MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 965 May 1987

C.B.I.P. Memo No. 024

INTERACTION OF DIFFERENT MODULES IN DEPTH PERCEPTION: STEREO AND SHADING

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

[1] Present Address of HAM:
Institut für Zoologie III (Biophysik)
Johannes Gutenberg-Universität, D-6500 Mainz, FRG

ABSTRACT

A method has been developed to measure the perceived depth of computer generated images of simple solid objects. Computer graphic techniques allow for independent control of different depth cues (stereo, shading and texture) and enable the investigator thereby to study psychophysically the interaction of modules for depth perception. Accumulation of information from shading and stereo and vetoing of depth from shading by edge information have been found. Cooperativity and other types of interactions are discussed. If intensity edges are missing, as in smooth-shaded surfaces, the image intensities themselves could be used for stereo matching. The results are compared with computer vision algorithms for both single modules and their integration for 3D-vision.

© Massachusetts Institute of Technology 1987

Acknowledgments. This report describes research done within the Artificial Intelligence Laboratory and the Center for Biological Information Processing (Whitaker College) at the Massachusetts Institute of Technology. Support for the A.I. Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014–85–K–0124. Support for this research is also provided by a grant from the Office of Naval Research, Engineering Psychology Division and a research fellowship from the Deutsche Forschungsgemeinschaft (Ma 1038/1-1) to HAM.

1. Summary

Depth information can be derived from a number of different cues (stereo, shading, texture and motion, to name a few). Possible types of interaction include accumulation, veto, cooperation, disambiguation, etc. To distinguish between these interactions experimentally, we studied the depth perceived from computer generated images (smooth- or flat-shaded ellipsoids of revolution with different elongations along the viewing axis) containing different combinations of depth cues. The cues could be either consistent or contradictory. Perceived depth was measured by interactively adjusting a depth probe to the surface of the ellipsoid. Depth perception is almost correct when disparity information can be derived from the relative locations of intensity edges in stereo images. If edges are missing, as in a smooth-shaded sphere, stereo depth information can still be derived from the image intensities themselves. If shading is the only information available, the perceived depth may be as low as 30% of the correct depth and is almost independent of the elongation. From this we can draw the following conclusions:

(1) The more information is available, the larger is the perceived depth (accumulation). It increases in the following sequence of cues: shading, stereo without edge information, stereo with edge information.

(2) Since the perceived depth of non-disparate flat-shaded surfaces is zero, we may conclude that edge-based stereo overrides shading (veto).

(3) If no intensity edges are present, depth can still be derived (intensity-based stereo).

(4) Intensity-based stereo cannot be due to intensity peak matching alone. It performs best in the vicinity of the peak but uses distributed information as well (patch correlation).

Both integration of depth modules and binocular shape–from–shading are compared to recently developed ideas in computer vision (intensity-based stereo matching and Markov Random Fields).

2. Introduction

The problem of deriving a description of a three-dimensional scene from its two-dimensional images on the retina is the inverse of classical optics, wherein one has to find the two-dimensional image (brightness distribution) of a three-dimensional object. While the optics problem can be solved straightforwardly, the inverse problem is much harder to attack because a unique solution does not always exist. Furthermore the solution has to be stable, i.e., depend continuously on the image intensities. Computational studies have provided in recent years promising, although far from complete, theories of the processes necessary to solve the ill-posed problem of deriving a three-dimensional scene description from two-dimensional images. It has become clear that a single module is not sufficient to solve this problem. Stereo and motion algorithms, for example, can work well under laboratory-controlled conditions (random dot stereograms and moving sinewave patterns), but quite often make severe errors under more natural conditions where specularity, inhomogeneous illuminations, and occlusion are common. We therefore argue that the analysis of the information processing involved should rely on complex natural images rather than non-complex synthetic images.

2.1. Complex vs Non-Complex Images

The human visual system extracts 3-D information much more reliably for complex natural images than for non-complex synthetic images. For example it can analyze complex shapes in a natural scene under quite different viewing conditions but produces often ambiguous solutions for simple line drawings like the Necker cube. Similar observations can be made for other vision modules like color, stereo and motion. Many illusions occur when only single or a few cues are available but are rare in complex natural situations because the interaction of different cues can avoid false interpretations. In psychophysics, the study of this interaction can be facilitated by the use of computer graphic systems which allow convenient control of different cues in complex synthetic images. Shading, for example, can be computed for arbitrary objects, and ray-tracing and texture mapping techniques allow the computation of synthetic images of three-dimensional scenes which cannot be distinguished from natural images (photographs).

Most studies of depth cues, both in psychophysics and in computer vision, deal with the reconstruction of a three-dimensional scene from one isolated cue, the most intensively studied one being stereo (for example, Julesz 1971, Marr & Poggio 1979, Mayhew & Frisby 1981). From the computational point of view, there also exist a number of studies on how to evaluate texture information (Bajcsy & Lieberman 1976, Kender 1979, Witkin 1981, Pentland 1986), shading (Koenderink & van Dorn 1980, Ikeuchi & Horn 1981, Pentland 1984), and motion (Braunstein 1976, Ullman 1979, Hildreth 1983). There is, however, little

knowledge of how the information from these cues can be integrated by the human visual system.

2.2. Classification of Depth Cues

Three types of cues may be distinguished from the large number of cues from which depth information may be inferred (for review see Braunstein 1976):

- Primary depth cues that provide "direct" depth information, such as convergence of the optical axes of the two eyes, accommodation, and unequivocal disparity cues.
- Secondary depth cues that may also be present in monocularly viewed images. These include shading, shadows, texture gradients, motion parallax, kinetic depth effect, occlusion, 3D-interpretation of line drawings, structure and size of familiar objects.
- Cues to flatness, inhibiting the perception of depth. Examples are frames surrounding pictures, or the uniform texture of a poorly resolving CRT-monitor.

In the scope of computational vision, an alternative approach to a classification of depth cues could rely on the observation that different cues require a different amount of preprocessing. For example, convergence and accommodation can be evaluated straightforwardly, whereas stereo disparity requires the previous extraction of some matching primitives from the image. To evaluate occlusion or the apparent size of familiar objects, even more preprocessing is required. In a complex scene, an object may be detected by a disparity discontinuity. Once it is defined, it may appear to be partly occluded by other objects and thus depth information would be gained from a higher level scene description. Only recently, attempts have been made to find general strategies for the integration of all this information in computer vision, e.g., by Poggio and Gamble (see Poggio 1987).

2.3. Interaction of Depth Cues

In principle, there are several types of possible interactions between different depth cues, which are not mutually exclusive:

- **Accumulation:** Information from the different modules could be accumulated in a way similar to the (non-linear) summation known from spatial frequency channels (probability summation).
- **Veto:** There can be unequivocal information from one cue that should not be challenged by others. In general primary depth cues should override secondary depth cues.
- **Cooperation:** Especially in the case of poor or noisy cues, the modules might work synergistically.

