Integration of depth modules: stereo and shading

Heinrich H. Bülthoff and Hanspeter A. Mallot*

Center for Biological Information Processing, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, E25-201, Cambridge, Massachusetts 02139

Received November 9, 1987; accepted June 3, 1988

We studied the integration of image disparities, edge information, and shading in the three-dimensional perception of complex yet well-controlled images generated with a computer-graphics system. The images showed end-on views of flat- and smooth-shaded ellipsoids, i.e., images with and without intensity discontinuities (edges). A map of perceived depth was measured by adjusting a small stereo depth probe interactively to the perceived surface. Our data show that disparate shading (even in the absence of disparate edges) yields a vivid stereoscopic depth perception. The perceived depth is significantly reduced if the disparities are completely removed (shape-from-shading). If edge information is available, it overrides both shape-from-shading and disparate shading. Degradations of depth perception corresponded to a reduced depth rather than to an increased scatter in the depth measurement. The results are compared with computer-vision algorithms for both single cues and their integration for three-dimensional vision.

1. INTRODUCTION

Computer vision has had a long-standing concern in the human visual system as an existence proof for high-performance vision. Conversely, human psychophysics has obtained from computer vision new insights into the computational structure of problems with which any visual system has to deal. In fact, this computational theory is much the same for both human and machine vision. It is best elaborated in the field of depth perception, which deals with the ill-posed problem of reconstructing a three-dimensional scene from two-dimensional images.

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[1-3] and shading.[4-7] From a more computational point of view, there are also a number of studies on depth from texture,[8-11] line drawings,[12] surface contours,[13,14] and structure from motion.[15-17] These implementations are quite successful with synthetic images but less reliable with natural images. However, 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 a method for a quantitative measurement of depth perception with complex yet well-controlled images.

A. Integration of Multiple Depth Cues

In principle, there are several types of useful interaction between depth cues that are not mutually exclusive:

- *Accumulation*: Information from different cues could be accumulated in different ways, such as probability summation and the linear summation model for the integration of stereo and proximity luminance covariance proposed by Dosher *et al.*[18] A more computational approach to accumulation is the notion of joint regularization, in which constraints from different cues are accounted for by means of a common cost function.[19,20]
- *Cooperation*: Especially in the case of poor or noisy cues, modules might work synergistically. Here we think of the nonlinear interactions of different cues, which can be treated with the coupled Markov random field approach.[21]
- *Disambiguation*: A particular case of a nonlinear interaction is the case in which information from one cue is used to disambiguate locally a representation derived from another one (e.g., stereo can disambiguate shading[22]).
- *Veto*: There can be unequivocal information from one cue that should not be challenged by others.

B. Psychophysical Depth Measurements

The perception of a three-dimensional scene 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. In our experiments we mapped perceived depth 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 with 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. In ongoing work[23] we are investigating the contribution of depth cues to global percepts such as the orientation and shape of objects.

In this paper we study the relation of shading (with and without highlights) and stereo disparities in the three-dimensional perception of smooth and polyhedral surfaces. For polyhedral objects, disparities were associated with localized features, i.e., the intensity changes at the facet boundaries, while for the smooth surfaces only shading disparities occurred. In addition, contours such as rings and lines could be drawn on the smooth surfaces to provide

0740-3232/88/101749-10$02.00

sparse edge information. The objects were ellipsoids of revolution viewed end-on; they were chosen for the following reasons:

- As we show in Subsection 4.A below, images of Lambertian-shaded smooth ellipsoids with moderate eccentricities do not contain Laplacian zero crossings when they are illuminated centrally with parallel light.
- Since the surfaces are closed, they are naturally outlined by an occluding contour. This contour was placed in the zero-disparity plane and did not provide any depth information.
- Convex objects such as ellipsoids do not generate cast shadows or reflections on their own surfaces. Therefore shading (attached shadows) could be studied without interference from cast shadows.
- End-on views of ellipsoids can be thought of as a model for the depth interpolation of a surface patch between sparse edge data.

2. METHODS

A. Computer Graphic Psychophysics

Images of smooth-shaded ellipsoids and flat-shaded polyhedral ellipsoidal objects were generated either by ray-tracing techniques or with a solid modeling software package (S-Geometry, Symbolics Inc.). The smooth objects were ellipsoids of revolution, the axis of revolution being perpendicular to the display screen, i.e., the objects were viewed end-on. Simple contours (rings) could be mapped onto the surface. The polyhedral objects were derived from quadrangular tesselations of the sphere along meridian and latitude circles. The objects were elongated along the axis in the equatorial plane perpendicular to the display screen. Thus the two types of object differed mainly in the absence or presence of edges. 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 what follows, all semi-diameters (elongations) are given as multiples of 6.7 cm.

