{d} = \sum_{d,V,C} dP(d, V, C). \tag{10}$$

This implies that the minimum variance estimator is given by the mean field value.

Statistical Mechanics and Mean Field Theory

One can estimate the most probable states of the probability distribution (equation 7) by, for example, using Monte Carlo techniques (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953) and the simulated annealing (Kirkpatrick, Gelatt & Vecchi, 1983) approach. The drawback of these methods is the amount of computer time needed for the implementation.

There are, however, a number of other techniques from statistical physics that can be applied. They have recently been used to show (Geiger & Girosi, 1989; Geiger & Yuille, 1989) that a number of seemingly different approaches to image segmentation are closely related.

There are two main uses of these techniques: (1) we can eliminate (or average out) different fields from the energy function to obtain effective energies depending on only some of the fields (hence relating this framework to previous theories), and (2) one can obtain methods for finding deterministic solutions.

There is an additional important advantage in eliminating fields—one can impose constraints on the possible fields by only averaging over fields that satisfy these constraints. For example, Geiger and Yuille (1989) describe two possible energy function formulations of a winner-take-all network in which binary decision units determine the "winner" from a set of inputs. The constraint that there is only one winner can be expressed by (1) introducing a term in the energy function to penalize configurations with more than one winner, or (2) computing the mean fields by averaging only over configurations with a unique winner. The second method is definitely preferable in general because it enforces the constraint more strongly. Moreover, it leads to a very simple solution of the winner-take-all problem.

For the first level theory it is possible to eliminate the disparity field to obtain an effective energy $E_{eff}(V_{ij})$ depending only on the binary matching field V_{ij}, which is related to cooperative stereo theories (Dev, 1975; Marr & Poggio, 1976). Alternatively, one can eliminate the matching fields to obtain an effective energy $E_{eff}(d)$ depending only on the disparity. The second approach seems to be better since it incorporates the constraints on the set of possible matches implicitly rather than imposing them explicitly in the energy function (as the first method does).

One can also average out the line process fields or the matching fields or both for the second and third level theories. This leaves us again with a theory depending only on the disparity field.

Alternatively, one can use mean field theory methods to obtain deterministic algorithms for minimizing the first level theory $E_{eff}(V_{ij})$. These differ from the standard cooperative stereo algorithms and should be more effective (though not as effective as using $E_{eff}(d)$) since they can be interpreted as performing the cooperative algorithm at finite temperature, thereby smoothing local minima in the energy function.

Psychophysical Support for the Bayesian Framework

The experiments discussed earlier show that depth can be derived from images with disparate shading even in the absence of disparate edges. The perceived depth, however, was weaker for shading disparities (70 percent of the true depth). Putting in edges or features helped improve the accuracy of the depth perception. But in some cases these additional features appeared to decouple from the intensity and were perceived to lie above the depth surface generated from the intensity disparities.

These results are in general agreement with the Bayesian framework. The edges give good estimates of disparity and so little a priori smoothness is required and an accurate perception results. The disparity estimates from the intensity, however, are far less reliable (small fluctuations of intensity might yield large fluctuations in the disparity). Therefore, more a priori smoothness is required to obtain a stable result. This gives rise to a weaker perception of depth.

The use of the peak as a matching feature is vital (at least for the edgeless case) since it ensures that the image intensity is accurately matched (some stereo theories based purely on intensity give an incorrect match for these stimuli [M. Gennert, personal communication; see figure 20.15]). For these images, however, the peak is difficult to localize and depth estimates based on it are not very reliable. Thus the peak is not able to pull the rest of the surface to the true depth.

Pulling up did occur for the edgeless case if a feature (ring) was added at the peaks of the images (figure 20.16). This is consistent with our theory since, unlike the peaks, features are easily localized, and matching them would give a good depth estimate. Our present theory, however, is not consistent with a perception that sometimes occurred for this stimulus. In some cases the dots were perceived as lying above the surface rather than being part of it. This may be explained by the extension of our theory to transparent surfaces (Yuille, Yang & Geiger, 1990).