- **Disambiguation:** Information from one module can be used locally to disambiguate a representation derived from another module. Also, a global ambiguity of depth-order (convex-concave) can occur from cues like shadows or kinetic depth (Braunstein et al. 1986).
- **Hierarchy:** Information derived from one cue may be used as raw data for another one.

2.4. Representation of Depth

In principle, there are many different ways to represent depth information. The most straightforward way is to produce a depth-map of all the points in the field of view. Another way is to segment the scene into distinguishable objects and describe the shape of the objects in more abstract terms. For the latter way, different approaches have been tried in the last decade. For example, Marr (1978, 1982) proposed the $2\frac{1}{2}$D-sketch which includes rough distances to surface patches as well as their orientations, and Koenderink & van Doorn (1979, 1980) used the tools of differential geometry and related their ideas to Gestalt theories of perception.

For a psychophysical approach to these questions, we studied the depth perceived from computer generated images containing different combinations of depth cues. The shading and stereo cues could be either consistent or contradictory. In contrast to other studies of shape perception (Todd & Mingolla 1983, Mingolla & Todd 1986), we did not try to describe the shape by measuring the surface orientation of the displayed objects, but rather tried to infer the shape from direct depth measurements of the surface of the objects. This was done interactively by adjusting a depth probe to the surface of an ellipsoidal object as described in the next chapter.

3. Methods

3.1. Computer Graphic Psychophysics

Images of smooth-shaded ellipsoids and flat-shaded polyhedral ellipsoidal objects were generated 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. Textures and simple figures could be mapped onto the surface. The polyhedral objects were derived from quadrangular tesselations of the sphere along meridian and latitude circles. These were

elongated along an axis in the equatorial plane, the axis of elongation again being perpendicular to the display screen. 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, or 4.0. With an original radius of 6.67 cm, this corresponds to depth values between 3.33 and 26.68 cm. In the following, all semi-diameters will be given as multiples of 6.67 cm.

The imaging geometry used in the computations is shown in Figure 1. 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 $\mathbf{e}_l$ and $\mathbf{e}_r$, 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 corresponds to an interval from 4.3 cm to 15.2 cm in front of the screen.

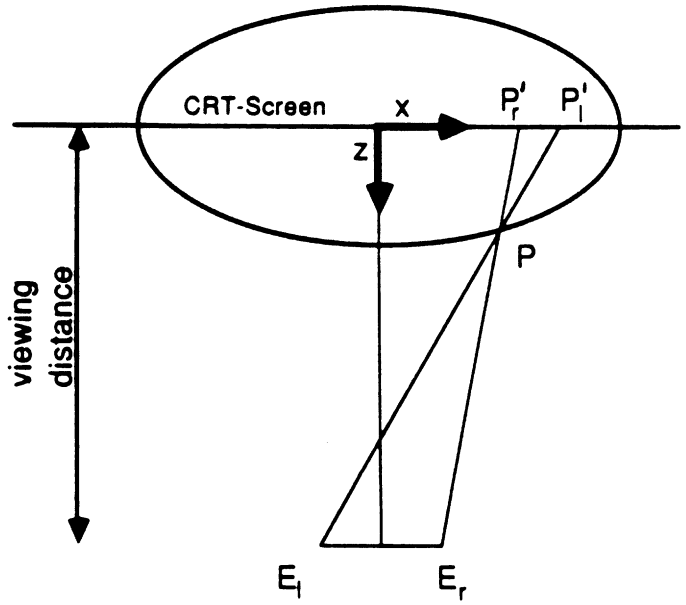

Figure 1 *Imaging geometry.* Projection onto the x-z-plane. Viewing distance is 120 cm. $\mathbf{e}_l, \mathbf{e}_r$: nodal points of the left and right eye, respectively. The distance between $\mathbf{e}_l$ and $\mathbf{e}_r$ is 6.5 cm. A point $\mathbf{p} \in \mathbf{R}^3$ is imaged at $\mathbf{p}'_l$ 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. We write the equation of the ellipsoid as

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

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

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

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

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

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 $\mathbf{x}^T\mathrm{A}\mathbf{x} = 1$, the surface normal is simply $\mathrm{A}\mathbf{x}/\|\mathrm{A}\mathbf{x}\|$. The viewing direction and the axes of illumination and of revolution of the ellipsoid were aligned. Since our objects were convex, no cast shadows or repeated scattering had to be considered.

3.2. Experimental Procedure

We displayed either a pair of disparate images or one single (monocular) 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 $\times$ V) 1024 $\times$ 874 pixels; bandwidth $\pm$3dB between 50 Hz and 50 MHz, short persistence phosphore). 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. Both disparate and monocular images were viewed 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 resolution of 254 gray-levels. The background was colored in half saturated blue.

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. Thus, the accommodation was the same for viewing both the surface and the probe. 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 ellipsoidal surface. After some training, subjects felt comfortable with this procedure and achieved reproducible depth measurements. All stimuli were viewed binocularly. Subjects included the authors (corrected vision) and one naive observer.

3.3. Data Evaluation

The above procedure leads to a local depth map at 45 positions in the image plane. To obtain more global measures of perceived elongation and shape, we first performed a principle component analysis on all data sets, treating each one as a point in 45-space. Variance of

the perceived shapes was found mainly (0.95) along two principal axes. In Figure 2, these are shown together with two analytical surfaces which allow an appropriate interpretation of these components. The first principle component is very close to an ideal ellipsoid (or sphere) which appears in Figure 2c. A model of the second principle component is derived from the depth gradient of the sphere which, in cylindrical coordinates, is $z = r/\sqrt{1-r^2}$. This 45-vector is orthogonalized (Gram-Schmidt) with respect to the sphere. The result is shown in Figure 2d; it provides a reasonable fit of the second component. In what follows, we will use this theoretical frame derived from the ellipsoids depth and depth gradient rather than the actual principle components. The corresponding coefficients will be called *perceived elongation* and *deformation*, respectively. Since they are derived from all 45 measurements of a set, their scatter is very small. The results were confirmed by other methods of evaluation, such as computing a least squares fit of an ellipsoid to the data.

It can be seen from the eigenvalues associated with the principle components ($\lambda_1 = 0.94$, $\lambda_2 = 0.01$) that the main difference of the perceived surfaces is in their elongation rather than in their shapes. This is partly due to the fact that stimuli with different elongations were used in the first place. Slight variations in the deformation will be discussed later.

4. Results

Four different image types were tested:

- Flat-shaded ellipsoid with disparity and edge information (D^+E^+)
- Smooth-shaded ellipsoid with disparity but without edge information (D^+E^-)
- Flat-shaded ellipsoid without disparity but with edge information (D^-E^+)
- Smooth-shaded ellipsoid with neither disparity nor edge information (D^-E^-).

Each image type was tested for four different elongations (0.5, 1.0, 2.0, 4.0). The subjects did not know the elongation of the displayed objects. Altogether, 253 measurements were performed, each consisting of 45 adjustments of the depth probe to the perceived surface. Results were consistent in all three subjects, with differences mainly in the standard deviation. The 16 plots of Figure 3 show the averaged results of all subjects for the four types of experiments and the four different elongations.