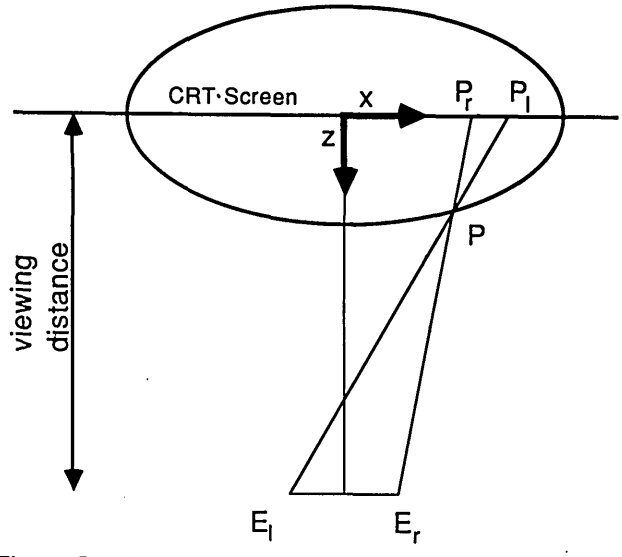

Fig. 1. Imaging geometry. Projection onto the x–z plane. The viewing distance is 120 cm. E_l, E_r: nodal points of the left and right eyes, respectively. The distance between E_l and E_r is 6.5 cm. A point $\mathbf{p} \in \mathbf{R}^3$ is imaged on the screen at P_l for the view from the left eye and at P_r for the view from the right eye.

The imaging geometry used in the computations is shown in Fig. 1. It differs from the usual camera geometry in that the image is constructed on a screen that 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 arcmin (cf. Ref. 24) corresponds to an interval from 4.3 to 15.2 cm in front of the screen.

For the computation of the smooth-shaded ellipsoids, a ray-tracing operation was performed.[25] The illuminant was modeled as an infinite point source (parallel illumination) centrally behind the observer. For a number of control experiments, oblique directions of illumination (upper left and lower right) were used. Surface shading was computed according to the Phong model,[26] which consists of an ambient, a diffuse (Lambertian), and a specular component. For Lambertian shading, the ambient and specular components were zero, while for specular shading (which produces highlights), 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.

Experiment 1: Disparity and Edge Information

In a first series of experiments, we crossed disparity and dense edge information in shaded images. Four different image types were tested [Figs. 2(a) and 2(b)]:

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

Experiment 2: Illuminant Direction

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 now (±14° azimuth and ∓13.6° elevation from the viewing direction).

Experiment 3: Edge versus Shading Disparity

The last 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

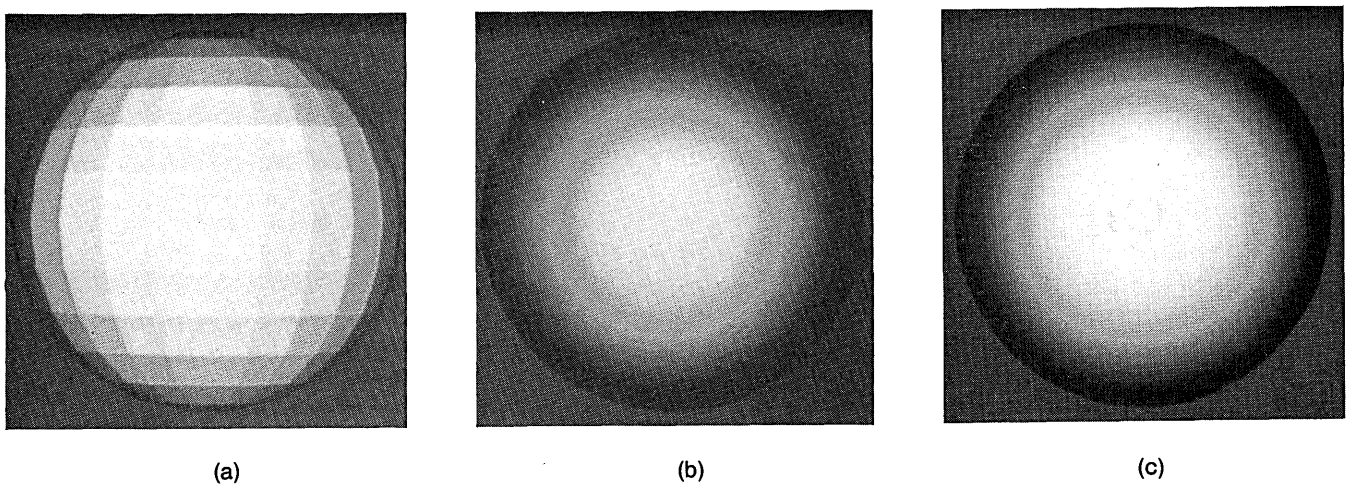

Fig. 2. Flat- and smooth-shaded surfaces. (a), (b) Discontinuous and smooth intensity variations in images of polyhedra and ellipsoids provide cues for edge-based stereo, shape-from-disparate-shading, and shape-from-shading (experiment 1). (c) A smooth ellipsoid with sparse edge information has been 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.

image). Disparities of shading and ring were varied independently, leading to the following combinations [Fig. 2(c)]:

- Disparate ring and disparate shading.
- Disparate ring and nondisparate shading.
- Nondisparate ring and disparate shading.
- Nondisparate ring and nondisparate shading.
- Disparate ring in front of a uniformly gray nondisparate disk.

