6 Integration of Stereo, Shading and Texture

Heinrich H. Bülthoff[1] and Hanspeter A. Mallot[2]

[1]*Center for Biological Information Processing, Massachusetts Institute of Technology, Cambridge, MA, USA*, and* [2]*Institut für Zoologie III, Johannes Gutenberg-Universität, Mainz, FRG*

INTRODUCTION

Integration of cues is one of the key features of natural vision that underlie its performance and robustness. In this chapter, we investigate the integration of various depth cues into different percepts related to three-dimensional structure.

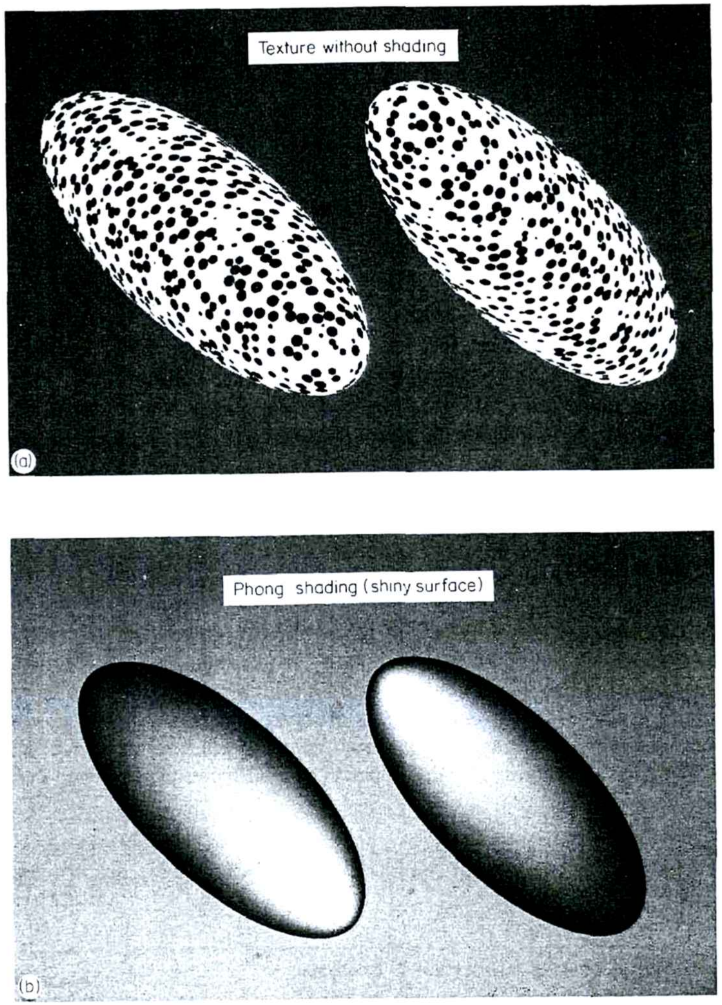
Most of the depth cues known in psychophysics have been formalized in terms of computational theory and have been implemented as single modules in machine vision systems. Mutually related studies from psychophysics and computational theory exist mainly for stereo (Julesz, 1971; Marr & Poggio, 1979; Mayhew & Frisby, 1981) and shading (Ikeuchi & Horn, 1981; Pentland, 1984; Blake *et al.* 1985; Mingolla & Todd, 1986). There are also a number of studies on depth from texture (Bajcsy & Lieberman, 1976; Kender, 1979; Witkin, 1981; Pentland, 1986), line drawings (Barrow & Tenenbaum, 1981), surface contours (Stevens, 1981; Stevens & Brooks, 1987), and structure-from-motion (Ullman, 1979; Longuet-Higgins & Prazdny, 1981; Grzywacz & Hildreth, 1987). Machine implementations are quite successful for synthetic images but less reliable for natural images. On the contrary, the human visual system deals much better with natural images and multiple depth cues than with single depth cues in synthetic images (e.g. random-dot stereograms). In order to study the superior performance of human vision in the integration of multiple depth cues, we developed methods for quantitative measurement of depth perception with complex yet well-controlled images.

*Now at Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912, USA.

AI and the Eye Edited by A. Blake and T. Troscianko

Integration of Multiple Depth Cues

The visual system derives a variety of information about the three-dimensional structure of the environment from different depth cues. This is illustrated in Figure 1 where three pairs of ellipsoids are shown whose axes of elongation are orthogonal to each other. The orthogonal orientation is best seen in Figure 1(c), where texture and specular shading provide sufficient 3D information. If texture is used without shading (Figure 1a), the orientation of the objects can usually be perceived correctly while the objects themselves appear flat. Vice versa, if shading is the only cue (Figure 1b), the objects appear nicely curved but it is difficult to see them orthogonal to each other. We therefore argue that at least at a low level of abstraction, multiple representations of three-dimensional structure exist, which will be called *3D-descriptors* in this chapter. These 3D-descriptors are sufficient for simple visual behavior and it is unclear whether a single complete representation of visible surfaces exists at all.



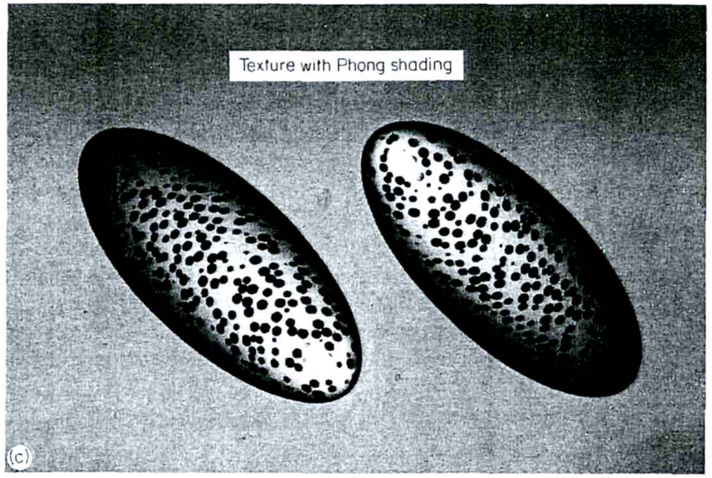


Figure 1. (a–c) Different depth cues provide information about different 3D-descriptors.

Our approach is schematically described in Figure 2.