4.1. Accumulation of Depth Information

The perceived elongation in the consistent images depends on the amount of information available. As can be seen from Figure 4, the perceived elongation is almost correct when

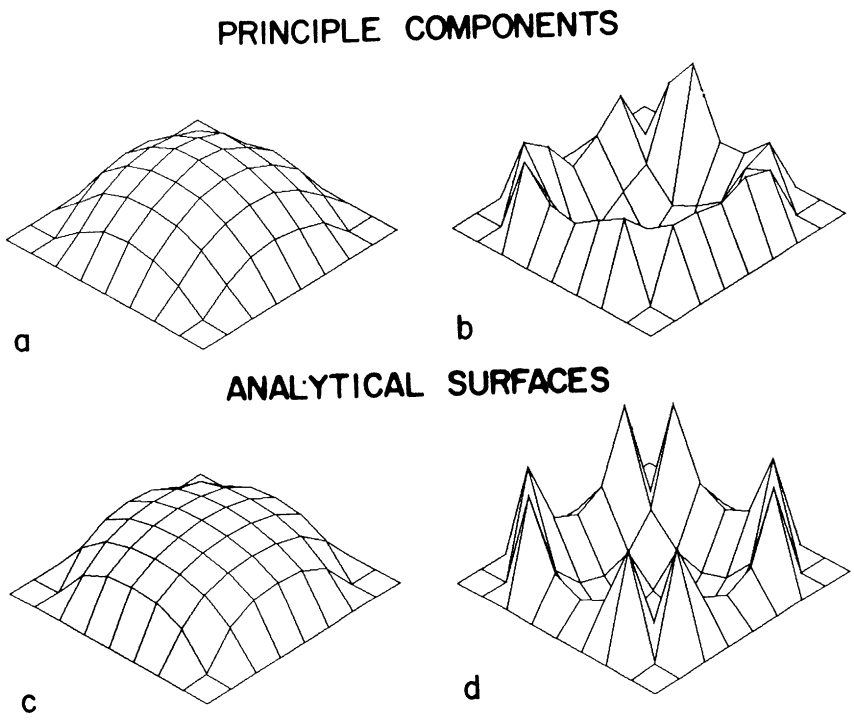

Figure 2 *Classification of the perceived surfaces.* **a,b.** Principle components. **a.** First component, $\lambda_1 = 94\%$. **b.** Second component, $\lambda_2 = 1.4\%$. **c,d.** Analytical surfaces that can be used to interpret the principle component data. **c.** An ideal ellipsoid is almost identical to the first component. The associated coefficient is used as a measure of the *perceived elongation.* **d.** The depth gradient of the ellipsoid leads to an analytical model of the second component. The associated coefficient describes the deviation of the perceived surface from an ellipsoid; it will be called *deformation.* Negative deformations correspond to a more cone-like percept, positive to a more cylindrical surface.

shading, intensity-based and edge-based disparity informations 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 (but see Section 4.4).

4.2. Edge-Based Stereo Vetoes Shading

In experiment D^-E^+, two identical images (no disparity) of flat-shaded 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

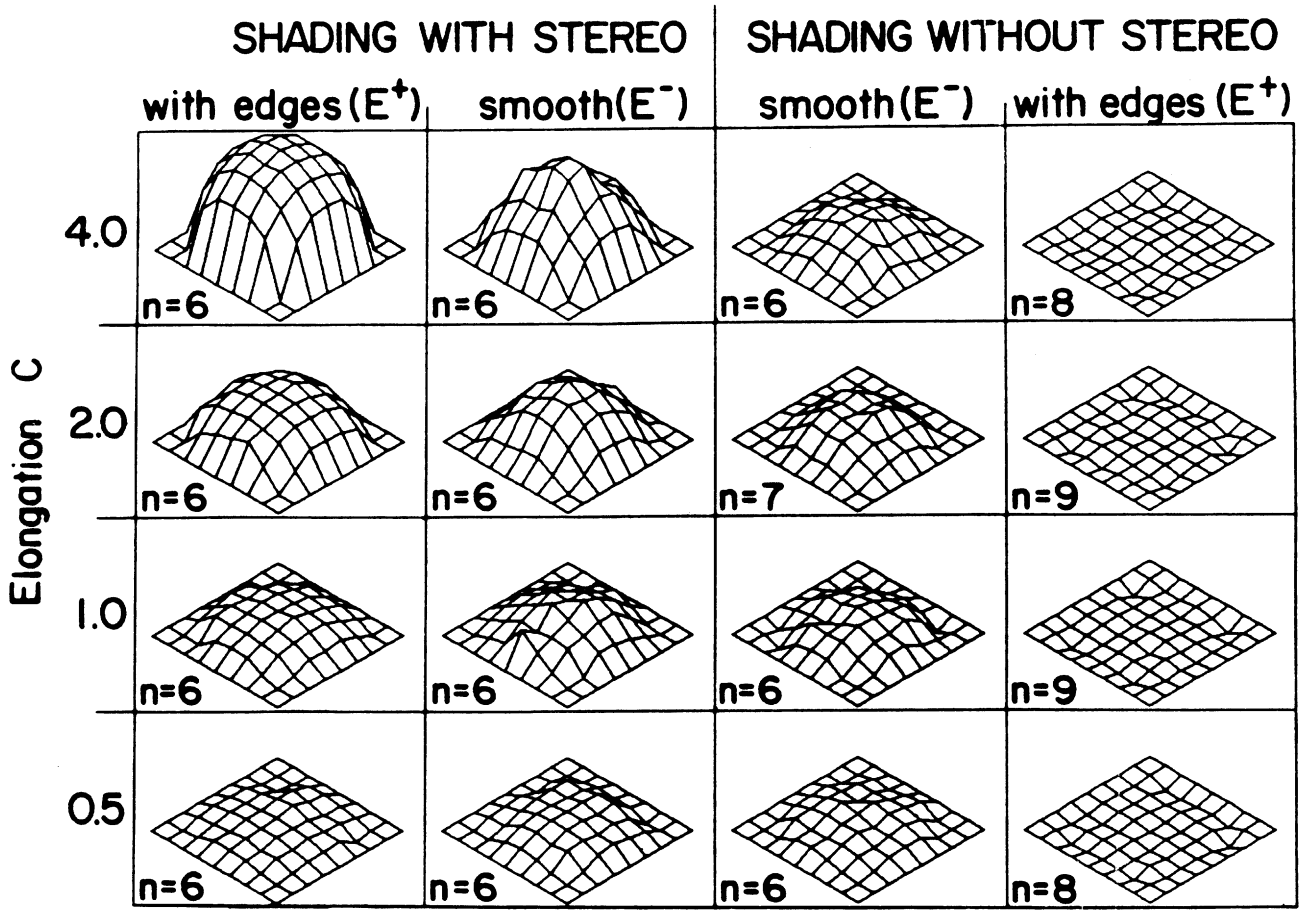

Figure 3 *Perceived surfaces* (depth not drawn to scale) Each plot shows the average of 6 - 9 sessions from three subjects. Perceived depth decreases with the following sequence of cue-combinations: disparity, edges and shading (D^+E^+); disparity and shading but no edges (D^+E^-); shading only (D^-E^-); contradictory disparity and shading (D^-E^+). The elongation of the displayed objects is denoted by c.

zero. Since the perceived depth does not change with elongation, we may conclude that edge-based stereo matching overrides shading. This is an example of the veto-relationship mentioned in the introduction. This finding is confirmed by an additional experiment where a small stereo marker was attached to the smooth surface (cf. Section 6.1). Note, however, that this veto-relationship might occur only in the locally derived depth map. The global percept of the polyhedral ellipsoid is not flat but convex.

