International Journal of Computer Vision, 4, 153–162 (1990)

Linear Shape From Shading

ALEX P. PENTLAND
Vision and Modeling Group, The Media Lab, Massachusetts Institute of Technology, Room E15-387, 20 Ames St., Cambridge, MA 02138

Abstract
In many situations the reflectance function of a surface is approximately linear, and there is an efficient closed-form solution to the shape-from-shading problem. When boundary conditions (e.g., edges, singular points) are not available, good estimates of shape may still be extracted by using the assumption of general viewing position. An improved method for estimating the illuminant direction is also presented.

1 Introduction

The extraction of shape from shading has a relatively long history within the field of computer vision. There have been two general classes of algorithm developed: global algorithms, which propagate information across a shaded surface starting from points with known surface orientation, and local algorithms, which attempt to estimate shape from local variations in image intensity.

Global algorithms, primarily due to Horn and his students [5, 6, 7], use assumptions about surface shape—primarily that the surface is smooth in some sense—in order to extract estimates of surface orientation. The smoothness assumption is used to relate adjoining points, enabling spatially isolated information about absolute surface orientation (which must be derived using some other technique) to be iteratively propagated across the surface. The use of a smoothness assumption, however, implies that the algorithms will not produce exact solutions except under certain conditions [8]. A subsequent integration step is normally required to convert estimated surface orientation into an estimate of surface shape.

Local algorithms, originally suggested by Pentland [1], also use assumptions about surface shape in order to extract estimates of surface orientation from the shading information within a small image neighborhood. As with the global estimation algorithms, integration is normally required to obtain the surface shape. These local methods of estimating surface orientation have been shown capable of producing good estimates of shape [1, 2, 3], however, they do not produce exact estimates except in quite limited situations [2, 4].

In this article I develop a new type of shape-from-shading algorithm, one that uses assumptions about the reflectance function, rather than using assumptions about the surface shape. The basic idea of this algorithm is to construct a linear approximation to the true reflectance function, thus permitting an extremely efficient closed-form solution for the surface shape. One important characteristic of the solution is that good estimates of shape can be obtained even when boundary conditions are not available, by use of the assumption of general viewing position.

2 The Imaging of Surfaces: Linear Reflectance Functions

The first step is to review the physics of how image shading is related to surface shape. As an example, let us start by considering a distantly illuminated surface whose shape is defined by the function $z = z(x, y)$, and with reflectance function $I(x, y) = R(p, q, \mathbf{L})$ where p and q are the slope of the surface along the x and y image directions respectively, for example,

$$p = \frac{\partial}{\partial x} z(x,y) \qquad q = \frac{\partial}{\partial y} z(x,y) \tag{1}$$

and $\mathbf{L} = (x_L, y_L, z_L)$ is a unit vector in the mean illuminant direction. For mathematical simplicity, I will also assume orthographic projection onto the x, y plane, that the surface is not self-shadowing, and that $z < 0$ within the region of interest.

Given an image region with mean surface orientation (p_0, q_0), we can form a linear approximation to the reflectance function by taking a Taylor expansion of R around $(p, q) = (p_0, q_0)$ [9, 12],

$$I(x, y) \approx R(p_0, q_0, \mathbf{L}) + (p - p_0)\left.\frac{\partial R(p, q, \mathbf{L})}{\partial p}\right|_{p=p_0, q=q_0} + (q - q_0)\left.\frac{\partial R(p, q, \mathbf{L})}{\partial q}\right|_{p=p_0, q=q_0} \quad (2)$$

Given a point on a smooth surface it is always possible to choose a neighborhood over which this approximation is quite accurate, because one can always find a neighborhood that contains a small range of (p, q) values, as is illustrated in figure 1a. Over a sufficiently small range of (p, q) values the true reflectance function $R(p, q, \mathbf{L})$ can always be accurately approximated by equation (2), as is illustrated by figure 1b.

Further, when the illuminant is at a large angle to the viewer (as in figure 1b) equation (2) provides a good approximation to the true reflectance function over *most* of the range of p and q. In general the range of (p, q) that can be accurately modeled by a single linear approximation becomes larger as the illuminant becomes more oblique, and smaller as the illuminant moves closer to the viewer.