Two Aspects of Integration

Raw data from depth cues such as shading, texture or disparity can be thought of as a trivial, or zero-order, representation of the spatial structure of a scene. Based on these data, higher-order descriptors are derived that make interesting spatial properties of the viewed scene explicit. The question of what an interesting 3D-descriptor is, has to be answered in the light of the action that it should subserve. For example, a pointwise depth map is useful for threading a needle, while curvatures might be sufficient for the recognition of complex 3D shapes (such as faces). Eventually, this process may or may not lead to a single complete representation of visible surfaces as was proposed by Marr & Nishihara (1978). In this framework, integration involves two largely independent processes:

1. *Assignment of descriptors to cues.* Which cues are relevant to one particular 3D-descriptor? For example, occlusion contributes more readily to depth ordering than to surface curvature. Shading contributes more qualitatively to curvature than quantitatively to a depth map, or texture more to object orientation than

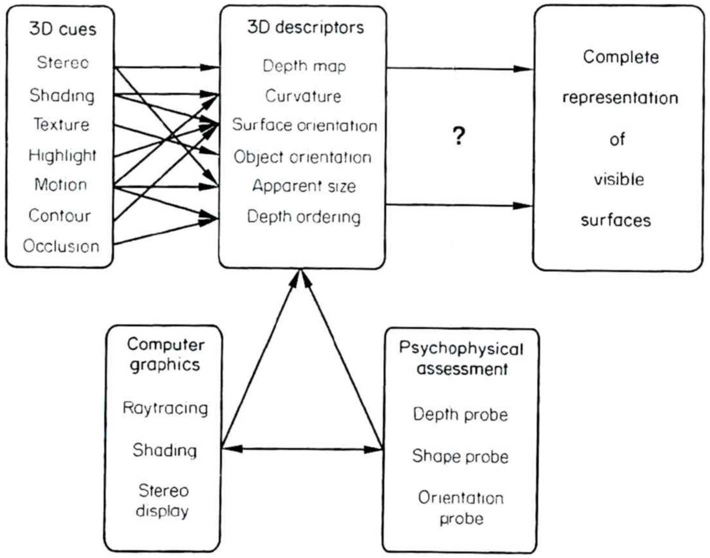


Figure 2. Integration of depth cues. 3D structure perceived from 2D images can be represented at different levels of abstraction. The depth cues themselves constitute multiple *zero-order* representations. Higher-order representations, i.e. the *3D-descriptors*, can be derived from interaction and integration of several of these zero-order representations. Different psychophysical experiments (much as computer vision tasks) involve various combinations of the 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.

to object form. The latter two cases can be qualitatively verified by observing Figure 1.

2. *Cue interaction per descriptor*. From the computational point of view we can ask how are the cues contributing to the computation of a particular descriptor combined? In principle, there are several types of useful interactions which are not mutually exclusive:

(i) *Accumulation*. Information from different cues could be accumulated in different ways such as probability summation or the linear summation model for the integration of stereo and proximity luminance covariance proposed by Dosher *et al.* (1986). A more computational approach to accumulation is joint regularization, where constraints from different cues are accounted for by means of a common cost function (Poggio *ei al.*, 1985; Terzopoulos, 1986).

(ii) *Cooperation*. Especially in the case of poor or noisy cues, modules might work synergistically. Here we think of the non-linear interactions of different cues which can be treated, for example, with the coupled Markov random field approach (Marroquin *et al.*, 1987).

(iii) *Disambiguation*. A particular case of a non-linear interaction is the case where information from one cue is used to locally disambiguate a representation derived from another one [e.g. stereo can disambiguate shading (Braunstein *et al.*, 1986)] or specular highlights can disambiguate the convex-concave ambiguity of shading (Blake & Bülthoff, 1989).

(iv) *Veto*. There can be unequivocal information from one cue that should not be challenged by others.

Most computational approaches to integration have focused on the second problem, i.e. the combination of different data types in one representation which is often thought to be unique. In our psychophysical experiments, we addressed both of the above questions by (a) measuring the contribution of individual cues in different matching tasks that correspond to certain 3D-descriptors, and (b) comparing these contributions in each of the matching tasks quantitatively.

Psychophysical 3D Measurements

The perception of three-dimensional scenes relies on many different depth cues and leads to various descriptions of that scene in terms of distance, surface orientation, and curvature, shape or form. We addressed various of these 3D-descriptors (depth map, curvature and object orientation) and depth cues (stereo, shading, highlight, texture).

3D-descriptors and matching tasks

1. *Perceived depth* was mapped with a small probe or cursor that was interactively adjusted to the perceived surface. The depth of this probe was defined by edge-based stereo disparities 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. Each adjustment of the probe gives a graded measurement of distance, or local depth, i.e. this experiment corresponds to the 3D-descriptor *depth map* mentioned in the scheme of Figure 2.
2. *Global shapes* of two objects with different combinations of depth cues were compared directly. Since all images showed end-on views of ellipsoids with different elongation, this measurement corresponds to *curvature* or *form* as a 3D-descriptor.
3. *Object orientation* can be measured in a matching task where long ellipsoids of different orientation have to be compared. While surface orientation is apparently hard to determine for human observers (Todd & Mingolla, 1983; Mingolla & Todd, 1986), the orientation of entire objects (e.g. orientation of *generalized cylinders*) can be measured easily in a matching task.

Depth cues and computer graphics

The relation of shading (with and without highlights), stereo and texture in the 3D perception of smooth and polyhedral surfaces was studied with computer graphics psychophysics. For polyhedral and textured objects, stereo disparities were associated with localized features, i.e. the intensity changes at the facet or texel boundaries, while for the smooth surfaces only shading disparities occurred. In addition, contours such as rings or lines could be drawn on the smooth surfaces to provide sparse edge information. The objects (ellipsoids of revolution viewed end-on) were chosen for the following reasons:

1. As is shown later, in the section on "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.
2. 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 provide any depth information.
3. 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 or mutual illumination.
4. End-on views of ellipsoids can be thought of as a model for the depth interpolation of a surface patch between sparse edge data.

METHODS

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.). The polyhedral objects were derived from quadrangular tesselations of the sphere along meridian and latitude circles. The objects were elongated along an axis in the equatorial plane of the tesselated 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 semi-diameters (elongations) are given as multiples of 6.7 cm. In Experiments 1 and 2, all objects were viewed end-on, i.e. the axis of rotational symmetry was orthogonal to the display screen. In Experiment 3, objects could be rotated around a diagonal axis in the display plane. As an example, the objects displayed in Figure 1 are rotated around that axis by plus and minus 45°, respectively.

The imaging geometry used in the computer graphics is shown in Figure 3. It differs from the usual camera geometry in that the image is constructed on a screen which is not perpendicular to the optical axis of the eyes. Note that the imaging geometry, and therefore the image itself, does not depend on the fixation point as long as the nodal points of the two eyes remain fixed at the positions E_l and E_r, respectively. Images were computed for a viewing distance of 120 cm and an interpupillary separation of 6.5 cm. When a point 10 cm in front of the center of the screen is fixated, Panum's fusional area of ± 10 min of arc (cf. Ardity, 1986) corresponds to an interval from 4.3 cm to 15.2 cm in front of the screen.

