Communicated by Steven Zucker

A Possible Neural Mechanism for Computing Shape From Shading

Alex Pentland
Vision Sciences Group, E15-410, The Media Laboratory,
Massachusetts Institute of Technology
20 Ames Street, Cambridge, MA 02138, USA

A simple neural mechanism that recovers surface shape from image shading is derived from a simplified model of the physics of image formation. The mechanism's performance is surprisingly good even when applied to complex natural images, and is even able to extract significant shape information from some line drawings.

1 Introduction

Shading is the variation in image intensity due to changes in surface shape, and has long been recognized as one of the most important visual cues to surface shape. Leonardo Da Vinci, for instance, wrote in his notebooks: "Shading appears to me to be of supreme importance in perspective, because, without it opaque and solid bodies will be ill-defined." Despite its importance, however, relatively little is known about how people extract shape from shading.

Perhaps the major obstacle to understanding is the lack of a good theoretical model. For although the physics is well understood, the mathematical problem is so underconstrained that no solution is possible without the use of simplifying assumptions (Horn 1975; Pentland 1984). Examination of the physics shows that there are three types of simplifications that might be useful. These are assumptions about the surface shape (e.g., smoothness), the distribution of illumination (e.g., a single light source direction), or about the reflectance function (e.g., Lambertian reflectance). Of these three categories, assumptions about illumination are the least controversial: almost all research has accepted the hypothesis that people assume a single, distant illuminant within "fairly large" image regions.

Assumptions about surface shape have received the most attention in recent research. There have been two types of simplifying assumptions that have been found useful: (1) smoothness assumptions, first employed by Horn (1975), and (2) assumptions about local surface curvature, first employed by Pentland (1984). The resulting shape-from-shading techniques have estimated surface orientation, so that integration is required to recover depth. Techniques employing smoothness have the additional disadvantage that they require dozens of iterations to converge to an answer. Despite considerable effort neither type of surface assumption has resulted in the sort of fast, locally-accurate recovery of shape typical of human vision.

Research concerning the final type of simplifying assumption — assumptions about the reflectance function — has concentrated on the role of highlights and specularities rather than on the more diffuse type of reflection which dominates in most image regions. In this paper, I develop a new type of shape-from-shading theory based on a simple characterization of this diffuse type of reflection, describe a simple neural mechanism that computes shape from shading, and then evaluate this mechanism on natural imagery.

2 The Imaging of Surfaces

The first step is to recount the physics of how image shading is related to surface shape. Let $z = z(x, y)$ be a surface, and let us assume that the surface is Lambertian and everywhere illuminated by (possibly several) distant point sources. I will also assume orthographic projection onto the x, y plane.

I will let $\vec{L} = (x_L, y_L, z_L) = (\cos\tau \sin\sigma, \sin\tau \sin\sigma, \cos\sigma)$ be the unit vector in the mean illuminant direction, where τ is the *tilt* of the illuminant (the angle the image plane component of the illuminant vector makes with the x-axis) and σ is its *slant* (the angle the illuminant vector makes with the z-axis).

Under these assumptions the normalized image intensity $I(x, y)$ will be

$$I(x, y) = \frac{p\cos\tau \sin\sigma + q\sin\tau \sin\sigma + \cos\sigma}{(p^2 + q^2 + 1)^{1/2}} \tag{2.1}$$

where p and q are the slope of the surface along the x and y image directions respectively, e.g.,

$$p = \frac{\partial}{\partial x} z(x, y) \qquad q = \frac{\partial}{\partial y} z(x, y) \tag{2.2}$$

Equation (2.1) can now be converted to a form which will allow us to relate image and 3-D surface in terms of their Fourier transforms — or in terms of any other convenient set of linear basis functions. This is accomplished by taking the Taylor series expansion of $I(x, y)$ about $p, q = 0$ up through the quadratic terms, to obtain

$$I(x, y) \approx \cos\sigma + p\cos\tau \sin\sigma + q\sin\tau \sin\sigma - \frac{\cos\sigma}{2}(p^2 + q^2) \tag{2.3}$$

Numerical simulation has shown that this expression yields a good approximation of the image intensities except for large values of p or q.