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

B. Experimental Procedure

We displayed either a pair of disparate images (stereo pair) or a nondisparate view of the object as seen from between the two eyes on a cathode-ray tube color monitor [Mitsubishi UC-6912 high-resolution color-display monitor, resolution ($H \times V$) 1024 × 874 pixels; bandwidth ±3dB between 50 Hz and 50 MHz, short-persistence phosphor]. The disparate images were interlaced (even lines for the left-hand image and odd lines for the right-hand image) with a frame rate of 30 Hz. Both disparate and nondisparate images were viewed binocularly through shutter glasses (Stereo-Optic Systems, Inc.), which were triggered by the interlace signal to present only the appropriate images to the left and right eyes. The objects were shown in black and white with a true resolution of 254 gray levels by using a 10-bit digital-to-analog converter. The background was uniformly colored in half-saturated blue. The screen was viewed in complete darkness.

Perceived depth was measured by adjusting a small red square-shaped (4 × 4 pixel) depth probe to the surface interactively (with the computer mouse). This probe was displayed in an interlaced mode together with the disparate images. This is a computer graphics version of a binocular rangefinder developed by Gregory,[27] 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 (Fig. 3). Subjects included the authors (corrected vision) and one naïve observer, all with normal stereo vision as tested with natural and random-dot stereograms.

C. Data Evaluation

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 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 with the first principal component, as a measure of perceived depth for a given cue combination (Fig. 4).

3. RESULTS

A. Experiment 1: Disparity and Edge Information

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 for all three subjects and were pooled since the differences were noticeable only in the standard deviation. The 16 plots of Fig. 3 show the averaged results of all subjects for the four types of experiment and four different elongations of Lambertian-shaded ellipsoids.

The perceived elongation in the images with consistent

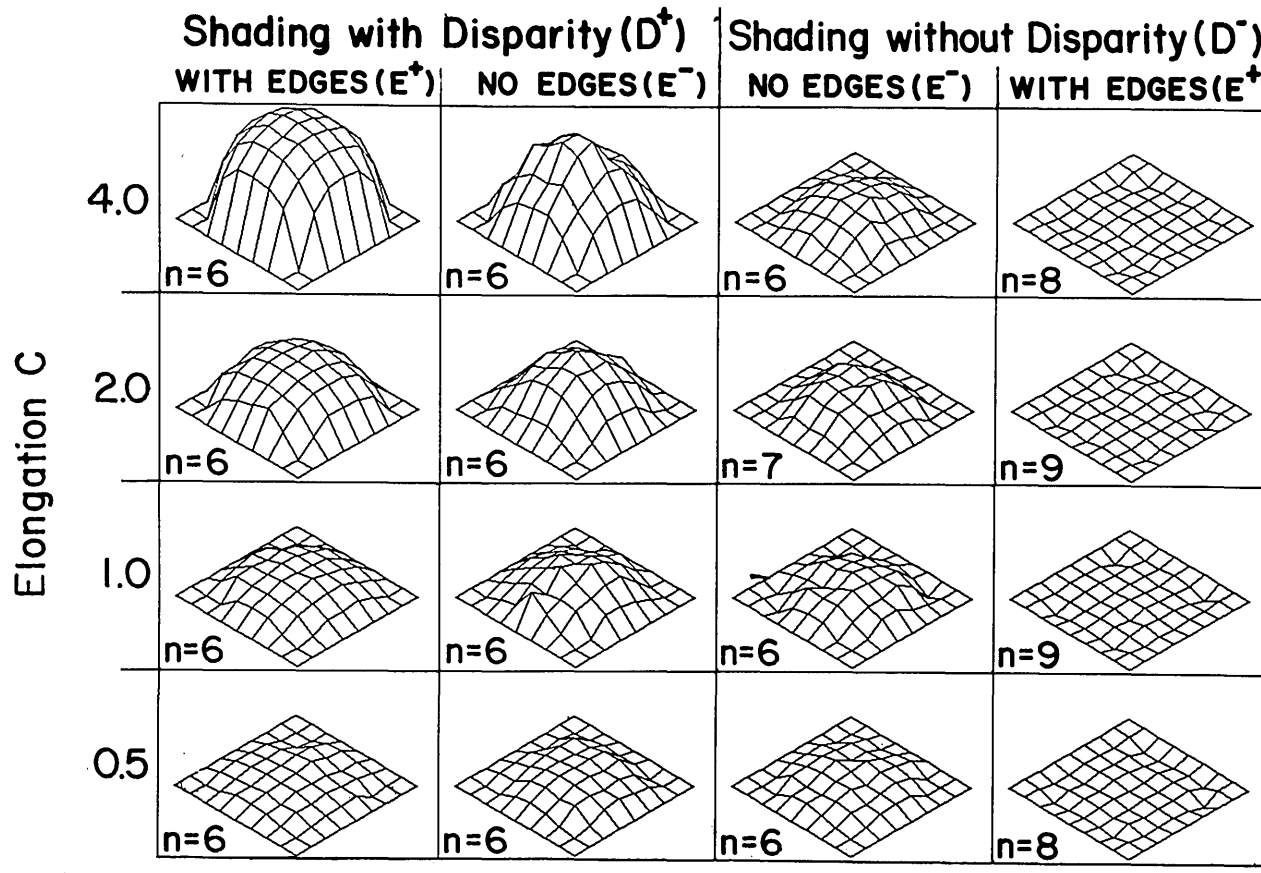

Fig. 3. Perceived surfaces (experiment 1; depth not drawn to scale). Each plot shows the average of 6–9 sessions from three subjects. Perceived depth decreases with the following sequence of cue combinations: disparity, edges, and shading (D^+E^+); disparity and shading but no edges (D^+E^-); shading only (D^-E^-); contradictory disparity and shading (D^-E^+).

cue combinations depends on the amount of information available. As can be seen from Fig. 4, the perceived elongation is almost correct when shading and 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 from the displayed elongation. Phong shading (highlights) used instead of Lambertian shading did not change perceived depth significantly.