As an example, for a Lambertian reflectance function and $(p_0, q_0) = (0, 0)$, the linear approximation to image intensity $I(x, y)$ will be

$$I(x, y) \approx \rho\lambda[\cos\sigma + p\cos\tau\sin\sigma + q\sin\tau\sin\sigma] \quad (3)$$

where ρ is the albedo of the surface, λ is the strength (flux density) of the illuminant at the surface, 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), so that $\mathbf{L} = (x_L, y_L, z_L) = (\cos\tau\sin\sigma, \sin\tau\sin\sigma, \cos\sigma)$. When $\sigma > 45°$ (i.e., the illuminant is more than 45 degrees from the viewer) this approximation to the Lambertian reflectance function is accurate to within 10% for $-0.2 < p, q < 0.2$, a range typical of mountainous terrain in aerial imagery. It is accurate over the range $-1 < p, q < 1$ when $\sigma > 75°$.

Thus, within a constrained region it is often the case that the true surface reflectance function can be accurately approximated by a linear reflectance function. This can allow us to use specialized algorithms for solving the shape-from-shading problem that are much more efficient than techniques that solve the problem in complete generality. It can also free us from the need to know the surface's reflectance function, as we merely need to have a sufficiently restricted range of (p, q) or sufficiently oblique illumination for the linear approximation to be a good one.

3 Shape Recovery

Given a linear reflectance function, we have that

$$I(x, y) = k_1 + pk_2 + qk_3 \quad (4)$$

for some constants k_1, k_2 and k_3. I will refer to the vector (k_2, k_3, k_1) as the "generalized illuminant direction,"

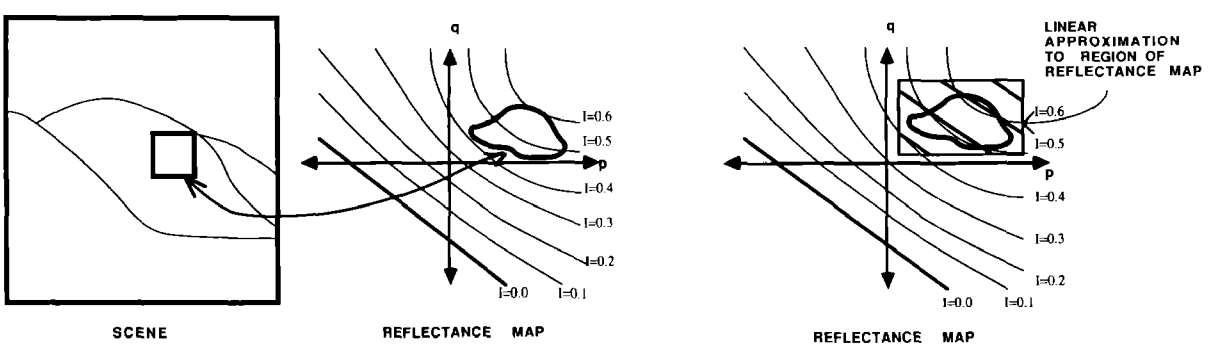

Fig. 1. (a) For smooth surfaces, a small image patch corresponds to a small range of (p, q) values. (b) Over a small range of (p, q) values, there is always a good linear approximation to the true reflectance function.

by analogy to the case of a Lambertian surface with mean normal $(p_0, q_0) = (0, 0)$ where we have that

$$k_1 = \cos \sigma$$
$$k_2 = \cos \tau \sin \sigma$$
$$k_3 = \sin \tau \sin \sigma \quad (5)$$

and so $\mathbf{L} = (k_2, k_3, k_1)$.

Equation (4) may be transformed into the Fourier domain in order to obtain a convenient and efficient solution. Let $F_z(f, \theta)$ be the complex Fourier spectrum of $z(x, y)$ (where f is radial frequency and θ is orientation), then, because p and q are the partial derivatives of $z(x, y)$, their Fourier transforms $F_p(f, \theta)$ and $F_q(f, \theta)$ are simply

$$F_p(f, \theta) = 2\pi \cos(\theta) f e^{i\pi/2} F_z(f, \theta) \quad (6)$$

$$F_q(f, \theta) = 2\pi \sin(\theta) f e^{i\pi/2} F_z(f, \theta) \quad (7)$$

Equation (4) can now be rewritten in the Fourier domain as

$$F_I(f, \theta) = H(f, \theta) F_z(f, \theta) \quad (8)$$

where $F_I(f, \theta)$ is the Fourier spectrum of the image and $H(f, \theta)$ is a linear transfer function which relates the Fourier transform of the image to that of the surface. Ignoring the singular DC term, $H(f, \theta)$ is simply

$$H(f, \theta) = 2\pi f e^{i\pi/2} [k_2 \cos \theta + k_3 \sin \theta] \quad (9)$$