Note that because the Taylor expansion is applied at each image pixel individually, no smoothing or approximation of the surface shape is involved; we are simplifying only the surface's reflectance function.

Now let the complex Fourier spectrum $F_z(f, \theta)$ of $z(x, y)$ be

$$F_z(f, \theta) = m_z(f, \theta)e^{i\phi_z(f,\theta)} \tag{2.4}$$

where $m_z(f, \theta)$ is the magnitude at position (f, θ) on the Fourier plane, and ϕ_z is the phase. Since p and q are partial derivatives of $z(x, y)$, their transforms F_p and F_q are simply related to F_z, for example,

$$F_p(f, \theta) = 2\pi \cos\theta f m_z(f, \theta)e^{i(\phi_z(f,\theta)+\pi/2)} \tag{2.5}$$

$$F_q(f, \theta) = 2\pi \sin\theta f m_z(f, \theta)e^{i(\phi_z(f,\theta)+\pi/2)} \tag{2.6}$$

Under the condition $|p|, |q| < 1$ the linear terms of equation (2.3) will dominate the image intensity function except when the average illuminant is within roughly $\pm 30^\circ$ of the viewers' position. When either $p, q \ll 1$ or the illumination direction is not near the viewing direction the quadratic terms will be negligible.

Thus when either $p, q \ll 1$ or the illumination direction is oblique, so that the quadratic terms are negligible, then the Fourier transform of the image $I(x, y)$ is (ignoring the DC term):

$$F_I(f, \theta) = 2\pi \sin\sigma f m_z(f, \theta)e^{i(\phi_z(f,\theta)+\pi/2)}[\cos\theta\cos\tau + \sin\theta\sin\tau]. \tag{2.7}$$

That is, *the image intensity surface $I(x, y)$ is a linear function of the height surface $z(x, y)$.*[1]

When the quadratic terms of equation (2.3) dominate (e.g., when the illumination and viewing directions are similar) then the relationship between surface and image is substantially more complex. The quadratic terms, p^2 and q^2, are everywhere positive whereas p and q are both positive and negative. Thus the transforms of p^2 and q^2 will show a *frequency doubling* effect, that is, the image $I(x, y)$ of a surface $z(x, y) = \sin(x)$ with $\vec{L} = (0, 0, 1)$ will be (roughly) $I(x, y) = \sin(2x)$. Although the actual situation is considerably more complex than this, the notion of frequency doubling provides a good qualitative description of the contribution of the quadratic terms to the overall image intensity function.

Human Psychophysics: The above mathematics makes it clear that a useful simplifying assumption is that the surface reflectance function is a linear function of surface orientation, for in this case it is possible to obtain a simple linear relationship between surface shape and image intensity. Such an assumption has the additional virtue of being an accurate description of the Lambertian reflectance function under a wide variety of common situations. This assumption has been tested in psychophysical experiments (Pentland 1988), with clear results: the hypothesis of a

[1] It is interesting to note that the moon's surface has exactly this reflectance function (Horn 1975).

linear reflectance function correctly predicts people's judgment of shape, even when people's judgment is incorrect. In order to model human performance our theory will therefore assume that the linear model given by equation (2.7) correctly describes how surfaces are imaged.

3 Recovery of Shape

Examining equation (2.7) shows that if given the illuminant direction then the Fourier transform of the surface can be recovered directly, except for an overall scale factor and the DC term. That is, letting the Fourier transform of the image be

$$F_I(f,\theta) = m_I(f,\theta)e^{i\phi_I(f,\theta)}, \tag{3.1}$$

then the Fourier transform of the z surface is simply

$$F_z(f,\theta) = \frac{m_I(f,\theta)e^{i(\phi_I(f,\theta)-\pi/2)}}{2\pi\sin\sigma f[\cos\theta\cos\tau + \sin\theta\sin\tau]}. \tag{3.2}$$