In experiment D^-E^+, two identical images (no disparity) of polyhedral ellipsoids (edges) were shown. Although shading alone provided some depth information, as shown for 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 is still available. A comparison of the results (Fig. 4) 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.

B. Experiment 2: Illuminant Direction

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 at the upper left and the lower right in front of the object (±14° azimuth and ∓13.6° elevation from the viewing direction). The results of these experiments (41 measurements, data pooled from all subjects) are depicted in Fig. 5. The slight

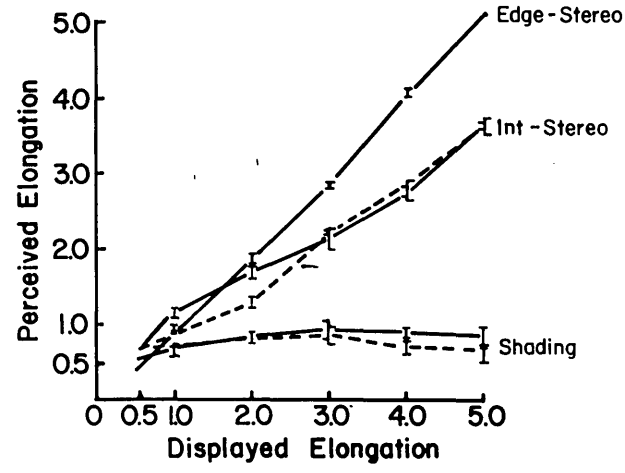

Fig. 4. Perceived elongation. Depth perception decreases as fewer cues are available. The significant separation between the middle and lower curves (smooth shading with and without any disparity) illustrates the influence of disparity information even in the absence of edges. Solid lines, Lambertian shading; dashed lines, Phong shading. Int, intensity.

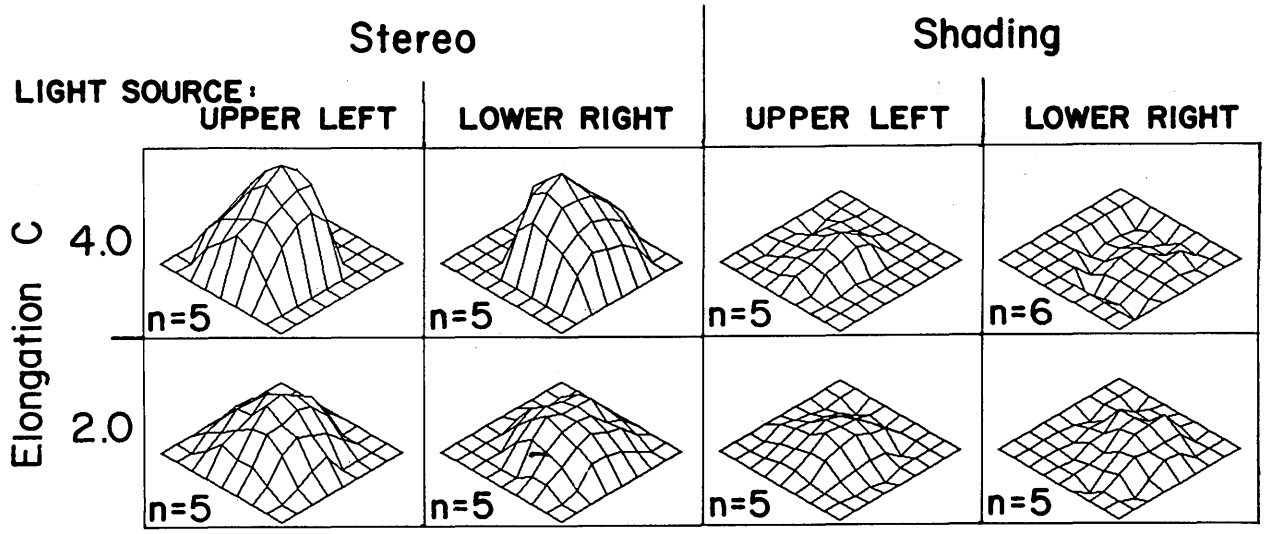

Fig. 5. Perceived surfaces for oblique illuminations (experiment 2). The data confirm the relevance of disparate shading and show the independence of the findings of experiment 1 from the lighting conditions. No depth was measured in the self-shadow regions.

asymmetries present at elongation 4.0 result 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 nondisparate shading (shape-from-shading). Furthermore, when illumination is from the lower right, stereo prevents depth inversion, which occasionally occurred in the nondisparate images (cf. the negative depth in Fig. 5; shading, lower right).

C. Experiment 3: Edge versus Shading Disparity

In contrast to the original measurements with polyhedral objects in which 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 4 different elongations. Figure 6 shows the results for the ring at zero disparity combined with nondisparate (top) and disparate (bottom) shading. 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 Fig. 7: In Fig. 7(a), edge and shading disparities are consistent, and the percept is between the results of D^+E^+ and D^+E^- from experiment 1. If the disparate ring appeared on a nondisparately 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 nearly transparent than those produced by the other cues and was also perceived as a conelike *subjective surface* when a ring floated in front of a uniformly gray disk [Fig. 7(b)].