4.3. Intensity-Based Stereo

Depth can still be perceived when no disparate edges are present. This is not surprising, since

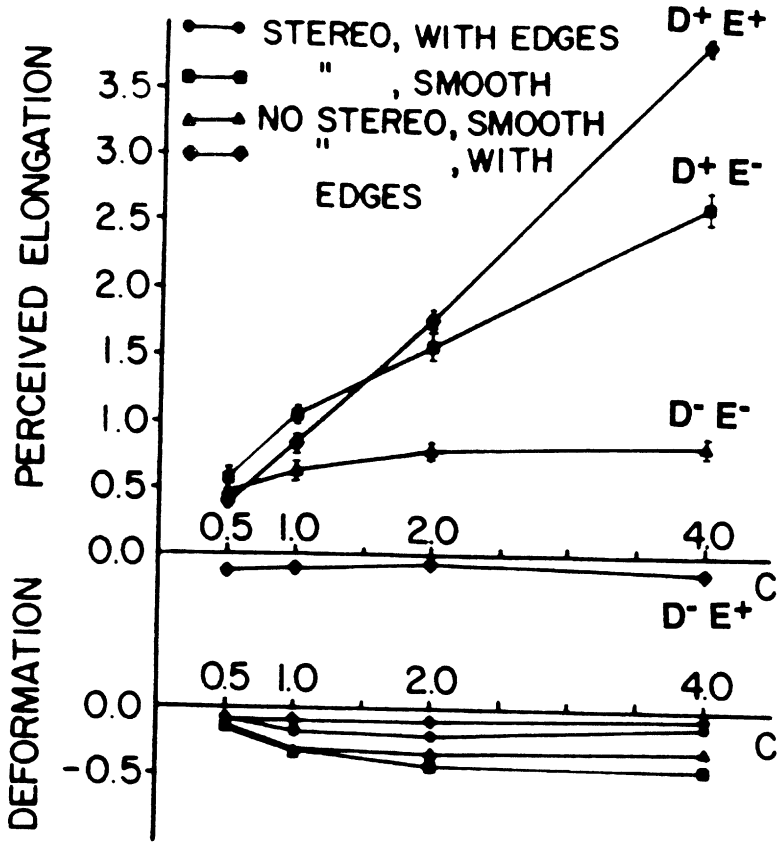

Figure 4 *Perceived elongation and deformation.* **top:** Depth perception improves as the number of available cues increases. The significant separation of the second and third curve (smooth shading with and without disparity) illustrates the influence of disparity information even in the absence of edges. **bottom:** Deformation (cf. Figure 2b). In the experiments with disparate edges, the coefficients are negligible. In all other experiments, the coefficients are negative, i.e., a more conical surface is perceived.

shading information was still available. A comparison of the results (Figure 4) for smooth-shaded images with and without disparity information, however, establishes a significant contribution of intensity-based disparity information. The curves for D^+E^- and D^-E^- are significantly separated for all elongations except 0.5. We therefore conjecture an intensity-based stereo mechanism that does not rely on edge information. This effect is almost as strong as edged-based stereo. A significant smaller depth perception is elicited only for larger elongations. Note, that for these elongations the ellipsoid does not fit into Panum's fusional area. One could argue that even in the smooth-shaded images one salient edge is present, namely the occluding contour. However, this boundary was placed in the zero-disparity plane in all experiments and therefore does not provide depth information. Note, that the self-shadow boundary coincides with the occluding contour since illumination was from the front. A control experiment with oblique lighting directions confirmed the findings

described here (cf. Section 6.1). For some general remarks on images without zero-crossings, see Section 5.1.

Preliminary results suggest that intensity-based stereo is vetoed by edge-based stereo, as is shade from shading. Thus, the two stereo mechanisms appear to be functionally separated.

4.4. Intensity-Based Stereo Does Not Veto Shading

If stereo matching can be performed without edge information, the depth cues in the experiment with smooth-shaded non-disparate images (D^-E^-) are contradictory in the sense that shading suggests some depth whereas stereo does not. A similar contradiction occurs in flat-shaded non-disparate images when edge-based stereo is considered. It appears that intensity-based stereo does not veto shading information, as did edge-based stereo in experiment D^-E^+. The contradiction, however, may be the reason for the saturation in the perceived depth from shading (Figure 4).

5. Discussion

Problems in vision are usually classified as part of low-level (or "early") vision or part of high-level vision. Early vision is the set of visual modules that perform the first steps of recovering physical properties of surfaces from two-dimensional images. High-level vision deals with the "later" problems of object recognition and shape representation.

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 and exploited in its algorithms. On the other hand *changes of image intensity* convey often information about physical edges in the scene. The location of sharp change 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.

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 more dense depth data by using in addition other features than edges or even use a complete featureless mechanism (eg., intensity-based stereo) we computed images without sharp changes in image intensity.

5.1. Images without Zero-Crossings

For the discussion of intensity-based stereo, the absence of zero-crossings in the Laplacians of images of smooth ellipsoids is crucial. Here, we show that for an orthographically projected image of a sphere with Lambertian reflection function and parallel illumination, zero-crossings 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 $\mathbf{l} = (0, 0, 1)$ and the Lambertian reflectance function, we obtain the luminance profile

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

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

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

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 5a, 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 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.

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

5.2. Receptor Non-Linearities and Image Interpretation

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). We therefore recalibrated the CRT to compensate either for the CRT non-linearity only, or for the non-linearities of both the CRT and the retina.

Retinal non-linearities in both vertebrates (Naka & Rushton 1966, Dawis 1978) 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. Among other things, $I_{0.5}$ depends on the adaptation of the eye. We repeated experiments D^+E^- and D^-E^-, i.e., those involving smooth-shaded images, with compensation for either monitor non-linearities or the combination of monitor and retina non-linearities with four different choices of the constant $I_{0.5}$. The results did not show significant differences from those obtained without corrections.

Figure 5a shows the luminance profile for an ellipsoid with elongation 4.0, and the effect of a non-linearity given in Equation 9 for a number of choices of $I_{0.5}$. It turns out that in our experiments, the presumed receptor non-linearities tend to cancel the small 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 either in the original image or in its first neural representation cannot account for the intensity-based stereo performance found in our experiments.