Additional support for this framework comes from the experiment of Bülthoff and Fahle (1989; see also Bülthoff, Fahle & Wegman, 1990) in which perceived depth for different matching primitives and disparity gradients was precisely measured. The results of these experiments suggest that several types of primitives are used for correspondence, but that some primitives are better than others. Perceived depth decreased as a function of the disparity gradient. This effect was strongest for horizontal lines, strong for pairs of dots or similar features, and weak for dissimilar features and nonhorizontal lines. An explanation in terms of the Bayesian framework assumes that these effects are due to the matching strategy and are based on the second level theory. The idea is that the smoothness term is required to give unique matching but that its importance, measured by λ, increases as the feature become more similar. If the features are sufficiently similar, then smoothness (or some other a priori assumption) must be used to obtain a unique match, leading to biases towards the frontoparallel plane. The greater the similarity between features, the more the need for smoothness and hence the stronger the bias toward the frontoparallel plane. The fall-off of perceived depth with increasing disparity gradient is modeled in detail by means of the second level theory in Yuille et al. (1989).

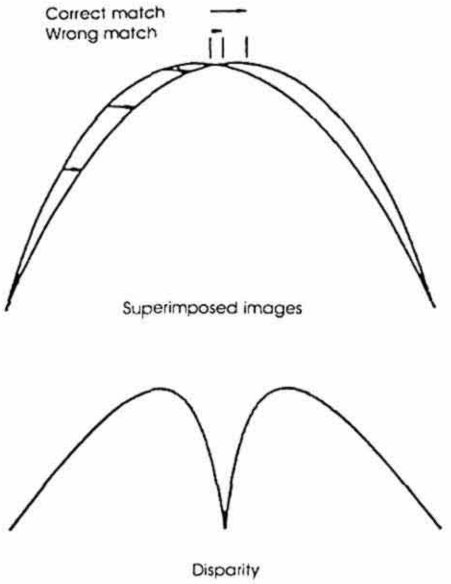
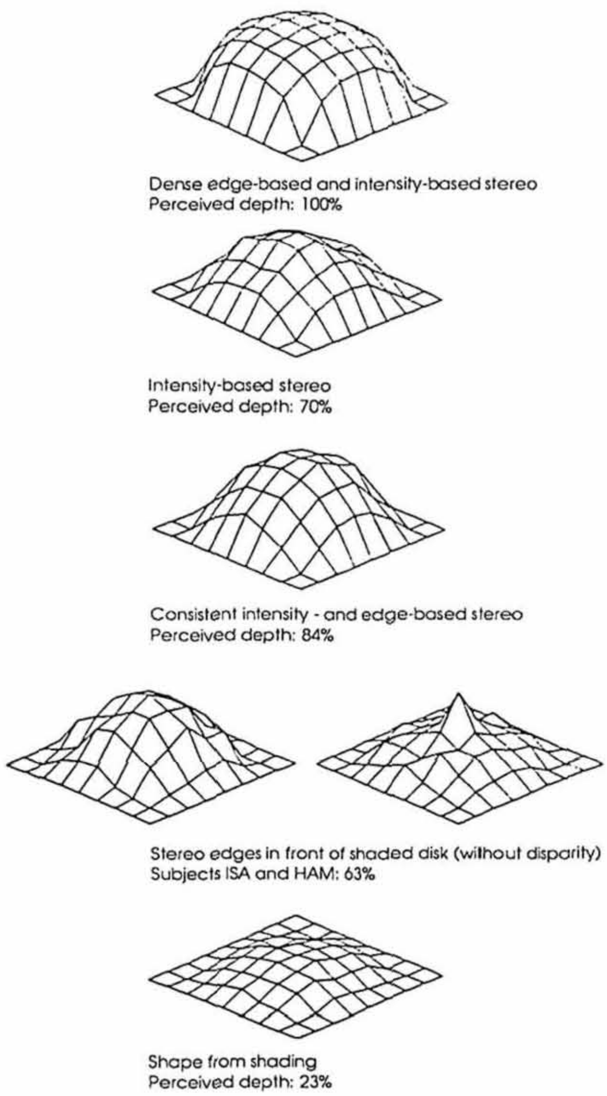


Fig. 20.15
False matching. The upper figure superimposes the left and right image and shows how the midpoints of the images (which have idientical intensity) are incorrectly matched by some intensity-based stereo theories, giving rise to the disparity profile shown in the lower figure with a dip in the center.