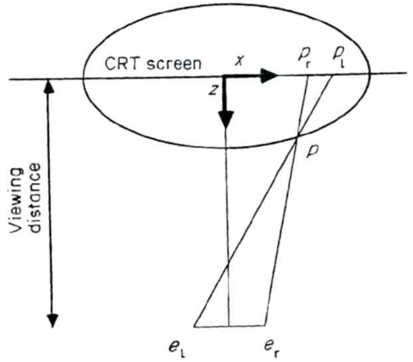


Figure 3. Imaging geometry. Projection onto the x–z-plane. Viewing distance is 120 cm. e_l, e_r: nodal points of the left and right eye, respectively. The distance between $\mathbf{e}_l$ and $\mathbf{e}_r$ is 6.5 cm. A point $\mathbf{p} \in \mathbf{R}^3$ is imaged on the screen at $\mathbf{p}_l$ for the view from the left eye and at $\mathbf{p}_r$ for the view from the, right eye.

For the computation of the smooth-shaded ellipsoids, a ray-tracing operation was performed.* The illuminant was modeled as an infinite point source (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), i.e. 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 (sometimes called highlight in the sequel), a combination of 30% ambient, 10% diffuse and 60% specular reflectance (specular exponent 7.0) was chosen. Since our objects were convex, no cast shadows or repeated reflections had to be considered.

* We write the equation of the ellipsoid as

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = 1, \qquad \mathbf{A} = \begin{pmatrix} a^{-2} & 0 & 0 \\ 0 & b^{-2} & 0 \\ 0 & 0 & c^{-2} \end{pmatrix}, \tag{1}$$

where a, b, c denote the semi-diameters. With $a = b = 1$, we have an ellipsoid of revolution. For a ray from e to p',

$$\mathbf{x} = \mathbf{e} + \mu(\mathbf{p}' - \mathbf{e}), \mu \in \mathbf{R}^+, \tag{2}$$

the ray-tracing amounts to the solution for μ of the quadratic equation:

$$(\mathbf{e} + \mu(\mathbf{p}' - \mathbf{e}))^T\mathbf{A}(\mathbf{e} + \mu(\mathbf{p}' - \mathbf{e})) = 1. \tag{3}$$

The image intensity at point p' was computed from this solution for an ideal Lambertian surface illuminated by parallel light from the z-direction. Note that for a point $\mathbf{x}$ on the surface of the ellipsoid $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$, the surface normal is simply $\mathbf{A}\mathbf{x}/\|\mathbf{A}\mathbf{x}\|$. The viewing direction and the axis of revolution of the ellipsoid were aligned.

Disparity and edge information (Experiment 1)

In a first series of experiments, we crossed disparity and dense edge information in shaded images. Four different image types were tested (Figure 4a, b):

1. Flat-shaded ellipsoid with disparity and edge information (D^+E^+).
2. Smooth-shaded ellipsoid with disparity but without edge information (D^+E^-). Both Lambertian and specular shading were tested.
3. Flat-shaded ellipsoid without disparity but with edge information (D^-E^+).
4. Smooth-shaded ellipsoid with neither disparity nor edge information (D^-E^-). Both Lambertian and specular shading were tested.

Illuminant direction (Experiment 2)

In a second series of experiments, we studied the influence of the illuminant direction in Lambertian shaded images with and without disparities (D^+E^-; D^-E^-). While in the first series illumination was from exactly behind the observer, we chose upper-left and lower-right directions ($\pm 14°$ azimuth and $\mp 13.6°$ elevation from the viewing direction).

Edge vs shading disparity (Experiment 3)

The third series of experiments addressed the interaction of smooth shading and sparse edge information provided by a small dark ring placed at the tip of the ellipsoid (contrast 0.11, radius 7.5 mm, covering less than 1% of the ellipsoid's image). Disparities of shading and ring were varied independently, leading to the following combinations (Figure 4c):

1. Disparate ring and disparate shading.
2. Disparate ring and non-disparate shading.
3. Non-disparate ring and disparate shading.
4. Non-disparate ring and non-disparate shading.
5. Disparate ring in front of uniformly grey non-disparate disk.

All experiments were performed with 4–6 different elongations (0.5–5.0) of the ellipsoids. The elongations were unknown to the observers.

Global shape comparison (Experiment 4)

The local depth probe technique used in the previous three experiments has some disadvantages with depth cues which have to be viewed preferably monocularly. Therefore, we developed a global shape comparison technique which allows the depth cues to be viewed monocularly and compared with a stereoscopically viewed shape reference.

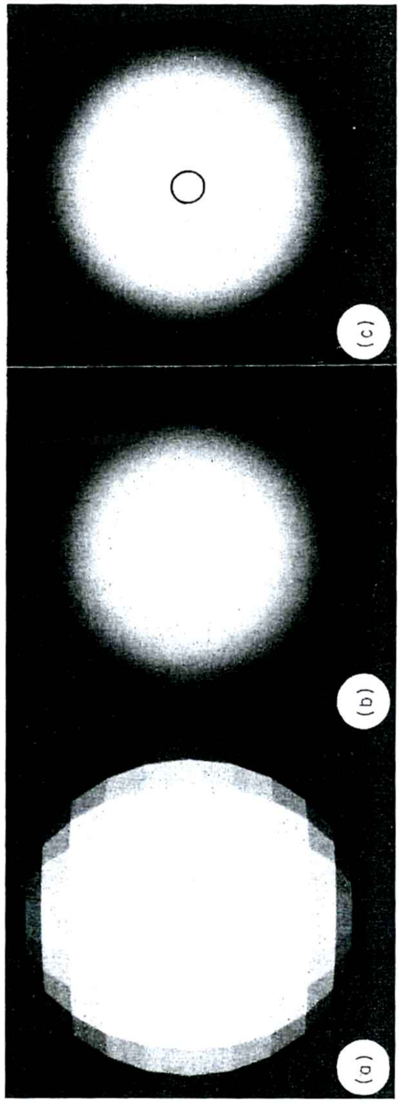


Figure 4. Flat- and smooth-shaded surfaces. (a), (b) Discontinuous and smooth intensity variations in images of polyhedra and ellipsoids provide cues for edge-based stereo, shape-from-disparate-shading and shape-from-shading (Experiment 1). (c) Smooth ellipsoid with sparse edge information has been used in experiments on the interaction of edge-based stereo and shape-from-shading (Experiment 3). All images could be displayed as stereograms or as pairs of identical images. In image (c), the disparities of shading and edge token (ring) could be varied independently.

Experimental Procedure

We displayed either a pair of disparate images (stereo pair) or a non-disparate view of the object as seen from between the two eyes on a CRT Color Monitor (Mitsubishi UC-6912 High-Resolution Color-Display Monitor, Resolution (H × V) 1024 × 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 the left and right views to be displayed at about the same location on the monitor and therefore treats any geometric distortion of the monitor equally to both eyes. Non-voluntary disparities are therefore avoided. Both disparate and non-disparate 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 grey-levels 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-shaped (4 by 4 pixel) depth probe to 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 which required monocular viewing. It is also useful for cues which are processed more globally and would be hindered by a too focused attention to 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.