6. Relation to Computational Studies

6.1. Edge-Based vs Intensity-Based Stereo

The major finding of this study, as far as single depth modules are concerned, is the strength of depth perception obtained from intensity-based stereo. In computational theory, most

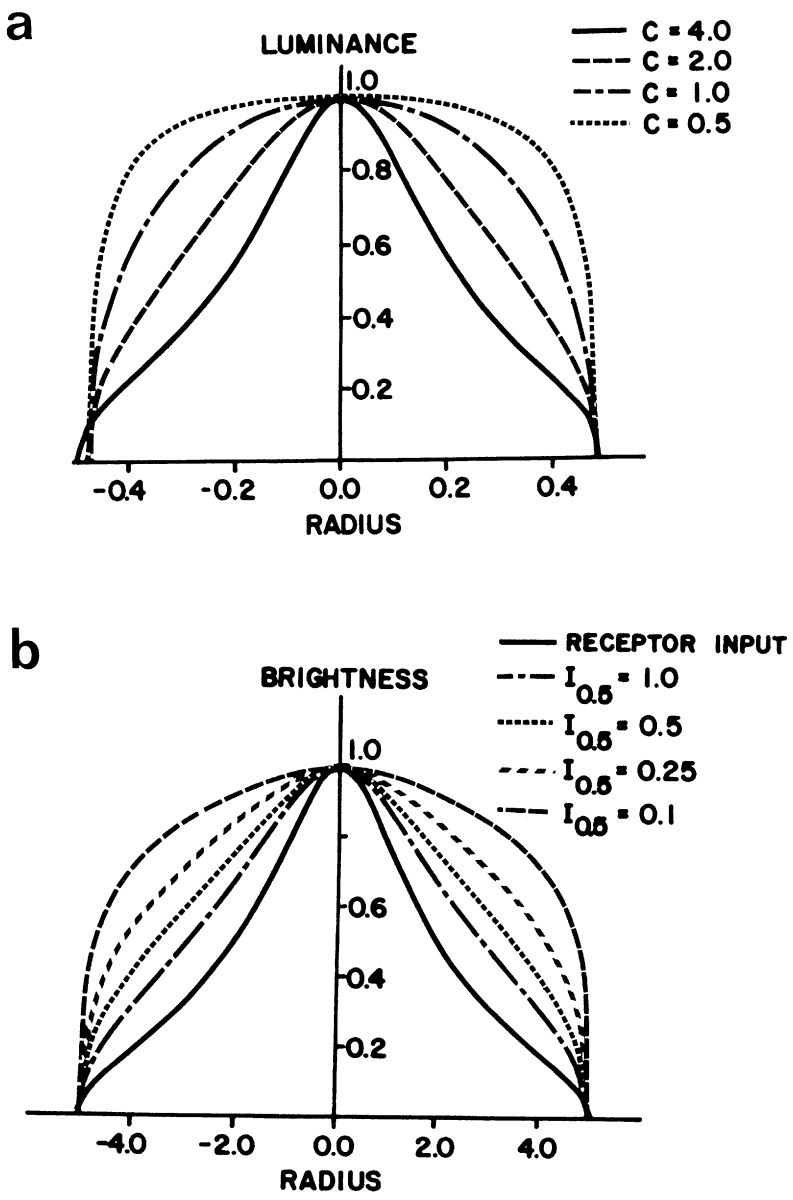

Figure 5 *Luminance and simulated brightness profiles.* **a.** Luminance of ellipsoids with different elongations. The functions differ from those given analytically in Equation 8 only in a slight distortion of the x-axis which is due to perspective rather than orthographic projection. Note that for elongations larger than 2.0, inflections occur. **b.** Simulated perceived brightness profiles for the ellipsoid with elongation 4.0 (the one with the pronounced inflections in Figure 5a). Receptor characteristics are accounted for by the non-linear compression described in Equation 9. The non-linear compression tends to cancel the inflections (which might give rise to zero-crossings) rather than to enhance them.

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 more reliable 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.

A number of additional experiments were performed to confirm the involvement of intensity-based stereo and to study its relationship to edge-based stereo. First, 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 ($\pm 14°$ azimuth and $\mp 13.6°$ elevation from the viewing direction). The results of these experiments are depicted in Figure 6. Note that no depth values were determined in the dark (shadowed) parts of the images. The results confirm the original finding that intensity-based stereo is present and is much stronger than pure shape–from–shading. Furthermore, when illumination is from the lower right, stereo prevents depth inversions which occasionally occurred in the non-disparate images. One has to keep in mind, however, that in the case of oblique illumination, the self-shadow boundary provides some edge information which improves depth perception in the stereo images and inhibits it in the non-disparate cases. Nevertheless, these data show that our original findings were not critically dependent on the special lighting conditions used.

In a second series of control experiments, we studied the interaction of intensity-based and edge-based stereo. In contrast to the original measurements with flat-shaded ellipsoids where edge-information was distributed all over the surface, we placed a small dark ring (Radius 7.5 mm, Contrast 0.11) at the tip of the ellipsoid. The stereo disparity of this ring could be chosen independently from the disparity of the shaded surface. Three cases were tested: consistent disparities in ring and shading, no disparities in ring and shading, and a disparate ring in front of a non-disparate shaded image. The first two cases (left and right columns in Figure 7) confirm the earlier findings of accumulation of depth information and vetoing. Although pure shape–from–shading yields some depth perception in the periphery, it is vetoed in the center by the non-disparate edge-information.

The third case, a stereo ring in front of a non-disparate smooth image (middle columns in Figure 7) provides information on the mechanisms involved in intensity-based stereo. One possibility is described by Mayhew & Frisby (1985) who propose a modification of the Marr-Poggio model (1979) where matches in the two images may occur before edge-detection is complete. In particular, they discuss peaks in image irradiance as additional matching primitives. However, it appears that their experimental data can be explained with level- (rather than zero-) crossings in the Laplacian of the image irradiance, or with a shift of the zero-crossings due to some prior filtering as well (Marr & Hildreth 1980, Hildreth 1983). Another possibility is that intensity-based stereo does not rely on matching primitives at all. For example, Gennert (1987) has developed a new intensity-based stereo algorithm that