To achieve reliable shape recovery using equation (3.2) it is necessary to add a stabilizing term to the denominator, so that its magnitude remains greater than approximately $0.5\pi\sin\sigma f$. Edge information often affects the perception of shape from shading; similarly, there are interactions between shading and other shape cues such as stereo. Boundary information derived from, for example, edges, can be inserted by setting the surface shape components along orientations approximately perpendicular to the illuminant direction (for example, by setting the components $F_z(f,\psi)$, where $\psi\perp\tau$) to produce the correct surface shape. Surface shape information, e.g., coarse stereo estimates, can be optimally combined with shading information as follows:

$$F_z(f,\theta) = \frac{\sigma^2_{stereo}F_z^{shading}(f,\theta) + \sigma^2_{shading}F_z^{stereo}(f,\theta)}{\sigma^2_{stereo} + \sigma^2_{shading}} \tag{3.3}$$

where σ^2_{stereo} and $\sigma^2_{shading}$ are the assumed variance of the two estimates of $F_z(f,\theta)$.

3.1 A Neural Mechanism. The ability to recover surface shape by use of equation (3.2) suggests a neural mechanism for the perception of shape from shading. It is widely accepted that the visual system's initial cortical processing areas contain many cells that are tuned to orientation, spatial frequency and phase. Although the tuning of these cells is relatively broad, one can still produce an estimate of shape by summing the output of these cells in a selective manner.

Figure 1 illustrates this mechanism. I assume that some transformation $\vec{T}$ of the image is produced by a set of filters that are *localized* in

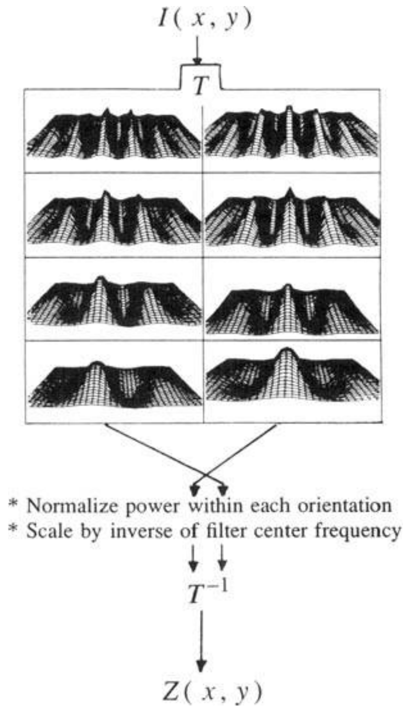

Figure 1: A Shape-from-shading mechanism. A transformation $\vec{T}$ produces localized measurements of sine and cosine phase frequency content, and then the inverse transformation is applied, switching sine and cosine phase amplitudes and scaling the filter amplitude in proportion to the central frequency. The output of this process is the recovered surface shape.

both space and spatial frequency. Such a transformation is widely believed to occur between the retina and striate cortex. Note that these filters are all centered on the same image location, and have the same overall response envelope. I will further assume that these filters exist in quadrature pairs, so that the local phase information is available. Exactly this sort of filter mechanism is central to many recent psychological theories (Adelson and Bergen 1985; Daugman 1980; Adelson et al. 1987).

I will for purposes of discursive clarity assume that these filters are sine and cosine phase Gabor filters, because the output of such a set of Gabor filters is exactly the Fourier transform of the image as seen through a Gaussian-shaped windowing function and thus the above equations are directly applicable. In order to recover surface shape from this filter set, the transformations indicated in equation (3.2) must be performed. These transformations are (1) phase-shift the filter responses by $\pi/2$, (2) scale the filter amplitude by $1/f$, where f is the filter's central spatial frequency, (3) bias the filters to remove variation due to illumination direction, and (4) reconstruct a depth surface from the scaled amplitudes of the filter set.

The final step, reconstruction, can be accomplished by a process nearly identical to that by which one would reconstruct the original signal, that is, by summing Gabor filter basis functions with an amplitude proportional to the Gabor filter's activity. This reconstruction process is indicated by the transformation $\vec{T}^{-1}$ in figure 1; it is the inverse of what is believed to happen between the retina and striate cortex.

The only differences between the shape recovery process and simple reconstruction are (1) the role of the sine and cosine phase filters are switched, so that cosine phase functions are added together with amplitude proportional to the response of the sine phase filters, and *vice versa*. This accomplishes a $\pi/2$ phase shift. (2) Each filter's amplitude is reduced in proportion to its central frequency, thus accomplishing the $1/f$ frequency scaling. (3) The average filter amplitude is normalized within each orientation. This normalization removes the directional biasing effects of the illuminant.