Final Remarks

In this chapter we discussed different modules for shape perception and their interaction. One can categorize these interactions in two broad classes. In one, the cues are *consonant* (noncontradictory). For example, consider view-



Dense edge-based and intensity-based stereo
Perceived depth: 100%

Intensity-based stereo
Perceived depth: 70%

Consistent intensity - and edge-based stereo
Perceived depth: 84%

Stereo edges in front of shaded disk (without disparity)
Subjects ISA and HAM: 63%

Shape from shading
Perceived depth: 23%

Fig. 20.16
Surface interpolation. The upper three figures show surface interpolation for: (1) dense edges and intensity-based stereo, (2) intensity-based stereo, and (3) sparse edges and consistent shading. The next two figures illustrate pulling: an edge token in front of intensity patterns with no relative disparity (no intensity-based stereo) pulls the surface up (*left*) but can sometimes cause a transparency percept (*right*) of the token lying in front of the intensity surface. The final figure shows the perceived depth without the edge token. (Redrawn from Bülthoff & Mallot, 1990.)

ing a golf ball with both eyes. There will be consistent depth information from stereo, shading, and texture cues. Viewing an image of the same golf ball in a photograph, however, puts the stereo cue into *conflict* with shading and texture.

Psychophysicists have attempted to deal with the first case by taking weighted linear combinations with some success (Bruno & Cutting, 1988; Dosher, Sperling & Wurst, 1986). Some experiments discussed in this chapter, however, do not seem consistent with such a model.

The case of conflicting cues seems to require significant nonlinearity and usually requires a different, and independent, mechanism. For example, this case is explicitly excluded in the statistical framework for fusion of depth information proposed by Maloney and Landy (1989).

Workers in computer vision have tended to use an alternative viewpoint. A recent book on sensor fusion (Clark & Yuille, 1990) proposed a distinction between *weak* methods in which modules compute depth independently and combine their results (perhaps by linear combination) and *strong* methods in which two modules interact during computation, usually in a very nonlinear way. They argue that strong methods are preferable since individual modules may be using conflicting assumptions. These theories also seem rich enough to encompass both the categories defined by psychophysicists.

These theories are expressed in a Bayesian framework that can be used both for describing the individual modules and for their integration. Although there are many other methods for dealing with individual modules, the Bayesian approach subsumes a number of these methods by isolating the key assumptions used by these theories.

Acknowledgments

The work in the first part of this chapter was done in close collaboration with Hanspeter Mallot, and parts of it were published in Bülthoff and Mallot (1988, 1990). The shape from highlights section is based on articles with Andrew Blake (Blake & Bülthoff, 1990, 1991; Bülthoff & Blake, 1989). The computational model for the integration of early vision models is based on a collaboration with Alan Yuille and Davi Geiger (Yuille, Geiger & Bülthoff, 1989). I thank all of them for many interesting discussions and a very enjoyable collaboration.

This chapter describes research done within the Center for Biological Information Processing (Whitaker College) at the Massachusetts Institute of Technology and in the Department of Cognitive and Linguistic Sciences, Brown University. Support for this work is provided by a grant from the Office of Naval Research, Cognitive and Neural Sciences Division and a NATO Collaborative Grant No. 0403/87.

Appendices

A. Computer Graphic Psychophysics

Images of smooth- and flat-shaded (polyhedral) ellipsoids of revolution were generated by either ray-tracing techniques or with a solid modeling software package (S-Geometry, Symbolics Inc.; figure 20.17). The polyhedral objects were derived from quadrangular tessellations of the sphere along meridian and latitude circles. The objects were elongated along an axis in the equatorial plane of the tessellated sphere. Thus the two types of objects differed mainly in the absence or presence of edges. As compared to spheres, the objects were elongated by the factors 0.5, 1.0, 2.0, 3.0, 4.0, and 5.0. With an original radius of 6.7 cm, this corresponds to depth values between 3.3 and 33.3 cm. In the following, all semidiameters (elongations) are given as multiples of 6.7 cm. In most experiments objects were viewed end-on, that is, the axis of rotational symmetry was orthogonal to the display screen. In an additional experiment, objects could be rotated around a diagonal axis in the display plane. As an example, the objects displayed in figure 20.2 are rotated around that axis by plus and minus 45°, respectively.