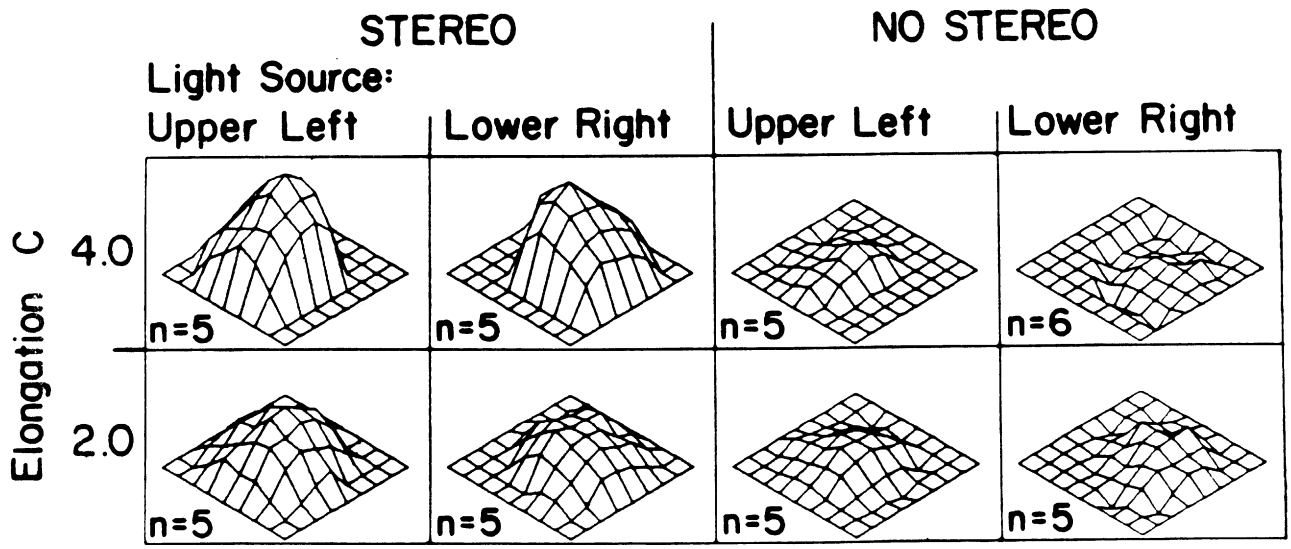

Figure 6 *Perceived surfaces for oblique illuminations* (Format as in Figure 3) Illumination from the upper left (first and third column) and from the lower right (second and fourth column). No depth was measured in the self-shadow regions. The data confirm the relevance of intensity-based stereo and show the independence of our findings from the lighting conditions.

makes use of a spatially varying linear transformation to relate gray-levels in the two images. A distributed mechanism of that kind would be especially useful in surface interpolation when matching primitives are sparse. Unfortunately this algorithm has specific problems with the particular images used in the psychophysical experiments. A severe matching error occurs where the intensity profiles of the left and right stereo images cross. The intensity at this point is the same and the algorithm matches these points leading to a zero or small disparity at a point where actually the maximum disparity should be expected. To avoid such a matching error information other than the image intensity alone has to be taken into consideration. For example, the slopes of the intensity profiles are different for these points where the image intensities are the same. To use the first derivative as an additional constraint could solve this matching problem without introducing too much noise into the system because the image intensity will still be the primary matching primitive.

The computer experiments with psychophysical images as shown above is a good example of the fruitful interaction between computational theory and psychophysics. Psychophysics cannot only be used as an existence proof for a solution of a computational problem, but as shown above, could also give hints to weak points in computer vision algorithms. This becomes even more clear if algorithms are tested with images, that the human visual system can easily deal with – natural images. Image intensity alone is certainly not

enough to compute a correct depth map. Higher order terms have to be taken into consideration and this is what the human visual system does when it uses more than one cue or matching primitive.

The principle component analysis of our original data (Figures 2 and 4) provides a first clue as to how intensity-based stereo might work. The coefficients corresponding to the depth gradient shown in Figure 2d are negative for intensity-based stereo indicating a somewhat cone-like percept. In edge-based stereo, these coefficients are zero. This finding suggests that in intensity-based stereo, perception is best in the vicinity of the intensity peak, as would be expected from intensity-peak-matching but not from the distributed mechanism described above.

The notion of intensity-peak-matching was tested with a disparate token displayed in front of a non-disparate background providing shading information only. Since the peak is replaced by a disparate edge token, the loss of global intensity disparities should not degrade the performance of a peak-matching mechanism. For the elongations 1.0 and 2.0, the results are in fact equal to those obtained with full stereo information; i.e., one salient stereo token in the center of the object (together with shape–from–shading) is sufficient to yield the same perception as a complete intensity stereo pair (Figure 7, middle columns). However, for the elongation 4.0, it seems that a single stereo match in the center of the object is not sufficient to produce the same percept as full intensity disparities. The difference between the results for the two subjects corresponds to an ambiguity which was experienced by both observers. For the large elongation, the object appears to consist of a solid base with about half the depth of the ring and a "glass dome" onto which the ring is drawn. While HAM adjusted the depth probe to this 'subjective surface', HHB measured the solid base. No such subjective surface was perceived in intensity-based stereo. We conclude that at least for large disparities, one single token such as the intensity peak is not sufficient to yield the full depth percept. Rather, the distributed disparity information seems to be utilized globally.

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. Until now, our experiments explored only purely Lambertian surfaces. We shall, however, include different reflectance functions and lighting conditions in future studies. At any rate, it is an interesting result that human observers are able to evaluate binocular shading information in the Lambertian case. From this we may conclude that a mechanism different from the one proposed by Grimson is involved.

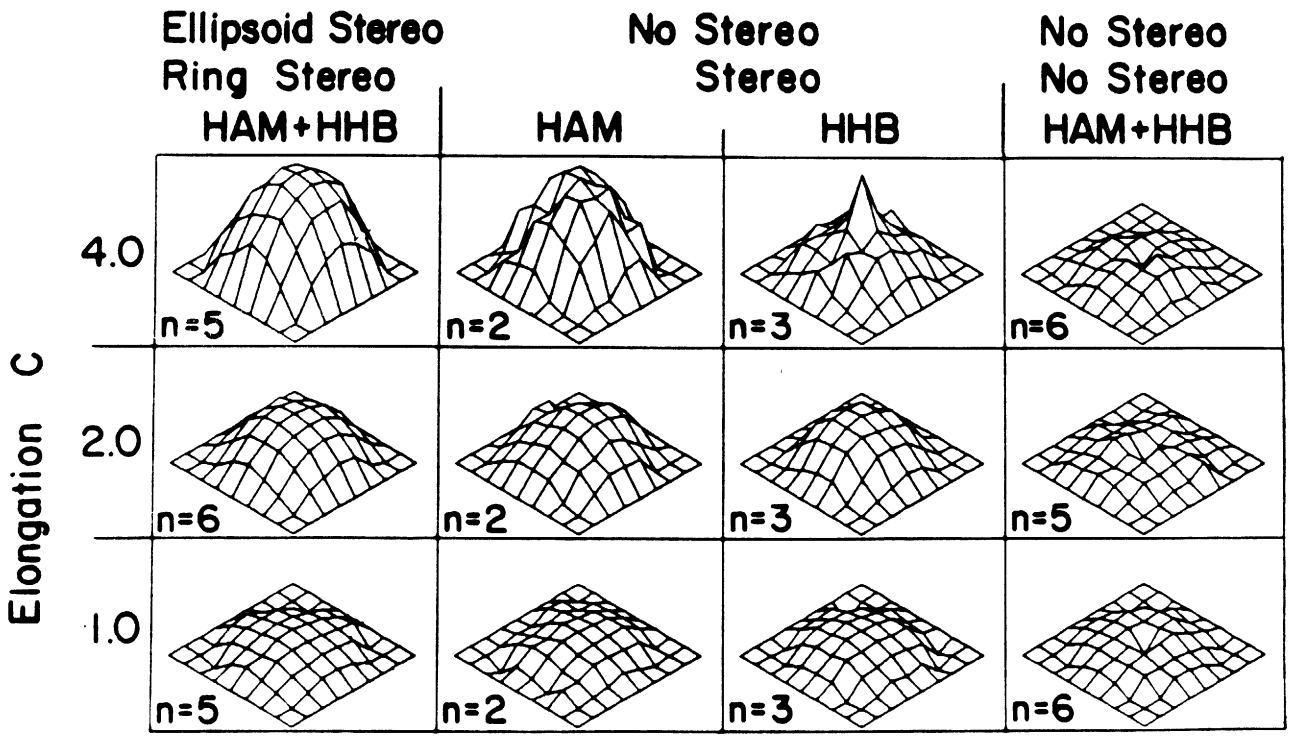

Figure 7 *Perceived surfaces for smooth shading combined with a small stereo token* (Format as in Figure 3) Edge-based stereo information cancels shape–from–shading (right column). When the token has the correct disparity, intensity-based stereo does not further improve the percept, at least for small elongations. For the elongation 4, the data are ambiguous. (For further discussion, see text.)