The result of this scaled summation will be the estimated surface shape within the windowed area of the image, that is, within the "receptive field" covered by the filters. If this theory were construed as applying to near parafoveal receptive fields of the Macaque monkey, then the patch over which shape is recovered would cover an area roughly the size of the monkey's fully extended hand. It is possible to smoothly link adjacent patches together to produce an extended surface, however a better alternative is to combine this small-scale shading information with large-scale structure that can come from other, more suitable sources such as stereo or motion.

Relation to Biology: In an actual biological implementation use of a Gabor filter transform $\vec{T}$ would be a poor choice because $\vec{T} \neq \vec{T}^{-1}$, so that two separate sets of filters are required, and because for Gabor filters calculation of $\vec{T}^{-1}$ requires a level of numerical accuracy inappropriate for biological mechanisms. In a biological mechanism one would expect an orthonormal filter set (which can be visually similar to Gabor filters) so that $\vec{T} = \vec{T}^{-1}$ and so that precise calculations are not required (Adelson et al. 1987). Further, the use of exactly sine and cosine phase filter pairs is unnecessary, as all the phase information is available from any two filters of different phase. When using filters of arbitrary phase one feeds a weighted average of the two input filters to the reconstruction filters rather than feeding the sine/cosine phase input filters exclusively to the cosine/sine phase reconstruction filters.

3.2 Surface Recovery Results. Having developed a theory and mechanism, it is time to evaluate the performance of that mechanism on real-world problems. This section and the following section present several