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%) 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% of the total variance. We therefore chose the overall elongation, i.e. the coefficient associated to the first principal component, as a measure of perceived depth for a given cue combination (Figure 6).

Shape comparison technique

The depth comparison data were averaged over different runs and over 2-4 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 10).

RESULTS

Disparity and Edge Information (Experiment 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 5 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. As can be seen from Figure 6, the perceived elongation is almost correct when shading, intensity-based and edge-based disparity information are available (D^+E^+). 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)

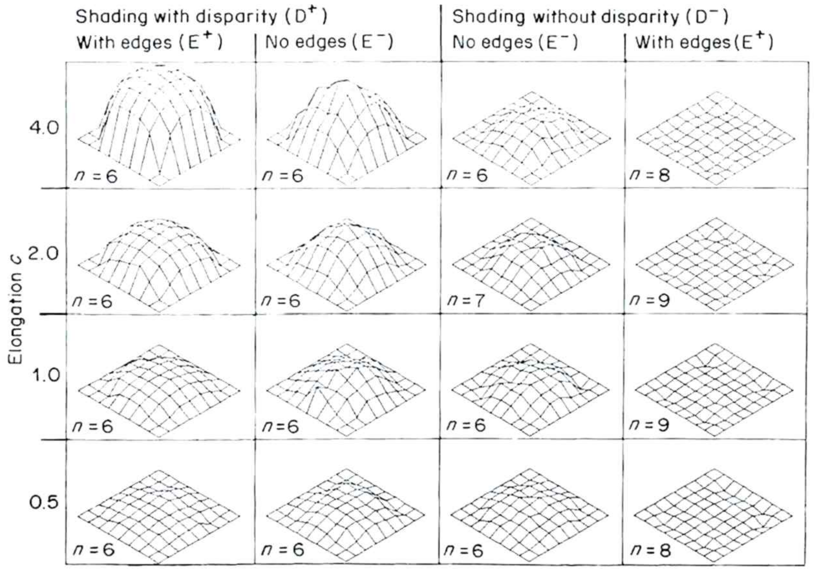


Figure 5. Perceived surfaces (Experiment 1). Each plot shows the average of 6–9 sessions from three subjects. Perceived depth decreases with the following sequence of cue combinations: disparity, edges and shading (D^+E^+); disparity and shading but no edges (D^+E^-); shading only (D^-E^-); contradictory disparity and shading (D^-E^+). The elongation of the displayed objects is denoted by *c* (depth not drawn to scale).

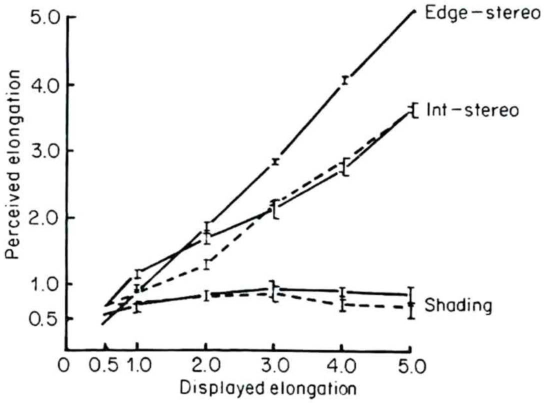


Figure 6. Perceived elongation. Depth perception decreases with fewer cues available. The significant separation between the middle and lower curves (smooth shading with and without disparity) illustrates the influence of disparity information even in the absence of edges. Solid lines: Lambertian shading; dashed lines: Phong shading.

instead of Lambertian shading did not change perceived depth significantly (dashed lines).

In experiment D^-E^+, two identical images (no 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 shading (and perspective) information and took slightly negative values which, however, were not significantly different from zero.

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 6) for smooth-shaded images with and without disparity information, however, establishes a significant contribution of shading disparities. The curves for D^+E^- and D^-E^- are significantly separated for all elongations except 0.5.

Illuminant Direction (Experiment 2)

Since the lighting conditions used in the preceding experiments were degenerate (no self-shadows) we measured smooth-shaded ellipsoids (D^+E^-, D^-E^-) with oblique directions of illumination. Light sources were placed in the upper-left and the lower-right in front of the object (±14° azimuth and ∓13.6° elevation towards the viewer). The results of these experiments (41 measurements, data pooled from all subjects) are depicted in Figure 7. The slight asymmetries present at elongation 4.0 result exclusively from the fact that no depth values were determined in the dark (shadowed) parts of the images. The data are in line with those of Experiment 1: shading disparities produce a significantly stronger depth perception than non-disparate shading (shape-from-shading). Furthermore, when illumination is from

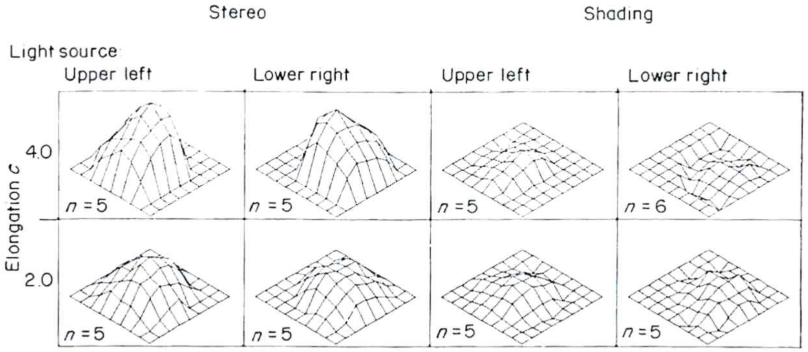
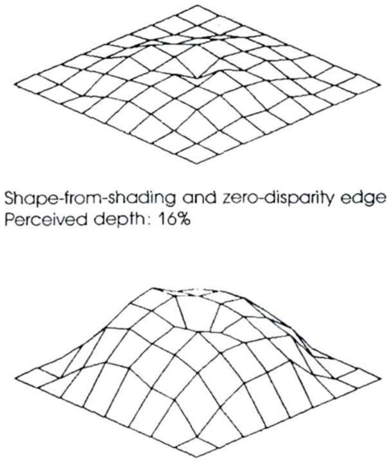


Figure 7. Perceived surfaces for oblique illuminations (Experiment 2). The data confirm the relevance of disparate shading and show the independence of the findings of Experiment 1 from the lighting conditions. No depth was measured in the self-shadow regions.

the lower-right, stereo prevents depth inversion which occasionally occurred in the non-disparate images (cf. negative depth in Figure 7; shading, lower-right).