4. DISCUSSION

A. Images without Zero Crossings

One of the most important constraints in early vision for recovering surface properties is that the physical process underlying image formation are typically smooth. The smoothness property is captured well by standard regularization[19] 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 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.[28]

The disadvantage of these feature-based stereo algorithms is that only sparse depth data (along the features) can be computed. 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 a Lambertian reflection function and parallel illumination, zero crossings in the Laplacian are missing.

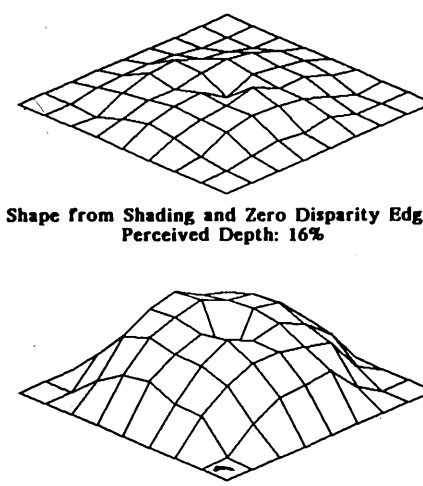

Fig. 6. Zero-disparity edge token overrides shading (experiment 3). Top, Shape-from-shading (n = 7). (Bottom) Shape-from-disparate-shading (n = 6). Only data for elongation 4.0 are shown.

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 is radial, i.e.,

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

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

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

which reduces to Eq. (3) for $c = 1$. In Fig. 8(a), where luminance profiles are plotted for the elongations $c = 0.5$, 1.0, 2.0, 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 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 rather than to a zero crossing in the image filtered with the difference of two Gaussians.

Independently from our own work, images of ellipsoids may be useful in the study of the psychophysical relevance of Laplacian zero crossings. We believe that images of ellipsoids are superior to the gratings or filtered images often used for this purpose.[29]

B. Receptor Nonlinearities in Early Vision

Since the visual system does not work directly on image intensities but on spatially and temporally filtered and compressed (nonlinear) signals, the effects of early visual processing in the retina have to be taken into account. Signal compression alone can significantly change image interpretation. Nonlinearity in the photoreceptors, for example, can lead to an illusory motion perception for time-varying signals that do not entail motion information.[30] In analogy, these nonlinearities 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 nonlinearity of the color monitor. If arbitrary nonlinearities are considered, zero crossings can be induced in every nonconstant image, however smooth (e.g., by discretization).

Retinal nonlinearities in both vertebrates[31,32] and invertebrates[33] have been modeled by saturation-type characteristics of the form

$$f(I) = \frac{I}{I + I_{0.5}}, \tag{6}$$

where $I_{0.5}$ is a constant, given by the luminance that produces 50% of the maximal excitation. Since $I_{0.5}$ depends on the adaptation of the eye, we repeated experiments D^+E^- and D^-E^-, i.e., those involving smooth-shaded images, compensating for the compression nonlinearity with the inverse of Eq. (6) and four different choices of the constant $I_{0.5}$.

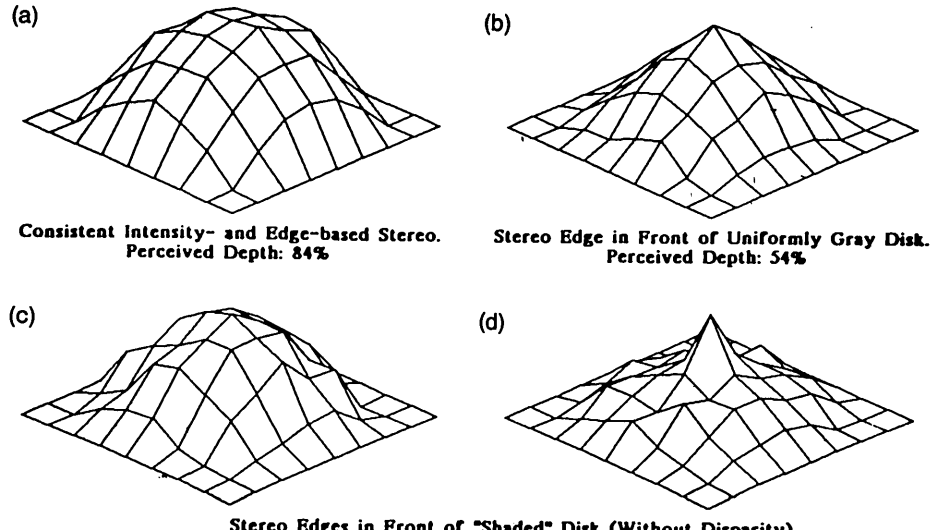

Fig. 7. Surface interpolation for sparse edge data (experiment 3). (a) Shape-from-disparate-shading plus disparate edge information leads to an almost correct percept ($n = 6$). (b) A single edge token in front of a uniformly gray disk yields a conelike subjective surface ($n = 6$). (c), (d) Shape-from-shading plus disparate edge information leads to an ambiguous perception ($n = 3 + 3$). Only data for elongation 4.0 are shown.