6.2. Shape from Shading

The case of pure shape–from–shading is studied in our experiment D^-E^-. Ikeuchi & Horn (1981) provide a computational theory of shape–from–shading. Their algorithm starts out from the occluding contour of a given object and successively computes first the surface-orientation and subsequently the depth within the surface. As an example, Ikeuchi & Horn 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 to our experiment. As can be seen from their Figure 15, the algorithm converges fastest in the vicinity of the occluding contour, i.e., in the periphery of the sphere, whereas errors persist for some iterations in the center. Interestingly the same dependence of the error on the position is found in our experiments and their algorithm underestimates depth in a similar way as the

human observer does. Note, however, that the distortion of shape in their algorithm depends on the regularization parameter λ. For a large value of λ, which would be appropriate for noisy image data, the smoothing of the surface would lead to a considerable underestimation of depth. Conversely, for small values of λ the smoothing would be less. The iterative scheme bocomes unstable, however, if the value of λ is reduced too much. In any case, it would be more desirable to compare the human performance with a shape–from–shading algorithm which does not depend so strongly on a single parameter. For an approach which avoids smoothing introduced by a regularization term see Horn and Brooks (1985).

The algorithm of Ikeuchi & Horn shows also other types of errors when the required knowledge on 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, the main errors were of different types. As can be seen from Figures 2 and 3, errors included underestimation of elongation and the deformation of the ellipsoidal shape to a more cone-like percept. Asymmetric deformations as reported by Ikeuchi & Horn did not occur even for the obliquely illuminated objects (Figure 6). Note, however, the asymmetry in shape perception for the two light source positions (upper left and lower right). The perceived surface for the lower right postion of the light source is neither convex nor concave. Interestingly, even for such simple shapes like ellipsoids observers seemingly neclect to force global consistency (R. Wildes, pers. communication).

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

Todd & Mingolla (1983, 1986) used psychophysical techniques to investigate how observers analyze shape by use of shading cues. According to their results, the human observer makes errors up to 50% in estimating shape–from–shading. 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 shape. 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 psychophysical tools? Obviously everybody can describe the shape of a vase in a photograph even without any texture on it. In principle, shading can provide only information about surface orientation and not absolute depth measurements. But as Todd and Mingolla have shown, a long training phase is required for subjects to point out the surface normal on simply shaded rigid bodies. And even after the training phase subjects make a lot of errors. A precise measurement of surface-slant and tilt does not seem to be necessary for humans

to describe shape. If we do not use slant of surfaces ($2\frac{1}{2}$D-sketch) it seems likely that we use other cues to construct a depth-map of an object.

In the study reported here, we tried to answer this question by measuring 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. Surprisingly, this method worked for shading cues alone (no disparity). This is not obvious, since it involves a cross comparison of supposedly more or less independent modules and also comparison of local (depth probe) versus global (shading) information. On the other hand, our depth probe requires binocular viewing even for non-disparate images (pure shape–from–shading). The rivalry between shape–from–shading and intensity-based stereo (cf. Section 4.4) may be partly responsible for the poor shape–from–shading performance. To avoid this we are currently developing a paradigm to measure shape–from–shading monocularly. With this paradigm we can analyze also other cues, eg. texture gradients and occluding contours which would show similar problems with a local stereo depth probe.

6.4. Interaction of Depth Modules

Concrete predictions as to what types of interactions should occur between different depth cues are still difficult to obtain from computational studies. Therefore, we hope that psychophysical studies will in turn provide useful hints for computational investigations as to how an integration of depth information could work. In this section, we try to relate our results to some of the emerging concepts of visual integration.

Accumulation is a simple type of interaction that can be implemented in a number of different ways. Consider for example Marr's $2\frac{1}{2}$D-sketch (Marr & Nishihara 1978, Marr 1982). Information on surface orientation can be collected from different modules such as shading, texture (density- and deformation-gradient), or 3D interpretations of line drawings. It seems natural that performance improves when more information is available.

Similar results should be obtained with the approach of regularization theory (Poggio et al. 1985). Originally introduced as a unified theory of a number of different modules in early vision, it is equally suited to model the integration of different modules by joint optimization of different sets of data (Terzopoulos 1986). Depending on the choice of the particular loss-functions, the described interaction types of accumulation and cooperativity are likely to occur. In fact, it should be possible to infer the form of the minimized functional from the particular type of summation found psychophysically between the involved modules.

More 'asymmetric' types of interaction, such as veto or disambiguation, can be expected from models of surface interpolation (Grimson 1982) that start with reliable depth information typically obtained from disparate edges and employ other modules, especially shading,

to improve the interpolation between the sites of the edges (R. Wildes, pers. communication). The combination of edge and shading information is thus similar to the combination of occluding contours and shading in Ikeuchi & Horn (1981). A similar relationship has been assumed between edge-based stereo and binocular shading (intensity-based stereo) (Grimson 1982).

Recently, Poggio (1985) proposed another formalism for the integration of different depth modules, based on a probabilistic approach to optimization by non-convex functionals (Marroquin 1984, Marroquin et al. 1986). 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 interactions are to be expected from the coupled Markov Random Field approach.

Acknowledgments. We are grateful to E. Hildreth and T. Poggio who contributed to this work and read the draft of the paper. We also like to thank B. Saxberg, and R. Wildes for helpful discussions, I. Bülthoff for being a subject, and J. Little for assistance on the Lisp-Machine. The work was performed during a postdoctoral fellowship of HAM at MIT which was made possible by T. Poggio and a grant from the Deutsche Forschungsgemeinschaft.