Edge vs Shading Disparity (Experiment 3)

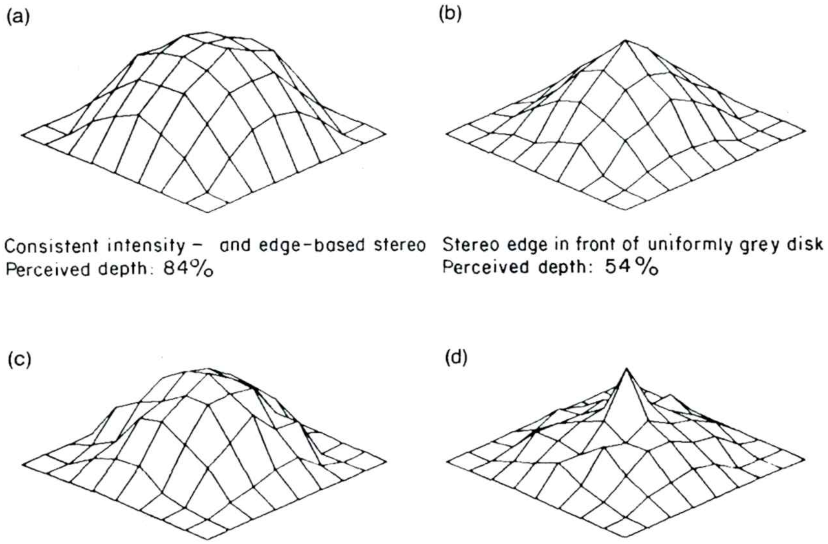
In contrast to the original measurements with polyhedral objects where edge information was distributed all over the surface, we now placed a small dark ring at the tip of the ellipsoid. Altogether, 126 measurements were performed with four different elongations. Figure 8 shows the results for the ring at zero disparity combined with non-disparate (a) and disparate shading (b). While the overall results resemble those of Experiment 1 (D$^+$E$^-$ and D$^-$E$^-$, respectively), zero depth is perceived in the vicinity of the ring. The cases with disparate edge information are summarized in Figure 9: in Figure 9(a), edge and shading disparities are consistent and the percept is in between the results of D$^+$E$^+$ and D$^+$E$^-$ from Experiment 1. If the disparate ring appears on a non-disparately shaded ellipsoid, two different perceptions were reported. Especially for large disparities, some observers saw the ring floating in front of a rather flat surface. Others fused the edge-token and the surround into one coherent surface passing through the ring. This surface looked more transparent than those produced by the other cues and was also perceived as a cone-like *subjective surface* when a ring floated in front of a uniformly grey disk (Figure 9b).



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

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

Figure 8. Zero-disparity edge token overrides shading (Experiment 3). (a) Shape-from-shading ($n = 7$). (b) Shape-from-disparate-shading ($n = 6$). Only data for elongation 4.0 are shown.



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 (without disparity)
Subjects ISA and HAM: 63% Subject HHB: 41%

Figure 9. Surface interpolation for sparse edge data (Experiment 3). (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 grey disk yields a cone-like subjective surface ($n = 6$). (c), (d) Shape-from-shading plus disparate edge information leads to an ambiguous perception ($n = 3 + 3$). Only data for elongation 4.0 are shown.

Shape Comparison (Experiment 4)

All images with single cues lead to large errors in perceived shape. With *shading* and *texture* curvature is underestimated (Figure 10a, b), with a highlight it is overestimated (Figure 10c). 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 ellipsoid and does not level off as in the case of the depth probe measurements. A highlight on the shaded surface also seems to have a much larger influence with this technique and leads to an overestimation of curvature. But the most interesting result is the strong interaction between shading and texture as shown in Figure 10(d), (e). If shading and texture cues can be used simultaneously the perceived shape is almost veridical with a small bias towards under- or overestimation depending on the shading model [highlight absent (Figure 10d), or present (Figure 10e)].

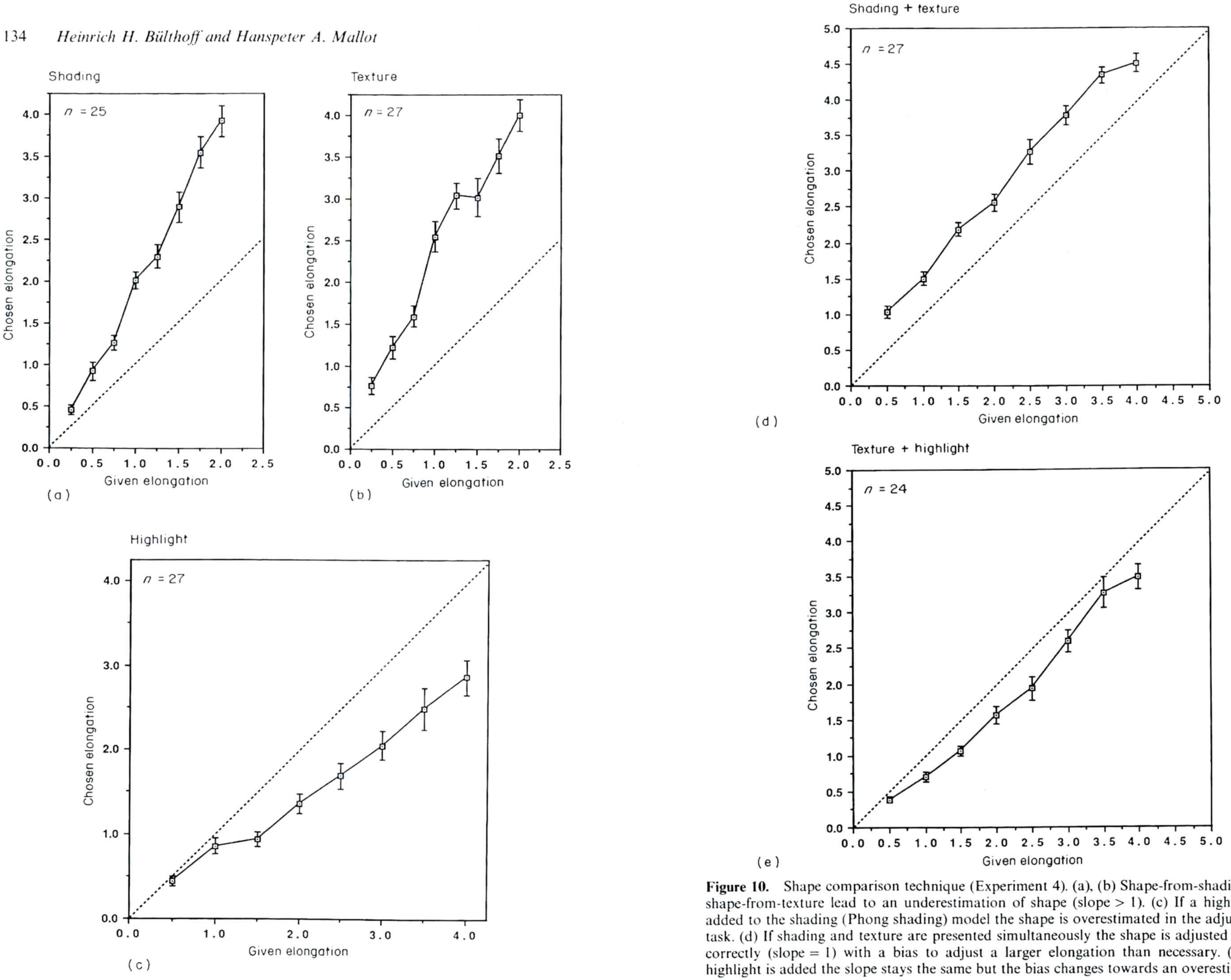


Figure 10. Shape comparison technique (Experiment 4). (a), (b) Shape-from-shading and shape-from-texture lead to an underestimation of shape (slope > 1). (c) If a highlight is added to the shading (Phong shading) model the shape is overestimated in the adjustment task. (d) If shading and texture are 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.