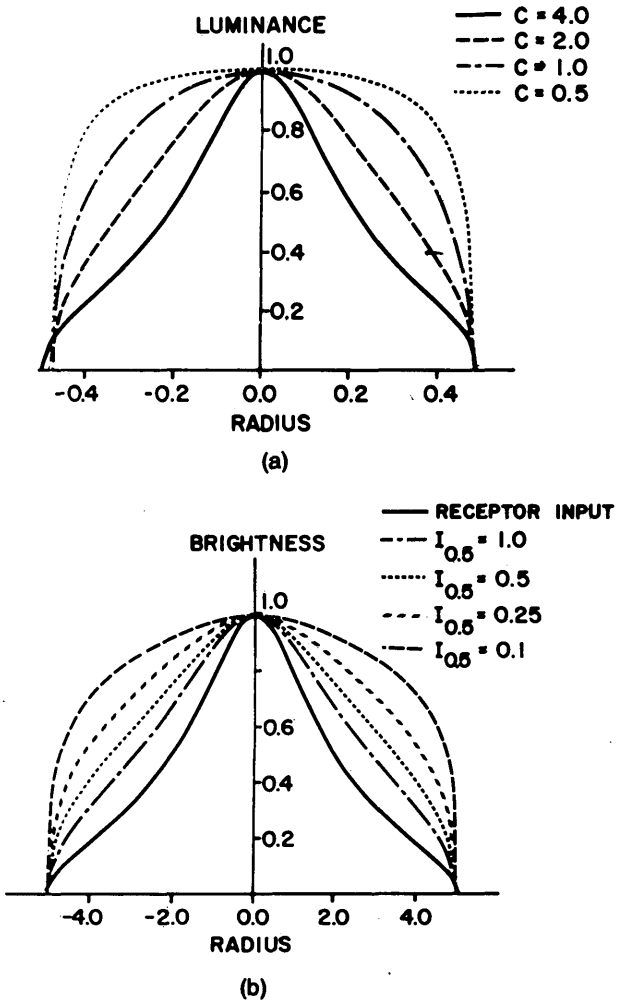

Fig. 8. Luminance and simulated brightness profiles. (a) Luminance profiles of ellipsoids with different elongations. Note that for elongations larger than 2.0, inflections occur. (b) Brightness profiles for the ellipsoid with elongation 4.0 [the one with the pronounced inflections in (a)]. The nonlinear compression [Eq. (6)] tends to cancel the infections that might give rise to zero crossings rather than enhancing them.

Monitor nonlinearities were compensated for as well. Depth perception from disparate shading was not affected significantly by this procedure.

Figure 8(b) shows the luminance profile for an ellipsoid with elongation 4.0 and the effect of the nonlinearity of Eq. (6) for the tested values of $I_{0.5}$. It turns out that in our experiments the presumed receptor nonlinearities 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 nonlinearities this also holds for the first neural representation of the zero-crossing-free images.

C. 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 a review, see Ref. 34). 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 intensity-based stereo. However, in the absence of edges and for surface interpolation, gray-level disparities appear to be more important than is usually appreciated.

Grimson[35] 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 with 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 and Phong shading. From this we may conclude that a mechanism different from the one proposed by Grimson is involved.

D. Shape from Disparate Shading: Is It Localized or Distributed?

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[3] and level crossings in the image filtered by the difference of two Gaussians, which, according to Hildreth,[36] might account for the data of Mayhew and Frisby 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 (Fig. 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 Fig. 7(a) with Figs. 7(b)–7(d)]. For small elongations (0.5–2.0; not shown in Fig. 7) the differences were not pronounced. We therefore conjecture that shape-from-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 (Fig. 6).

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 both in the data and in neural processing. Even in the absence of image noise, intensity-based algorithms (e.g., Ref. 37 can lead to severe matching errors when run on our stimuli.

E. Surface Interpolation and Subjective Surfaces

In the experiments with sparse edge information [Figs. 7(b)–7(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 [Fig. 7(d)]. The other interpretation was a transparent subjective surface onto which the ring was drawn [Fig. 7(c)]. In this case the depth of the entire surface was pulled toward the ring. Surprisingly, a subjective surface could also be perceived when the token was floating in front of a uniformly gray disk [Fig. 7(b)]. An interaction between shape-from-shading and edge-based stereo is therefore not necessary for perception of subjective surfaces.

F. Shape-from-Shading: Algorithms and Psychophysics

A computational theory for shape-from-shading is presented by Ikeuchi and Horn.[6] As an example, they discuss the image of a sphere with a Lambertian reflectance function, illuminated by parallel light from the viewing direction. This example can be directly compared with our experiment 1 (D^-E^-), in which about 25% of the correct depth was perceived by the observers. Interestingly, the algorithm of Ikeuchi and 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 that avoids smoothing by a regularization term, see Ref. 38.

The algorithm of Ikeuchi and Horn also shows other types of error when the light-source position and the reflectance properties of the surface are not known exactly. The types of error 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 and Horn are not present even for the obliquely illuminated objects (Fig. 5). Whether this is due to a correct judgment 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 (Fig. 4, dashed lines).