For the computation of the smooth-shaded ellipsoids, a ray-tracing operation was performed (Bülthoff & Mallot, 1988). The illuminant was modeled as a point source at infinity (parallel illumination) centrally behind the observer. For some control experiments, oblique directions of illumination (upper left and lower right) were used. Surface shading was computed according to the Phong model (Phong, 1975), consisting of an ambient, a diffuse (Lambertian), and a specular component. For Lambertian shading, the ambient and specular components were zero, while for specular shading (highlight), a combination of 30 percent ambient, 10 percent diffuse, and 60 percent specular reflectance (specular exponent 7.0) was chosen. Since our objects were convex, no cast shadows or repeated reflections had to be considered.

B. Experimental Procedure

We displayed either a pair of disparate images (stereo pair) or a nondisparate view of the object as seen from be-

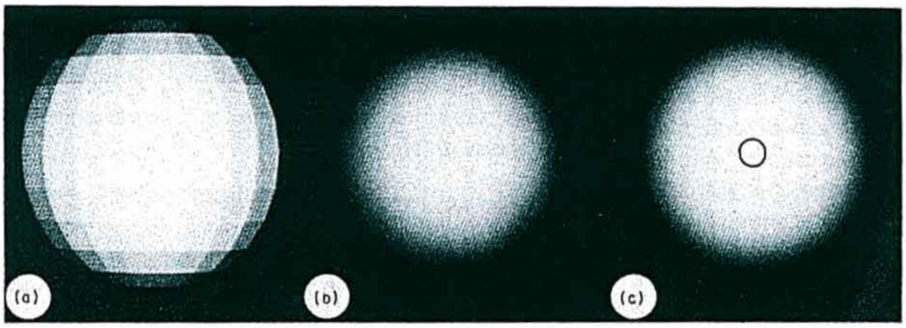


Fig. 20.17
Flat- and smooth-shaded surfaces. (a, b) Discontinuous and smooth intensity variations in images of polyhedra and ellipsoids provide cues for edge-based stereo, shape-from-disparate-shading, and shape-from-shading. (c) Smooth ellipsoids with sparse edge information have been used in experiments on the interaction of edge-based stereo and shape-from-shading. All images could be displayed as stereograms or as pairs of identical images. In image (c), the disparities of shading and the edge token (ring) could be varied independently. (Redrawn from Bülthoff & Mallot, 1988.)

tween the two eyes on a CRT Color Monitor (Mitsubishi UC-6912 High-Resolution Color-Display Monitor, Resolution (H × V) 1024 by 874 pixels; bandwidth ± 3 dB between 50 Hz and 50 MHz, short persistence phosphor). The disparate images were interlaced (even lines for the left image and odd lines for the right image) with a frame rate of 30 Hz. This technique allows one to display the left and right view at about the same location on the monitor and therefore treats any geometric distortion of the monitor equally for both eyes. Errors in displayed disparities due to geometric distortions of the monitor are therefore avoided. Both disparate and nondisparate images were viewed binocularly through shutter glasses (Stereo-Optic Systems, Inc.) which were triggered by the interlace signal to present the appropriate images only to the left and right eye. The objects were shown in black and white with a true resolution of 254 graylevels using a 10-bit D/A-Converter. The background was uniformly colored in half-saturated blue. The screen was viewed in complete darkness.