Thus given a linear reflectance function, the surface shape can be estimated in closed form by use of the inverse transfer function, $H^{-1}(f, \theta)$,

$$F_z(f, \theta) = H^{-1}(f, \theta) F_I(f, \theta)$$
$$= (2\pi f e^{i\pi/2} [k_2 \cos \theta + k_3 \sin \theta])^{-1} F_I(f, \theta) \quad (10)$$

Assuming that the generalized illuminant direction is known, equation (10) may be used to recover the Fourier components of the surface shape. If only the ratio k_2/k_3 is known (i.e., the tilt component of the generalized illuminant direction), the surface's Fourier components may be recovered up to a multiplicative constant. Estimation of the generalized illuminant direction is discussed in section 3.1.

3.0.1 Boundary Conditions. The Fourier components of the surface that are exactly perpendicular to the illuminant cannot be seen in the image data, and must either be obtained from other information sources or simply set to some default value. When boundary conditions—information about surface shape from contours, singular points, or other sources—are available, they can be used to determine the Fourier components perpendicular to the illuminant; in this manner an exact recovery of surface shape can be achieved.

When boundary conditions are not available (as is often the case), the assumption of general viewing position may be invoked to argue that these unseen Fourier components should be assumed to be zero, because if they were large then small variations in viewing geometry would produce large changes in the estimated surface shape. I have found that in practice these default boundary conditions produce good estimates of shape whenever the surface is complex and irregular, however for regular geometric forms the estimated surface shape can be substantially in error.

3.0.2 Noise Sensitivity. The recovery process can be improved by use of Weiner filtering to remove noise and nonlinear components of the image intensity pattern [10]. If the contaminating noise $N(f, \theta)$ is modeled as being proportional to $\|(k_2 \cos \theta + k_3 \sin \theta)\|$ (for example, as a fixed fraction of the spectral power along each image orientation), and the surface $S(f, \theta)$ is modeled as a fractal Brownian function [11, 12] whose power spectrum is proportional to f^{-4} (or, equivalently, as a second-order Markov random field or as a "thin-plate" model) then the optimal RMSE estimate of surface shape is

$$F_z(f, \theta) = H^{-1}(f, \theta) \|H(f, \theta)\|^2 \left[\|H(f, \theta)\|^2 + \frac{\|N(f, \theta)\|}{\|S(f, \theta)\|}\right]^{-1} F_I(f, \theta)$$
$$= \{2\pi \sin(\sigma) f e^{i\pi/2} [sd + (k_2 \cos \theta + k_3 \sin \theta)]\}^{-1} F_I(f, \theta) \quad (11)$$

where $s = \text{Sign}[\cos(\tau - \theta)]$ and $0.5 < d < 0.75$. In actual practice equation (11) has been found to perform much better than equation (10).

3.1 Estimating the Illuminant Direction

Pentland [13] introduced a method of estimating illuminant direction from the distribution of image derivatives as a function of image direction. The method works by assuming a statistically uniform distribution of surface orientations, and then performing a maximum-likelihood analysis to estimate the cosine variation in image gradient magnitude induced by the directionality of the illuminant. In summary, the result is that

$$(x_L^*, y_L^*) = (\beta^T\beta)^{-1}\beta^T(dI_1, dI_2, \ldots, dI_n) \quad (12)$$

where (x_L^*, y_L^*) are the unnormalized x and y components of the illuminant direction, β is a $2 \times n$ matrix of directions (dx_i, dy_i) and dI_i is the mean magnitude of $dI(x, y)/dx_i + dI(x, y)/dy_i$.

Given (x_L^*, y_L^*), the complete illuminant direction is simply

$$x_L = x_L^*/k, \quad y_L = y_L^*/k, \quad z_L = \sqrt{1 - x_L^2 - y_L^2}, \quad (13)$$

where

$$k = \sqrt{E(dI^2) - E(dI)^2} \quad (14)$$

and $E(dI)$ is the expected value of $dI/dx_i + dI/dy_i$ over all directions i.