DISCUSSION

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 *et al.*, 1985) and exploited in its algorithms. On the other hand, *changes of image intensity* often convey information about physical edges in the scene. The location of sharp changes in image intensity correspond very often 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 (along the features) can be computed. This forces an additional stage in which sophisticated algorithms (Grimson, 1982; Blake & Zisserman, 1987) allow the interpolation of the surface between data points. In order to test for the ability of human stereo vision to get denser depth data by using in addition 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

$$z = \sqrt{1 - r^2}. \tag{4}$$

In the special case of a sphere, the surface normal simply equals the radius, i.e.

$$\mathbf{n} = (r \cos \varphi, r \sin \varphi, \sqrt{1 - r^2}). \tag{5}$$

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

$$\mathbf{I}(r) = I_0(\mathbf{l} \cdot \mathbf{n}) = I_0\sqrt{1 - r^2}, \tag{6}$$

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

$$\nabla^2 I(r) = I''(r) - \frac{1}{r} I'(r) = -I_0 \frac{r^2}{(1 - r^2)^{3/2}}. \tag{7}$$

This is a non-positive 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

$$I_c(r) = I_0 \frac{\sqrt{1 - r^2}}{\sqrt{1 - (1 - c^2)r^2}}, \tag{8}$$

which reduces to equation (6) for $c = 1$. In Figure 11(a), where luminance profiles are plotted for the elongations $c = 0.5$, 1.0, 2.0 and 4.0, it can be seen that for $c \geq 2$ the curves 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 the oblique illuminations used in Experiment 2, we found numerically that the self-shadow boundary corresponds to a level-crossing rather than a zero-crossing in the DOG-filtered image.

Independent from our own work, images of ellipsoids may be useful in the study of the psychophysical relevance of Laplacian zero-crossings. We feel that images of ellipsoids are superior to the gratings or filtered images often used for this purpose (Daugman, 1985).

Receptor Non-linearities in Early Vision

Since the visual system does not work directly on image intensities, but on spatially and temporally filtered and compressed (non-linear) signals, the effects of early visual processing in the retina have to be taken into account. Signal compression alone can significantly change image interpretation. Non-linearity in the photoreceptors, for example, can lead to an illusory motion perception for time-varying signals that do not entail motion information (Bülthoff & Götz, 1979). In analogy, these non-linearities could induce edge information that is not present in smooth-shaded images. An additional source of zero-crossings not present in our image arrays is the non-linearity of the color monitor. If arbitrary non-linearities are considered, zero-crossings can be induced in every non-constant image, however smooth (e.g. by discretization).

Retinal non-linearities in both vertebrates (Naka & Rushton, 1966; Hemilä, 1987) and invertebrates (Kramer, 1975) have been modeled by saturation-type characteristics of the form

$$f(I) = \frac{I}{I + I_{0.5}}, \tag{9}$$

where $I_{0.5}$ is a constant, given by the luminance which produces 50% of the maximal excitation. We repeated experiments D^+E^- and D^-E^-, i.e. those involving smooth-shaded images, compensating for the compression non-linearity with the inverse of equation (9). Since $I_{0.5}$ depends on the adaptation of the eye, four different choices of the constant $I_{0.5}$ were used. Monitor non-linearities were compensated as well. Depth perception from disparate shading was not affected significantly by this procedure.

Figure 11(b) shows the luminance profile for an ellipsoid with elongation 4.0, and the effect of the non-linearity of equation (9) for the tested values of $I_{0.5}$. It turns out that in our experiments, the presumed receptor non-linearities tend to cancel the shallow zero-crossings rather than to create new ones. This is further support for our assumption that edges cannot be extracted from the smooth-shaded images. Mechanisms relying on zero-crossings in the original image cannot account for the shape-from-disparate-shading performance found in our experiments. Under the assumption of compression-type non-linearities, this holds also for the first neural representation of the zero-crossing free images.

Shape-from-Disparate-Shading

The major finding of this study, as far as single depth cues are concerned, is the strength of depth perception (70%) 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 our 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, grey-level disparities appear to be more important than is usually appreciated.

Grimson (1984) makes explicit use of binocular shading differences for the interpolation of surfaces between good matches (i.e. between edges). Unfortunately, his model is not directly comparable to our study for the following reasons. First, the information that Grimson's algorithm recovers from shading is the surface orientation along zero-crossings. In our experiments with smooth ellipsoids, the only zero-crossing contour is the occluding contour of the object where the surface orientation does not depend on the total elongation of the object; it is always perpendicular to the image plane. Second, Grimson's model requires a specular component in the reflectance function of the object. Quite to the contrary, we did not find significant differences between Lambertian or Phong shading. From this we may conclude that a mechanism different from the one proposed by Grimson is involved.

Shape-from-Disparate-Shading: Is it Localized or Distributed?

Are there features other than zero-crossings which can account for the shape-from-disparate-shading performance found in our experiments? Possible candidates

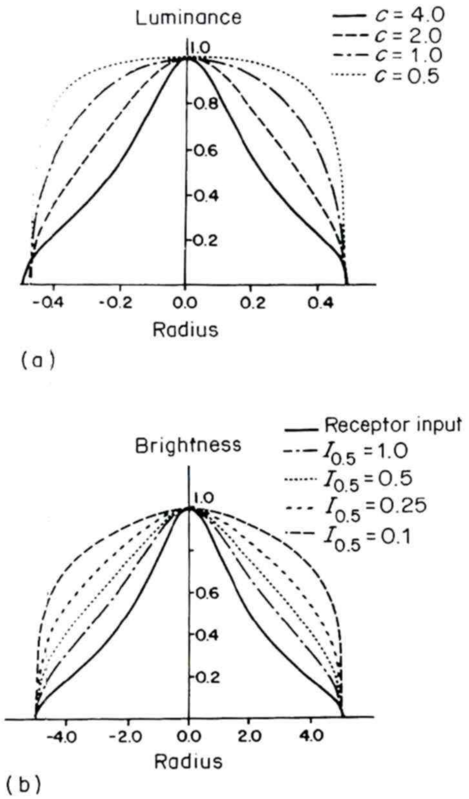


Figure 11. Luminance and simulated brightness profiles. (a) Luminance profiles of ellipsoids with different elongations. Note that for elongations larger than 2.0 inflections occur. (b) Brightness profiles for the ellipsoid with elongation 4.0 (the one with the pronounced inflections in Figure 11a). The non-linear compression (Equation 9) tends to cancel the inflections which might give rise to zero-crossings, rather than to enhance them.

include the intensity peak as proposed by Mayhew & Frisby (1981) and level-crossings in the DOG-filtered image which, according to Hildreth (1983), might account for Mayhew & Frisby's (1981) 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 non-disparate background (Figure 9). 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 Figure 9(a) with Figures 9(b)–(d)]. For small elongations (0.5–2.0; not shown in Figure 9) 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 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.