Local Depth Probe Technique

Perceived depth was measured by adjusting a small, red, square (4 by 4 pixel) depth probe to match the perceived depth of the surface interactively (with the computer mouse). This probe was displayed in interlaced mode together with the disparate images. This is a computer graphics version of a binocular rangefinder developed by Gregory (1966) called "Gregory's Pandora's Box" by some investigators with the additional advantage that the accommodation cue is eliminated. Measurements were performed at 45 vertices of a Cartesian grid in the image plane in random order. The initial disparity of the depth probe was randomized for each measurement to avoid hysteresis effects. Subjects were asked to move the cursor back and forth in depth until it finally seemed to lie directly on top of the displayed test surface. After some training sessions, subjects felt comfortable with this procedure and achieved reproducible depth measurements. Subjects included the authors (corrected vision) and one naive observer, all with normal stereo vision as tested with natural and random dot stereograms.

Global Shape Comparison Technique

The global shape comparison technique was used mainly for those cues that required monocular viewing. It is also useful for cues that are processed more globally and would be hindered by focussed attention on the local probe. Depending on the task, this technique was used in two different ways. To measure shape from shading and/or texture with the global probe we displayed a stereoscopically viewed reference object in the same orientation as the probe. The task of the subject was to change the shading or the texture (or both together) in order to match the shape with the reference object. This could be done almost in real-time by fast recall from computer memory of precomputed images of different shapes and/or orientations. The reference object did not contain any shading or texture cue beside the disparate rings on its surface to avoid any cross-comparison with the depth cues to be tested.

C. Data Evaluation

Depth Probe Technique

The depth probe technique leads to a depth map measured locally at 45 positions in the image plane. In order to derive a global measure of perceived depth we performed a principal component analysis on all data sets, treating each one as a point in 45-space. Variance of the perceived shapes was found mainly (94 percent) along the first principal axis, whose corresponding loading was very close to an ideal ellipsoid (or sphere). The second component accounted for only 1.4 percent of the total variance. We therefore chose the overall elongation, namely, the coefficient associated with the first principal component, as a measure of perceived depth for a given cue combination (see figure 20.5).

Global Shape Comparison Technique

The depth comparison data were averaged over different runs and over two to four subjects. The mean number of runs was about 180 and the average correlation between displayed and estimated shape was 0.83. In order to distinguish easily between over or underestimation of depth we give the mean slope for each depth cue. A slope of 1.0 is naturally the veridical perception and a slope > 1 is an underestimation of curvature (see figure 20.6).

D. A Bayesian Framework for Stereo

The First Level: Matching Field and Disparity Field

The basic idea is that there are a number of possible primitives that could be used for matching and that these all contribute to a disparity field $d(x)$. This disparity field exists even where there is no source of data. The primitives considered here are features such as edges in image brightness. Edges typically correspond to object boundaries, and other significant events in the image. Other primitives, such as peaks in the image brightness or texture features, can also be added. In the following, the theory is described for the one-dimensional case.

One can assume that the edges and other features have already been extracted from the image in a preprocessing stage. The matching elements in the left eye consist of features at positions x_{i_L}, for $i_L = 1, \ldots, N_l$. The right eye contains features at positions x_{a_R}, for $a_R = 1, \ldots, N_r$. A *matching field* is defined as a set of binary matching elements $V_{i_L a_R}$, such that $V_{i_L a_R} = 1$ if point i_L in the left eye corresponds to point a_R in the right eye, and $V_{i_L a_R} = 0$ otherwise. A *compatibility field* $A_{i_L a_R}$ is defined over the range [0, 1]. For example, it is 1 if i_L and a_R are compatible (i.e., features of the same type), 0 if they are incompatible (an edge cannot match a peak),

One can now define a cost function $E(d(x), V_{i_L a_R})$ of the disparity field and the matching elements. There are several methods to estimate the fields $d(x)$, $V_{i_L a_R}$ given the data. A standard estimator is to minimize $E(d(x), V_{i_L a_R})$ with respect to $d(x)$, $V_{i_L a_R}$.

$$\begin{aligned} E(d(x), V_{i_L a_R}) = {} & \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 \\ & + \lambda \left\{ \sum_{i_L} \left(\sum_{a_R} V_{i_L a_R} - 1 \right)^2 \right. \\ & \left. + \sum_{a_R} \left(\sum_{i_L} V_{i_L a_R} - 1 \right)^2 \right\} \\ & + \gamma \int_M (Sd)^2 \, dx. \qquad (11) \end{aligned}$$

The first term gives a contribution to the disparity obtained from matching i_L to a_R. The fourth term imposes a smoothness constraint on the disparity field imposed by a smoothness operator S.