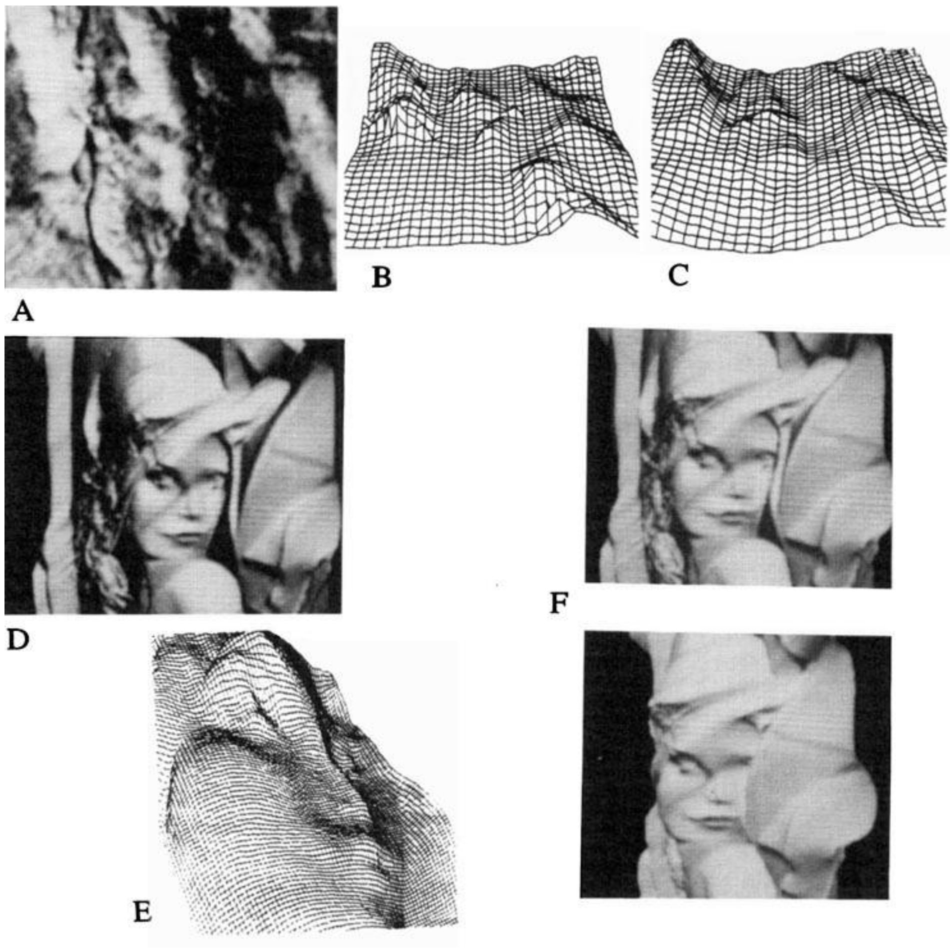

Figure 2: (a) An image of a mountainous region outside of Phoenix, Arizona, (b) a perspective view of a stereo-derived surface shape, and (c) a perspective view of the surface shape derived from image shading, (d) an image widely used in image compression research, (e) a perspective view of the recovered surface in the neighborhood of the woman's face, (f) two shaded views of the recovered surface; note the accurate recovery of eye, cheek, lip, nostril, and nose arch shape.

examples of using this approach to recover shape from shading. In these examples the illuminant direction was estimated from the Fourier transform of the image, as described in reference (Pentland 1988).

Figure 2a shows a high-altitude image of a mountainous region outside of Phoenix, Arizona. The Defense Mapping Agency has created a elevation map of this region using their interactive stereo system; figure 2b shows a perspective view of this stereo-recovered surface. Figure 2c shows a perspective view of the surface shape derived from the shading information in figure 2a. Comparisons of the stereo-derived surface (Fig. 2b) to the shading-derived surface (Fig. 2c) demonstrate that the recovery of shape from shading in this example is quite accurate. The major defect of the recovered surface shape is a low-frequency bowing of the entire surface, which appears to stem from slow variations in average surface reflectance and illumination direction.

A second example of shape recovery is shown in figures 2d and 2e. Figure 2d shows an image widely used in image compression research. Figure 2e shows a perspective view of the recovered surface in the neighborhood of the face. Figure 2f shows shaded versions of the recovered surface from viewpoints 30 and 45 degrees from the original image's viewpoint. The shape of the eyes, cheek, lips, the nose arch and nostrils are all recovered accurately.

3.2.1 Line Drawings

The role shading information has in interpreting line drawings has long been debated. To investigate this issue we applied our shape-from-shading technique to several line drawings. One example, shown in figure 3a, is a line drawing of a famous underground cartoon character named Zippy. Surface shape for a digitized version of this drawing was estimated using our shape-from-shading mechanism. A perspective view of the face region of the recovered surface is shown in figure 3b. Figure 3c shows shaded versions of the recovered surface from several points of view. As can be seen, the shape of the forehead, eyes, cheeks, shirt collar, and mouth are all recovered in a way that agrees closely with our perceptions.

4 Summary

I have shown that a simple neural mechanism based on the assumption of a linear model surface reflectance function allows shape information to be extracted from image shading in a robust and accurate manner. The filters used in this neural mechanism are similar to those that are believed to occur in biological visual systems, and the mechanism itself is a simple modification of decomposition-reconstruction networks pro-