Surface Interpolation and Subjective Surfaces

In the experiments with sparse edge information (Figure 9b–d), an interpolated surface was measured directly with the depth probe technique. If the depth separation between the ring and the shaded ellipsoid was large (elongation 4.0) an ambiguous perception was experienced. One interpretation consisted of a solid base at about the depth perceived from shape-from-shading alone with the ring floating in front of it (Figure 9d). The other interpretation was a transparent subjective surface onto which the ring was drawn (Figure 9c). In this case, the depth of the entire surface was pulled towards the ring. Surprisingly, a subjective surface could also be perceived when the token was floating in front of a uniformly grey disk (Figure 9b). An interaction between shape-from-shading and edge-based stereo is therefore not necessary to perceive subjective surfaces.

Shape-from-Shading: Algorithms and Psychophysics

A computational theory for shape-from-shading is presented by Ikeuchi & Horn (1981). As an example, they discuss the image of a sphere with Lambertian reflectance function, illuminated by parallel light from the viewing direction. This example can be directly compared to our Experiment 1 (D^-E^-) where about 25% of the correct depth was perceived by the observers. Interestingly, the algorithm of Ikeuchi & Horn underestimates depth if the input data are noisy. The distortion of shape in their algorithm depends on a regularization parameter λ. For a large value of λ, which would be appropriate for noisy image data, the smoothing of the surface leads to a considerable underestimation of depth On the other hand, the iterative scheme becomes unstable if the value of λ is reduced too much. For an approach which avoids smoothing by a regularization term, see Horn and Brooks (1985).

The algorithm of Ikeuchi & Horn also shows other types of errors when the light source position and the reflectance properties of the surface are not known exactly. The types of errors reported from numerical experiments are asymmetric distortions for false assumptions of the light source position and overestimation of depth when false reflectance functions are assumed. In our psychophysical studies, these errors did not occur. Asymmetric deformations as reported by Ikeuchi & Horn are not present even for the obliquely illuminated objects (Figure 7). Whether this corresponds to a correct judgement of the illuminant direction by the human observer is currently under investigation. Also, varying the reflectance function did not change the shape-from-shading performance as measured with our depth probe technique in Experiment 1 (Figure 6, dashed lines).

How Useful is Shading as a Cue for Depth?

Todd & Mingolla (1983) and Mingolla & Todd (1986) used psychophysical techniques to investigate how observers analyze shape by use of shading cues. According to their results, the human observer underestimates surface curvature by over 50% when using shading information. A similar result has been reported by Barrow & Tenenbaum (1978), showing that shading of a cylindrical surface can deviate substantially from natural shading before a change in the perceived shape can be detected. This is well in line with our psychophysical findings which suggest that non-disparate shading is a poor cue to depth. It is, however, in contrast to the intuition of artists who use shading as a primary tool to depict objects in depth.

Is it possible that we are not asking the right question when we try to analyze shape with the local depth probe? Obviously everybody can describe the shape of a vase in a photograph even without any texture on it. In principle, the information that can be obtained from shape-from-shading is surface orientation rather than absolute depth. However, as Todd and Mingolla have shown, the surface normal on simply shaded bodies is difficult to estimate in psychophysical experiments, and even after a training phase subjects make a lot of errors. A precise measurement of surface slant and tilt does not seem to be necessary for shape perception.

In the study reported here, we tried to measure the perceived depth directly with a stereoscopically viewed depth probe. This seems to be a much simpler task for the subjects, and indeed we did not need a long training phase to obtain consistent depth measurements. It is not obvious that this method worked for shading cues alone, since it involves a cross-comparison of supposedly more or less independent modules as well as a comparison of local (depth probe) and global (shading) information. Since our depth probe is defined by stereopsis it requires binocular viewing even for non-disparate images (pure shape-from-shading). To avoid this, we developed a paradigm to measure shape-from-shading monocularly (Bülthoff & Mallot, 1988). With this paradigm we can also analyze other cues, e.g. texture gradients and occluding contours, which would show similar problems with a local stereo depth probe.

Integration of Depth Modules

Concrete predictions of the types of interactions that should occur between different depth cues are still difficult to obtain from computational theory. Therefore, we hope that psychophysical studies will in turn provide useful hints for computational investigations into the integration of depth information.

Accumulation is a simple type of interaction that can be implemented in a number of different ways. Depth information can be collected from different cues and performance should improve as more information becomes available. Our data show that it is not the reliability which improves, but the perceived depth which increases. This result hints at regularization as the mechanism for the observed accumulation. Given a stereo contour surrounding a surface patch, the most conservative estimate would be a smooth interpolation as performed by computer vision algorithms (Grimson, 1982; Marroquin, 1984; Terzopoulos, 1986).

In our stimuli, the smoothest interpolation is a flat disk. In a tradeoff with the smoothness constraint, the visual system seems to use the available information to the extent of its reliability. This might explain why depth perception increases as more information becomes available.

Conflicting cues were presented in Experiments 1 and 3. Whenever visible, edge-based disparities were decisive for the perceived depth (Figure 5, D^-E^+, Figures 8 and 9). Except for the subjective surface (see the section on "Surface interpolation and subjective surfaces") the "veto effect" was restricted to a vicinity of the edge as can be seen from the sparse edge data in Figures 8 and 9. Edge-based stereo thus overrides both shape-from-shading and shape-from-disparate-shading. It is possible, however, that this veto relationship occurs only in the locally derived depth map. The global percept of the polyhedral ellipsoid in Experiment 1 (D^-E^+) is not flat but convex. A conflicting cue combination of shape-from-shading and shape-from-disparate-shading was presented in the experiment with smooth-shaded non-disparate 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%. An inhibitory interaction between the two cues may account for this poor shape-from-shading performance and the ceiling effect in Figure 6.