G. How Useful is Shading as a Cue for Depth?

Mingolla and Todd[4,39] 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 more than 50% when using shading information. A similar result has been reported by Barrow and Tenenbaum,[40] 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 nondisparate 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 Mingolla and Todd have shown, the surface normal on simply shaded bodies is difficult to estimate in psychophysical experiments, and even after a training phase subjects make many 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. Therefore our depth probe requires binocular viewing even for nondisparate images (pure shape-from-shading). To avoid this we are currently developing a paradigm to measure shape-from-shading monocularly. With this paradigm we will be able to analyze other cues also, e.g., texture gradients and occluding contours, which would show similar problems with a local stereo depth probe.

H. Integration of Depth Modules

Concrete predictions of the types of interaction that should occur among 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 that improves but the perceived depth that increases. This result hints toward 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.[20,41,42] In our stimuli, the smoothest interpolation is a flat disk. In a trade-off 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 edge-based disparities were visible, they were decisive for the perceived depth (Fig. 3, D^-E^+; Figs. 6 and 7). Except for the subjective surface (Subsection 4.E) the veto effect was restricted to a vicinity of the edge, as can be seen from the sparse edge data in Figs. 6 and 7. Edge-based stereo thus overrides both shape-from-shading and shape-from-disparate-shading. Note, however, that this veto relationship might occur 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 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%. An inhibitory interaction between the two cues may account for this poor shape-from-shading performance and the ceiling effect in Fig. 4.

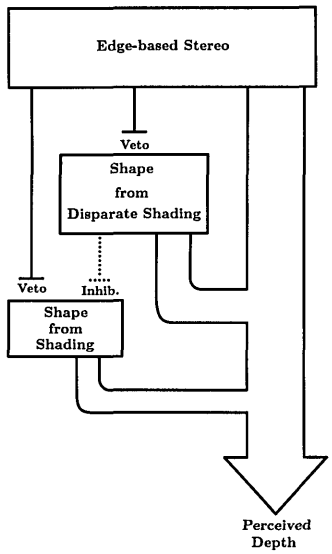

Fig. 9. Integration of depth cues. The sizes of the boxes and interaction channels reflect the contributions of the different depth cues for the overall perceived depth (accumulation). In contradictory cases, shape from both disparate and nondisparate shading is vetoed by edge-based stereo. An inhibitory influence of shape-from-disparate-shading on shape-from-shading is discussed in Subsection 4.H.

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[41] or on additional cues such as shading[5,6] and binocular shading of specular surfaces.[35] 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 in consistent and contradictory cases are summarized in Fig. 9.

Recently Poggio[43] proposed a new formalism for the integration of different vision modules, based on a probabilistic approach.[21,42] The advantage of this coupled Markov random-fields 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.

ACKNOWLEDGMENTS

We are grateful to E. Hildreth, T. Poggio, B. Saxberg, and R. Wildes for helpful discussions, I. Bülthoff for being a subject, and J. Little for assistance on the LISP computer. Comments by two anonymous reviewers were most valuable.

This paper describes research done within the Artificial Intelligence Laboratory and the Center for Biological Information Processing (Whitaker College), the Massachusetts Institute of Technology. Support for the Artificial Intelligence Laboratory's research is provided in part by the Advanced Research Projects Agency of the U.S. Department of Defense under Office of Naval Research contract N00014-85-K-0214. Support for this research is also provided by a grant from the U.S. Office of Naval Research, Engineering Psychology Division, and NATO Collaborative Grant 0403/87. The research of H. A. Mallot was supported by a grant from the Deutsche Forschungsgemeinschaft (Ma 1038/1-1,2).

* Present address, Institut für Zoologie III (Biophysik), Johannes Gutenberg-Universität, D-6500 Mainz, Federal Republic of Germany.

REFERENCES AND NOTES

1. B. Julesz, *Foundations of Cyclopean Perception* (U. Chicago Press, Chicago, Ill., 1971).
2. D. Marr and T. Poggio, "A computational theory of human stereo vision," Proc. R. Soc. London Ser. B **204,** 301–328 (1979).
3. J. E. W. Mayhew and J. P. Frisby, "Psychophysical and computational studies towards a theory of human stereopsis," Artif. Intell. **17,** 349–386 (1981).
4. E. Mingolla and J. T. Todd, "Perception of solid shape from shading," Biol. Cybern. **53,** 137–151 (1986).
5. A. Blake, Zisserman, and G. Knowles, "Surface description from stereo and shading," Image Vision Comput. **3,** 183–191 (1985).
6. K. Ikeuchi and B. K. P. Horn, "Numerical shape from shading and occluding boundaries," Artif. Intell. **17,** 141–184 (1981).
7. A. P. Pentland, "Local shading analysis," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6,** 170–187 (1984).
8. R. Bajcsy and L. Lieberman, "Texture gradient as a depth cue," Comput. Vision Graphics Image Process. **5,** 52–67 (1976).
9. J. R. Kender, "Shape from texture: an aggregation transform that maps a class of textures into surface orientation," in *Proceedings, International Joint Conference on Artificial Intelligence* Kaufmann, Los Angeles, Calif., 1979).
10. A. P. Witkin, "Recovering surface shape and orientation from texture," Artif. Intell. **17,** 17–47 (1981).
11. A. P. Pentland, "Shading into texture," Artif. Intell. **29,** 147–170 (1986).
12. H. G. Barrow and J. M. Tenenbaum, "Interpreting line drawings as three-dimensional surfaces," Artif. Intell. **17,** 75–116 (1981).
13. K. A. Stevens, "The visual interpretation of surface contours," Artif. Intell. **17,** 17–45 (1981).
14. K. A. Stevens and A. Brooks, "Probing depth in monocular images," Biol. Cybern. **56,** 355–366 (1987).
15. S. Ullman, *The Interpretation of Visual Motion* (MIT Press, Cambridge, Mass., 1979).
16. H. C. Longuet-Higgins and K. Prazdny, "The interpretation of moving retinal image," Proc. R. Soc. London Ser. B **208,** 385–397 (1981).
17. N. M. Grzywacz and E. C. Hildreth, "Incremental rigidity scheme for recovering structure from motion: position-based versus velocity-based formulations," J. Opt. Soc. Am. A **4,** 503–518 (1987).
18. B. A. Dosher, G. Sperling, and S. Wurst, "Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure," Vision Res. **26,** 973–990 (1986).
19. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," Nature **317,** 314–319 (1985).
20. D. Terzopoulos, "Integrating visual information from multiple sources," in *From Pixels to Predicates,* A. P. Pentland, ed. (Ablex, Norwood, N.J., 1986), pp. 111–142.
21. J. L. Marroquin, S. K. Mitter, and T. Poggio, "Probablistic solution of ill-posed problems in computational vision," J. Am. Stat. Assoc. **82,** 76–89 (1987).
22. M. L. Braunstein, G. J. Andersen, M. W. Rouse, and J. S. Tittle, "Recovering viewer-centered depth from disparity, occlusion, and velocity gradients," Percept. Psychophys. **40,** 216–224 (1986).
23. H. H. Bülthoff and H. A. Mallot, "Integration of depth modules: local and global depth measurements," Invest Ophthalmol. Vis. Sci. Suppl. **29,** 400 (1988).