The second and third term encourage features to have a single match, they can be avoided by requiring that each column and row of the matrix $V_{i_L a_R}$ contains only one 1. Minimizing the energy function with respect to $d(\mathbf{x})$ and $V_{i_L a_R}$ will cause the matching that results in the smoothest disparity field. The coefficient γ determines the amount of a priori knowledge required. If all the features in the left eye have only one compatible feature in the right eye then little a priori knowledge is needed and γ may be small. If all the features are compatible then there is matching ambiguity which the a priori knowledge is needed to resolve, requiring a larger value of γ and hence more smoothing. This gives a possible explanation for the depth reduction effects discussed in Bülthoff, Fahle & Wegman (1990).

The theory can be extended to two dimensions in a straightforward way. The matching elements $V_{i_L a_R}$ must be constrained to only allow for matches that use the epipolar line constraint. The disparity field will have an additional smoothness constraint perpendicular to the epipolar line which will enforce figural continuity.

Finally, and perhaps most importantly, a form for the smoothness operator S has to be chosen. Marr (1982) proposed that, to make stereo correspondence unambiguous, the human visual system assumes that the world consists of smooth surfaces. This suggests that one should choose a smoothness operator that encourages the disparity to vary smoothly spatially. In practice the assumptions used in Marr's two theories of stereo are somewhat stronger. Marr and Poggio I (1976) encourages matches with constant disparity, thereby enforcing a bias to the frontoparallel plane. Marr and Poggio II (1979) uses a coarse to fine strategy to match nearby points, hence encouraging matches with minimal disparity and thereby giving a bias towards the fixation plane.

An alternative approach is to introduce discontinuity fields that break the smoothness constraint. For these theories the experiments described in Bülthoff et al. (1989, 1990) are consistent with S being a first order derivative operator. This is also roughly consistent with Marr and Poggio I (1976). A default choice is therefore $S = \partial/\partial x$.

The Second Level: Adding Discontinuity Fields

The first level theory is easy to analyze but makes the a priori assumption that the disparity field is smooth everywhere, which is false at object boundaries. There are several standard ways to allow smoothness constraints to break (Blake, 1983; Geman & Geman, 1984; Mumford &

Shah, 1985). Here, a discontinuity field $l(x)$ is introduced which is represented by a set of curves C.

Introducing the discontinuity fields C gives an energy function

$$E(d(x), V_{i_L a_R}, C) = \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 + \lambda \left\{ \sum_{i_L} \left(\sum_{a_R} V_{i_L a_R} - 1 \right)^2 + \sum_{a_R} \left(\sum_{i_L} V_{i_L a_R} - 1 \right)^2 \right\} + \gamma \int_{M-C} (Sd)^2 \, dx + M(C), \qquad (12)$$

where smoothness is not enforced across the curves C, and $M(C)$ is the cost for enforcing breaks.

The Third Level: Adding Intensity Terms

The final version of the theory couples intensity based and feature based stereo. The psychophysical results suggest that this is necessary. The energy function becomes

$$E(d(x), V_{i_L a_R}, C) = \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 + \mu \int \left\{ L(x) - R(x + d(x)) \right\}^2 dx + \lambda \left\{ \sum_{i_L} \left(\sum_{a_R} V_{i_L a_R} - 1 \right)^2 + \sum_{a_R} \left(\sum_{i_L} V_{i_L a_R} - 1 \right)^2 \right\} + \gamma \int_{M-C} (Sd)^2 \, dx + M(C). \qquad (13)$$

If certain terms are set to zero in equation 13, it reduces to previous theories of stereo. If the second and fourth terms are kept, without allowing discontinuities, it is similar to work by Gennert (1987) and Barnard (1986).