This method has proven to be quite robust [2, 3, 13, 14], however the assumption of uniformly distributed surface orientations is disagreeably strong. This method can be substantially improved by observing that the illuminant produces a similar effect *in each frequency band*. Thus, if I make the much weaker assumption that the power in a particular spatial frequency band is uniformly distributed over orientation—or, more precisely, is not distributed in a way that is correlated with the illuminant effects—then I can use a similar method to estimate the illuminant direction, substituting the magnitude of the Fourier components for magnitude of the first derivatives. In particular, equation (12) becomes

$$(x_L^*, y_L^*) = (\beta^T\beta)^{-1}\beta^T(m_1, m_2, \ldots, m_n) \quad (15)$$

where the m_i are the magnitude of the Fourier components within the selected frequency band in direction (dx, dy).

When applied to an image region, this technique produces an estimate of what I have called the *generalized illuminant direction*, that is, an estimate of the orientation and magnitude of illumination effects within the region. For the purposes of shape recovery one can use equation (15) to determine (x_L^*, y_L^*), and then set $k_2 = x_L^*$, $k_3 = y_L^*$. The surface shape can then be estimated up to an overall multiplicative ambiguity.

3.2 A Biological Mechanism

The ability to recover surface shape by use of equation (11) suggests a parallel filtering mechanism for recovering shape from shading. Such a mechanism may be relevant to biological vision, as it is widely accepted that early stages of the human visual system can be regarded as being composed of filters tuned to orientation, spatial frequency, and phase [15, 16, 17]. Figure 2

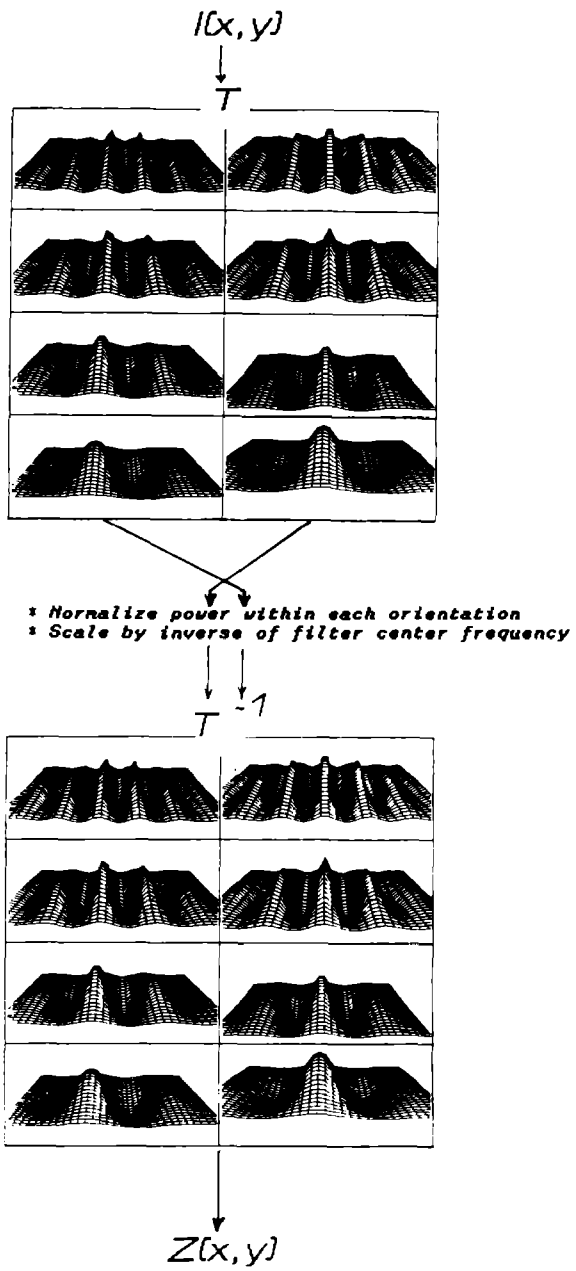

Fig. 2. A shape-from-shading mechanism: A transformation **T** produces localized measurements of sine and cosine phase frequency content; and then the inverse transformation is applied, switching sine and cosine phase amplitudes and scaling the filter amplitude in proportion to the central frequency. The output of this process is the recovered surface shape.

illustrates a mechanism based on filters with similar characteristics. The transformation **T** is a decomposition of the image using filters that form an orthonormal basis set and which are localized in both space and spatial frequency.

In order to recover surface shape from this filter set, the transformations indicated in equation (11) must be performed, as indicated in figure 2. These transformations are: (1) phase-shift the filter responses by $\pi/2$, accomplished by switching the outputs of the sine and cosine phase filters; (2) scale the filter amplitude by $1/f$, where f is the filter's central spatial frequency; (3) normalize average filter responses within each orientation to remove the illumination's directional bias; and (4) reconstruct an elevation surface from the scaled amplitudes of the filter set. The final step, reconstruction, can be accomplished by passing the signal through a second, identical set of filters. This produces the estimated surface shape within the windowed area of the image (the "receptive field" of the filters). For more detail see reference [14].