Asymmetric types of interaction, such as veto or disambiguation, can be expected from models of surface interpolation that start with reliable but sparse depth information typically obtained from disparate edges or occluding contours. Interpolation between the sites of the edges could rely on a smoothness constraint (Grimson, 1982) or on additional cues such as shading (Ikeuchi & Horn, 1981; Blake *et al.*, 1985) and binocular shading of specular surfaces (Grimson, 1984). Its distributed mechanism and the veto relationship to edge-based stereo make shape-from-disparate-shading especially suitable for surface interpolation in human vision. The interactions of different depth cues (as derived from our depth probe experiments) in consistent and contradictory cases are summarized in Figure 12. Another summary of our data which includes both depth probe and shape comparison techniques is shown in Figure 13. This representation is based on the idea (sketched in Figure 2) 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%). It is even stronger for the perception of shape (100%). 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 overestimate of shape while dull surfaces (Lambertain shading) were judged to be flat.

Recently, Poggio (1985) proposed a new formalism for the integration of different vision modules, based on a probabilistic approach (Marroquin, 1984; Marroquin *et al.*, 1987). The advantage of this coupled Markov random fields

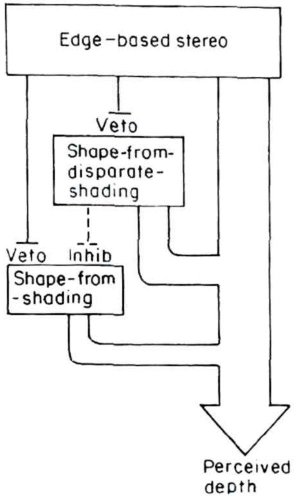


Figure 12. Integration of depth cues. The size of the boxes and interaction channels reflects the contribution of the different depth cues for the overall perceived depth (accumulation). In contradictory cases, shape from both disparate and non-disparate shading is vetoed by edge-based stereo. An inhibitory influence of shape-from-disparate-shading on shape-from-shading is discussed in "Integration of depth modules".

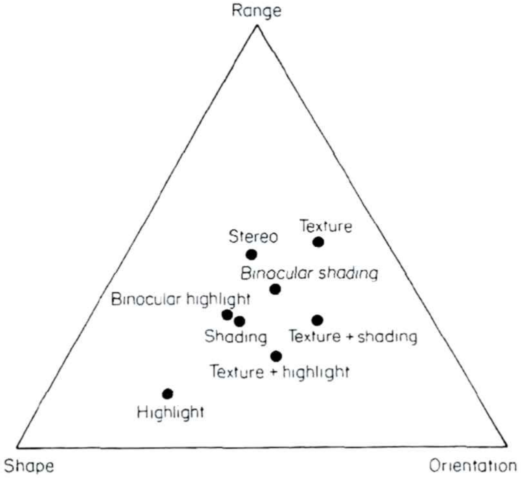


Figure 13. Depth triangle. This representation of our depth probe and shape comparison data shows the relative importance of depth cues (stereo, shading, texture) for different 3D-descriptors (range, shape, orientation); see also Figure 2. Shading has a stronger influence on the perceived shape, while texture seems to be more important for orientation (compare with Figure 1). 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.

approach over regularization theory lies in the possibility of simultaneous segmentation and (piecewise) smoothing of the image. As far as the experiments discussed here are concerned, the results should not be significantly different from those of regularization. However, if other cues such as occlusion are considered, more complex types of interaction are to be expected from the coupled Markov random field approach.

ACKNOWLEDGEMENTS

We are grateful to A. Blake, E. Hildreth, D. Kersten, B. Saxberg and T. Poggio for helpful discussions and I. Bülthoff for going through long and tedious experimental sessions.

This report describes research done within the Center for Biological Information Processing (Whitaker College) at the Massachusetts Institute of Technology. Support for this work is provided by a grant from the Office of Naval Research, Engineering Psychology Division and a NATO Collaborative Grant No. 0403/87. One of us (HAM) was supported by a grant from the Deutsche Forschungsgemeinschaft (Ma 1038/1-1,2).

REFERENCES

Ardity, A. (1986). Review of "Binocular Vision". In K. R. Boff, L. Kaufman & J. P. Thomas (Eds) *Handbook of Perception and Human Performance*, Vol. I, *Sensory Processes and Perception*, Chapter 23. John Wiley, New York.

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

Barrow, H. G. & Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds) *Computer Vision Systems*, pp. 3–26. Academic Press, New York.

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

Blake, A. & Bülthoff, H. H. (1989). Does the brain know the physics of specular reflection? *Nature* (in press).

Blake, A. & Zisserman, A. (1987). *Visual Reconstruction*. MIT Press, Cambridge, Mass.

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

Braunstein, M. L., Andersen, G. J., Rouse, M. W. & Tittle, J. S. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception and Psychophysics*, **40**, 216–224.

Bülthoff, H. H. & Götz, K. G. (1979). Analogous motion illusion in man and fly. *Nature*, **278**, 636–638.

Bülthoff, H. H. & Mallot, H. A. (1988). Integration of depth modules: local and global depth measurements. *Investigative Ophthalmology and Visual Science*, **29** (Suppl.), 400.

Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, **2**, 1160–1169.

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.

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

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

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

Grimson, W. E. L. (1984). Binocular shading and visual surface reconstruction. *Computer Vision Graphics and Image Processing*, **28**, 19–43.

Grzywacz, N. M. & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: position-based versus velocity-based formulations. *Journal of the Optical Society of America*, **A4**, 503–518.

Hemilä, S. (1987). The stimulus–response functions of visual systems. *Vision Research*, **27**, 1253–1261.

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

Horn, B. K. P. & Brooks, M. J. (1985). The variational approach to shape from shading, MIT Artificial Intelligence Laboratory Memo, No. 813, pp. 1–32.

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*. University of Chicago Press, Chicago.

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

Kramer, L. (1975). Interpretation of invertebrate photoreceptor potentials in terms of a quantitative model. *Biophysics of Structure and Mechanism*, **1**, 239–257.

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

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

Marr, D. & Nishihara, H. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society (London) Series B*, **200**, 269–294.

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

Marroquin, J. L. (1984). Surface reconstruction preserving discontinuities. MIT Artificial Intelligence Laboratory Memo, No. 792.

Marroquin, J. L., Mitter, S. K. & Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, **82**, 76–89.

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

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

Naka, K. I. & Rushton, W. A. H. (1966). S-potentials from color units in the retina of fish (*Cyprinidae*). *Journal of Physiology (London)*, **185**, 536–555.

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

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 Reviews of Neuroscience*, **7**, 379–412.

Poggio, T. (1985). Integrating vision modules with coupled MRFs. MIT Artificial Intelligence Laboratory Working Paper, No. 285.

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

Stevens, K. A. (1981). The visual interpretation of surface contours. *Artificial Intelligence*, **17**, 17–45.

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

Terzopoulos, D. (1986). Integrating visual information from multiple sources. In: A. P. Pentland (Ed.) *From Pixels to Predicates*, pp. 111–142. Ablex, Norwood, NJ.

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*. MIT Press, Cambridge MA.

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