Thus the cost function (13) reduces to well-known stereo theories in certain limits. It also shows how it is possible to combine feature and brightness data in a natural manner. In addition it can be modified to include monocular cues (Clark & Yuille, 1990).

A similar theory for integrating different cues for motion perception was proposed by Yuille and Grzywacz (1988), although this theory did not involve discontinuity fields.

References

Aliomonos, J. & Swain, M. J. (1985). Shape from texture. IEEE Joint Conference on Artificial Intelligence, 926–931.

Arnold, R. D. & Binford, T. O. (1980). Geometric constraints in stereo vision. *Society of Photo-Optical Instrumentation and Engineering, 238*, 281–292.

Bajcsy, R. & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Vision, Graphics, and Image Processing, 5*, 52–67.

Barnard, S. (1986). *Proceeding of the Image Understanding Workshop*, Los Angeles.

Barrow, H. G. & Tenenbaum, J. M. (1981). Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence, 17*, 75–116.

Beck, J. (1972). *Surface color perception*. New York: Cornell University Press.

Blake, A. (1983). The least disturbance principle and weak constraints. *Pattern Recognition Letters, 1*, 393–399.

Blake, A. (1985). Specular stereo. *Proceedings of the 9th IJCAI Conference*, 973–976.

Blake, A. & Brelstaff, G. J. (1988). Geometry from specularities. In *Proceedings of the International Conference on Computer Vision* (pp. 394–403). Washington, DC: IEEE.

Blake, A. & Bülthoff, H. H. (1990). Does the brain know the physics of specular reflection? *Nature, 343*, 165–168.

Blake, A. & Bülthoff, H. H. (1991). Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society London B, 331*, 237–252.

Blake, A. & Zisserman, A. (1987). *Visual reconstruction*. Cambridge, MA: MIT Press.

Blake, A., Zisserman, A. & Knowles, G. (1985). Surface description from stereo and shading. *Image and Vision Computing, 3*, 183–191.

Bruno, N. & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General, 117*, 161–170.

Buckley, D., Frisby, J. P. & Mayhew, J. E. W. (1989). Integration of stereo and texture cues in the formation of discontinuities during three-dimensional surface interpolation. *Perception, 18*, 563–588.

Bülthoff, H. H. & Blake, A. (1989). Does the seeing brain know physics. *Investigative Ophthalmology and Visual Science, Suppl., 30*, 262.

Bülthoff, H. H. & Fahle, M. (1989). Disparity gradients and depth scaling. *MIT Artificial Intelligence Memo, 1175*.

Bülthoff, H. H., Fahle, M. & Wegman, M. (1990). Perceived depth scales with disparity gradient. *Perception*, in press.

Bülthoff, H. H., Little, J. J. & Poggio, T. (1989). A parallel algorithm for real-time computation of motion. *Nature, 337,* 549–553.

Bülthoff, H. H. & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America, 5,* 1749–1758.

Bülthoff, H. H. & Mallot, H. A. (1990). Integration of stereo, shading and texture. In A. Blake & T. Troscianko (Eds.), *AI and the eye.* New York: John Wiley and Sons.

Bülthoff, H. H. & Yuille, A. (1990). Models for seeing surfaces and shapes. *Comments in Theoretical Biology,* in press.

Clark, J. & Yuille, A. (1990). *Data fusion for sensory information processing.* Norwell, MA: Kluwer Academic Press.

Cutting, J. E. & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General, 113,* 198–216.

Dev, P. (1975). Perception of depth surfaces in random-dot stereograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-2,* 333–340.

Dosher, B. A., Sperling, G. & Wurst, S. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973–990.

Gelb, A. (1974). *Applied optimal estimation.* Cambridge, MA: MIT Press.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6,* 721–741.

Geiger, D. & Girosi, F. (1989). Parallel and deterministic algorithms from MRFs: Integration and surface reconstruction. *MIT Artificial Intelligence Laboratory Memo, 1114.*

Geiger, D. & Yuille, A. (1989). *A common framework for image segmentation* (Harvard Robotics Laboratory Technical Report No. 89–7).