References

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

Barrow H.G. & Tenenbaum J.M. Recovering intrinsic scene characteristics from images. In: *Computer Vision Systems*, A. Hanson & E. Riseman (eds.), Academic Press, New York, San Francisco, London, 1978

Braunstein M.L. Depth Perception through Motion. Academic Press, New York, San Francisco, London, 1976

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

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

Dawis M. A model for light adaptation: producing Weber's law with bleaching–type kinetics. *Biol. Cybern.*, vol. 30, pp. 187–193, 1978

Geiger D. & Poggio T. Level crossings and Panum area. *MIT Artif. Intelligence Lab. Memo.* in prep., 1987

Gennert M. *MIT Artif. Intelligence Lab. Memo.* in prep., 1987

Grimson W.E.L. A computational theory of visual surface interpolation. *Phil. Trans. Roy. Soc. London B*, vol. 298, pp. 395–427, 1982

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

Hildreth E.C. The measurement of visual motion. *The MIT Press*, Cambridge, MA., and London, U.K., 1983

Hildreth E.C. The Detection of Intensity Changes by Computer and Biological Vision Systems. *Computer Vision, Graphics and Image Processing*, vol. 22, No.1, pp. 1–27, 1983

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

Julesz B. Foundations of cyclopean perception. *The University of Chicago Press*, Chicago and London, 1971

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

Koenderink J.J. & van Doorn A.J. The internal representation of solid shape with respect to vision. *Biol. Cybern.*, vol. 32, pp. 211–216, 1979

Koenderink J.J. & van Doorn A.J. Photometric invariants related to solid shape. *Optica Acta*, vol. 27, pp. 981–996, 1980

Kramer L. Interpretation of vertebrate photoreceptor potentials in terms of a quantitative model. *Biophys. Struct. Mechanism*, vol. 1, pp. 239–257, 1975

Marr D. Vision. Freeman and Company, San Francisco, 1982

Marr D., Nishihara H.K. Representation and recognition of the spatial organization of three–dimensional shapes. *Proc. Roy. Soc. Lond. B*, vol. 200, pp. 269–294, 1978

Marr D., Poggio T. A computational theory of human stereo vision. *Proc. Roy. Soc. London B*, vol. 204, pp. 301–328, 1979

Marr D., Hildreth E. Theory of Edge Detection. *Proc. Roy. Soc. Lond. B*, vol. 207, pp. 187–217, 1980

Marroquin J.L. Surface reconstruction preserving discontinuities. *MIT Artif. Intelligence Lab. Memo.* 792, 1984

Marroquin J.L., Mitter S. & Poggio T. Probabilistic solution of ill–posed problems in computational vision. *Proc. Image Understanding Workshop* Baumann (ed.) Scientific Applications International Corp., 1986

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

Mingolla E.& Todd J.T. Perception of solid shape from shading. *Biological Cybern.*, vol. 53, pp. 137–151, 1986

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

Pentland A.P. Local shading analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 170–187, 1984

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

Poggio G. & Poggìo T. The analysis of stereopsis. *Ann. Rev. Neurosci.*, vol. 7, pp. 379–412, 1984

Poggio T. Integrating vision modules with coupled MRFs. *MIT Artif. Intelligence Lab. Working Paper* 285, 1985

Poggio T. MIT Progress in Understanding Images. *Proc. Image Understanding Workshop* Baumann (ed.), Scientific Applications International Corp., 1987

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

Terzopoulos D. Integrating visual information from multiple sources. In: Pentland, A.P. (ed.) From pixels to predicates, Ablex Publishing Corp., Norwood, N.J. 1986

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

Ullman S. The interpretation of visual motion. *The MIT Press*, Cambridge, MA., and London, U.K., 1979

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

CS-TR Scanning Project
Document Control Form

Date : 5/26/95

Report # AIM-965

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☐ Technical Report (TR) ☒ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 24(30-IMAGES)
Not to include DOD forms, printer instructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☒ DOD Form (2) ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages(by page number):_____

Photographs/Tonal Material (by page number):_____

Other (note description/page number):

	Description :	Page Number:
(A)	IMAGE MAP (1) UN#'ED TITLE PAGE	
	(2-24) PAGES #'ED 1-23	
	(25-27) SCANCONTROL, DOD (2)	
	(28-30) TRGT'S (3)	
(B)	CUT+PAST FIG'S ON PAGES 5, 8-10, 14, 16, 18	

Scanning Agent Signoff:
Date Received: 5/26/95 Date Scanned: 6/7/95 Date Returned: 6/8/95

Scanning Agent Signature: Michael N. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: A. I. Memo 965/CBIP 024

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER: AD-A185607

4. TITLE (and Subtitle): Interaction of Different Modules in Depth Perception: Stereo and Shading

5. TYPE OF REPORT & PERIOD COVERED: Memo -- 5/86-5/87

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Heinrich H. Bulthoff and Hanspeter A. Mallot

8. CONTRACT OR GRANT NUMBER(s): A.I.: N00014-85-K-0124
CBIP: N00014-85-C-0038

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02139

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

12. REPORT DATE: May 1987

13. NUMBER OF PAGES: 23

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office):
Office of Naval Research
Information Systems
Arlington, VA 22217

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Stereo
Depth Perception
Shape from Shading

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A method has been developed to measure the perceived depth of computer-generated images of simple solid objects. Computer graphic techniques allow for independent control of different depth cues (stereo,shading, and texture) and enable the investigator thereby to psychophysically study the interaction of modules for depth perception. Accumulation of information from shading and stereo and vetoing of depth from shading by edge information have been found. Cooperativity and other types of interactions are discussed. If intensity edges are missing, as in smooth-shaded surfaces, the image intensities (cont.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

(1) REPORT NUMBER

(2) GOVT ACCESSION NO.

(3) RECIPIENT'S CATALOG NUMBER

(4) TITLE (and Subtitle)

Interaction of Different Modules in Depth Perception: Stereo and Shading

(5) TYPE OF REPORT & PERIOD COVERED

Memo -- 5/86-5/87

(6) PERFORMING ORG. REPORT NUMBER

(7) AUTHOR(s)

Heinrich H. Bülthoff and Hanspeter A. Mallot

(8) CONTRACT OR GRANT NUMBER(s)

A.I.: N00014-85-K-0124
CBIP: N00014-85-C-0038

(9) PERFORMING ORGANIZATION NAME AND ADDRESS

Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02139

(10) PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

(11) CONTROLLING OFFICE NAME AND ADDRESS

Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

(12) REPORT DATE

May 1987

(13) NUMBER OF PAGES

23

(14) MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

Office of Naval Research
Information Systems
Arlington, VA 22217

(15) SECURITY CLASS. (of this report)

UNCLASSIFIED

(15a) DECLASSIFICATION/DOWNGRADING SCHEDULE

(16) DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

(17) DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(18) SUPPLEMENTARY NOTES

None

(19) KEY WORDS (Continue on reverse side if necessary and identify by block number)

Stereo
Depth Perception
Shape from Shading

(20) ABSTRACT (Continue on reverse side if necessary and identify by block number)

A method has been developed to measure the perceived depth of computer-generated images of simple solid objects. Computer graphic techniques allow for independent control of different depth cues (stereo, shading, and texture) and enable the investigator thereby to psychophysically study the interaction of modules for depth perception. Accumulation of information from shading and stereo and vetoing of depth from shading by edge information have been found. Cooperativity and other types of interactions are discussed. If intensity edges are missing, as in smooth-shaded surfaces, the image intensities (cont.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 6/8/1995

M.I.T. Libraries
Document Services