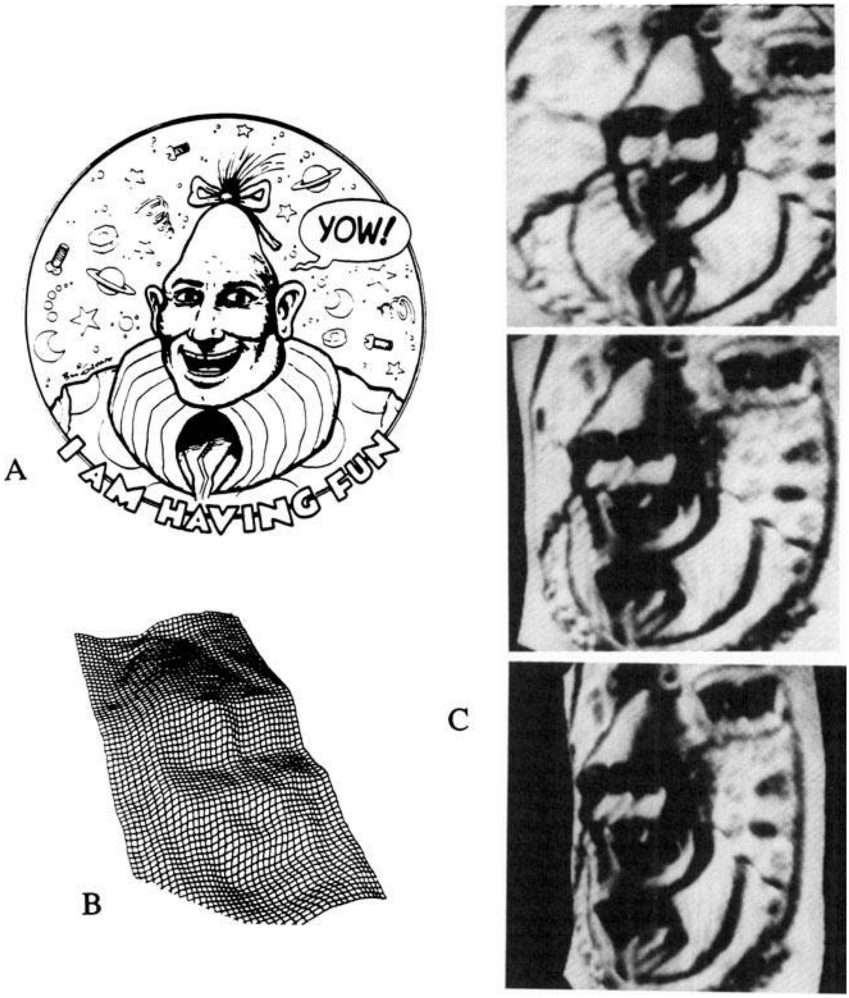

Figure 3: (a) A line drawing of a famous cartoon character, (b) a perspective view of the surface shape recovered by a shading analysis covering the face region, (c) several shaded views of the recovered surface.

posed for other types biological visual processing (Adelson and Bergen 1985; Daugman 1980; Adelson et al. 1987).

This shape-from-shading mechanism fails to give a good answer when the illumination is from behind the viewer, and when either the surface reflectance or the illumination changes sharply. However when the illumination is from behind the viewer (so that the mechanism proposed here fails) then image brightness is closely correlated with surface orientation so that very simple rules can be used to interpret the image shading (Pentland 1984). In either case practical extraction of shape information requires preprocessing to segment the image into regions of constant reflectance and illumination.

Perhaps the most surprising aspect of this shape-estimation mechanism is its performance when applied to line drawings. In the cases examined so far, it appears that a substantial amount — and in some cases *most* — of the 3-D information may be recovered by this simple shading analysis.

Acknowledgments

This research was made possible by National Science Foundation, Grant No. DAAL 03-87-K-0005. I wish to thank Berthold Horn and Ted Adelson for their comments and insights.

References

Adelson, E. and J. Bergen. 1985. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, **2(2)**, 284–299.

Adelson, E.H., E. Simonelli, and R. Hingorani. 1987. Orthogonal pyramid transforms for image coding. *SPIE Proceedings on Visual Communications and Image Processing II*, SPIE **845**, 50–58, Oct. 27–29, 1987, Boston, MA.

Daugman, J. 1980. Two-dimensional analysis of cortical receptive field profiles. *Vision Research*, **20**, 846–856.

Horn, B.K.P. 1975. Obtaining shape from shading information. *In:* The Psychology of Computer Vision, ed. P.H. Winston. McGraw-Hill.

Pentland, A. 1988. Shape from shading: A theory of human perception. *Proceedings of International Conference on Computer Vision*, IEEE Society, Dec. 5–8, Tarpon Springs, FL.

———. 1984. Local analysis of the image. *IEEE Transactions on Pattern Analysis and Machine Recognition*, 170–187.

Received 10 November; accepted 14 November 1988.

This article has been cited by:

1. José R.A. Torreão, João L. Fernandes. 2011. Linear-nonlinear neuronal model for shape from shading. *Pattern Recognition Letters* **32**:9, 1223-1239. [CrossRef]
2. Z LIU, J GASKA, L JACOBSON, D POLLEN. 1992. Interneuronal interaction between members of quadrature phase and anti-phase pairs in the cat's visual cortex. *Vision Research* **32**:7, 1193-1198. [CrossRef]