Gennert, M. A. (1987). A computational framework for understanding problems in stereo vision. *MIT Artifical Intelligence Laboratory Thesis.*

Gregory, R. L. (1966). *Eye and brain.* New York: McGraw-Hill.

Grimson, W. E. L. (1982). A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society London* B, *298,* 395–427.

Haynes, S. M. & Jain, R. (1987). A qualitative approach for recovering depth in dynamic scenes. *Proceedings of the of IEEE Workshop on Computer Vision,* Miami Beach, 66–71.

Hildreth, E. C. (1983). The detection of intensity changes by computer and biological vision systems. *Computer Vision, Graphics and Image Processing, 22,* 1–27.

Hildreth, E. C. (1984). Computations underlying the measurement of visual motion. *Artificial Intelligence Journal, 23,* 309–354.

Ikeuchi, K. & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence, 17,* 141–184.

Julesz, B. (1971). *Foundations of cyclopean perception.* London: The University of Chicago Press, Ltd.

Kender, J. R. (1979). Shape from texture: An aggregation transform that maps a class of textures into surface orientation. In *Proceedings, International Joint Conference on Artificial Intelligence,* Tokyo, Japan.

Kirkpatrick, S., Gelatt, C. D. Jr. and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220,* 671–680.

Koenderink, J. J. (1986). Optic flow. *Vision Research, 26,* 161–180.

Koenderink, J. J. & van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics, 21,* 29–35.

Landy, M. S. (1987). Parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A, 4,* 864–877.

Longuet-Higgins, H. C. & Prazdny, K. (1981). The interpretation of a moving retinal image. *Proceedings of the Royal Society London* B, *208,* 385–397.

Maloney, L. T. & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. In W. A. Pearlman (Ed.), *Visual Communications and Image Processing IV. Proceedings of the SPIE, 1199,* 1154–1163.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Marr, D. & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society London* B, *207,* 187–217.

Marr, D. & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194,* 283–287.

Marr, D. & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society London B, 204,* 301–328.

Mayhew, J. E. W. & Frisby, J. P. (1981). Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence, 17,* 349–386.

Mayhew, J. E. W. & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature, 297,* 376–379.

Metropolis, N. Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal Physical Chemistry, 21,* 1087–1091.

Mingolla, E. & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics, 53,* 137–151.

Mumford, D. and Shah, J. (1985). Boundary detection by minimizing functionals, I, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* San Francisco.

Parisi, G. (1988). *Statistical field theory.* Reading, MA: Addison-Wesley.

Pentland, A. P. (1984). Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6*, 170–187.

Pentland, A. P. (1985). A new sense for depth of field. *IEEE Joint Conference on Artificial Intelligence*, 988–994.

Pentland, A. P. (1986). Shading into texture. *Artificial Intelligence, 29*, 147–170.

Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM, 18*, 311–317.

Poggio, G. & Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience, 7*, 379–412.

Poggio, T., Torre, V. & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*, 314–319.

Prazdny, K. (1985). Detection of binocular disparities. *Biological Cybernetics, 52*, 93–99.

Stevens, K. A. (1981). The visual interpretation of surface contours. *Artificial Intelligence, 17*, 47–73.

Stevens, K. A. & Brookes, A. (1987). Probing depth in monocular images. *Biological Cybernetics, 56*, 355–366.

Stevens, K. A. & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research, 28*, 371–386.

Todd, J. T. & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 583–595.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception, 13*, 255–274.

Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence, 17*, 17–47.

Yuille, A. L. (1989). (Harvard Robotics Laboratory Technical Report 89–12).

Yuille, A. L., Geiger, D. & Bülthoff H. H. (1989). *Stereo integration, mean field theory and psychophysics* (Harvard Robotics Laboratory Technical Report 89–1).

Yuille, A. L. & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature, 333*, 71–74.

Yuille, A. L., Tong Yang, & Geiger, D. (1990). *Robust statistics, transparency and correspondence* (Harvard Robotics Laboratory Technical Report 90-7).

Zisserman, A., Giblin, P. & Blake, A. (1989). The information available to a moving observer from specularities. *Image and Vision Computing, 7*, 38–42.