24. Reviewed by A. Ardity, "Binocular vision," in *Sensory Process and Perception*, Vol. I of Handbook of Perception and Human Performance, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, New York, 1986), Chap. 23.
25. We write the equation of the ellipsoid as

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

where a, b, and c denote the semidiameters. With $a = b = 1$, we have an ellipsoid of revolution. For a ray from viewpoint $\mathbf{e}$ to a point $\mathbf{p}'$ on the surface,

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

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

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

The image intensity at point $\mathbf{p}'$ 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 $\mathbf{x}^T A\mathbf{x} = 1$, the surface normal is simply $A\mathbf{x}/\|A\mathbf{x}\|$. The viewing direction and the axis of revolution of the ellipsoid were aligned.
26. B. T. Phong, "Illumination for computer generated pictures," Commun. ACM **18**, 311–317 (1975).
27. R. L. Gregory, *Eye and Brain* (McGraw-Hill, New York, 1966).
28. D. Marr and E. Hildreth, "Theory of edge detection," Proc. R. Soc. London Ser. B **207**, 187–217 (1980).
29. J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," J. Opt. Soc. Am. A **2**, 1160–1169 (1985).
30. H. H. Bülthoff and K. G. Götz, "Analogous motion illusion in man and fly," Nature **278**, 636–638 (1979).
31. K. I. Naka and W. A. H. Rushton, "S-potentials from color units in the retina of fish (Cyprinidae)," J. Physiol. (London) **185**, 536–555 (1966).
32. S. Hemilä, "The stimulus-response functions of visual systems," Vision Res. **27**, 1253–1261 (1987).
33. L. Kramer, "Interpretation of invertebrate photoreceptor potentials in terms of a quantitative model," Biophys. Struct. Mechan. **1**, 239–257 (1975).
34. G. Poggio and T. Poggio, "The analysis of stereopsis," Annu. Rev. Neurosci. **7**, 379–412 (1984).
35. W. E. L. Grimson, "Binocular shading and visual surface reconstruction," Comput. Vision Graphics Image Process. **28**, 19–43 (1984).
36. E. C. Hildreth, "The detection of intensity changes by computer and biological vision systems," Comput. Vision Graphics Image Process. **22**, 1–27 (1983).
37. M. A. Gennert, "A computational framework for understanding problems in stereo vision," Ph.D. dissertation (Massachusetts Institute of Technology, Cambridge, Mass., 1987).
38. B. K. P. Horn and M. J. Brooks, "The variational approach to shape from shading," MIT Artif. Intell. Lab. Memo 813 (Massachusetts Institute of Technology, Cambridge, Mass., 1985), pp. 1–32.
39. J. T. Todd and E. Mingolla, "Perception of surface curvature and direction of illumination from patterns of shading," J. Exp. Psychol. Human Percept. Perform. **9**, 583–595 (1983).
40. H. G. Barrow and J. M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Computer Vision Systems*, A. Hanson and E. Riseman, eds. (Academic, New York, 1978), pp. 3–26.
41. W. E. L. Grimson, "A computational theory of visual surface interpolation," Philos. Trans. R. Soc. London Ser. B **298**, 395–427 (1982).
42. J. L. Marroquin, "Surface reconstruction preserving discontinuities," MIT Artif. Intell. Lab. Memo 792 (Massachusetts Institute of Technology, Cambridge, Mass., 1984).
43. T. Poggio, "Integrating vision modules with coupled MRFs," MIT Artif. Intell. Lab. Working Paper 285 (Massachusetts Institute of Technology, Cambridge, Mass., 1985).