4 Surface Recovery Results

I have applied equation (11) to both synthetic images of complex surfaces, such as is shown in figure 3a (this is a fractal Brownian surface with $D = 2.3$; max $(p, q) \approx 5.0$), as well as to complex natural images such as shown in figures 4 through 7. In these examples no knowledge of boundary conditions was employed; instead, the assumption of general viewing position was used to obtain default boundary conditions as described above.

4.1 Synthetic Imagery

The use of synthetic imagery is necessary to answer the two important questions concerning this method: One, is the Taylor series approximation a good one; and two, is the recovery stable and accurate? Figure 3b shows the distribution of intensity values obtained when the surface of figure 3a is illuminated from $\mathbf{L} = (1, 1, 1)/\sqrt{3}$. Figure 3c shows the distribution of errors

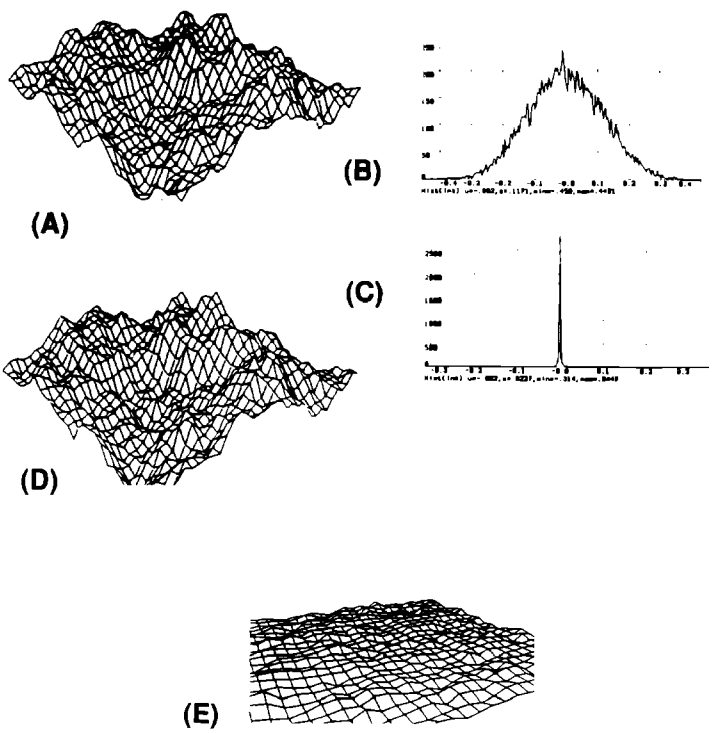

Fig. 3. A fractal Brownian surface. (b) The distribution of intensities within the image of the surface in (a). (c) The distribution of differences between the image and our linear-term-only Taylor series approximation. (d) The surface recovered from shading (compare to (a)). (e) The errors in the recovery process.

between the full imaging model and the Taylor series approximation using only the linear terms. As can be seen, the approximation is a good one, even though this surface is often steeply sloped (for example, max $(p, q) \approx 5.0$).

Figure 3d shows the surface recovered by use of equation (11). Figure 3e shows the differences between the original surface and the recovered surface. As can be seen, the recovery errors are uniformly distributed across the surface. These errors have a standard deviation that is approximately 5% of the standard deviation of the original surface. It appears that these errors can be attributed to the linear approximation breaking down for steeply sloped regions of the surface.

4.2 Natural Imagery

The first example using natural imagery is one where the true reflectance function is already nearly linear, so that we can expect accurate surface recovery. Figure 4a shows an image of a plaster cast of a nickel, together with a range image showing the surface shape extracted from this image by the shape-from-shading mechanism. As part of the recovery process, the illuminant direction was estimated from the Fourier transform of the image, as described above.

Because most people find it difficult to interpret range images I have also illustrated the recovered surface shape in another manner, as shown in figure 4c. This image is generated by using standard computer graphics techniques to first render a shaded, perspective view of the recovered surface shape, and then to project straight lines onto the surface. These straight lines are seen as bending surface contours in the resulting image. This image gives most viewers an accurate impression of the actual recovered surface shape.

By comparing figure 4c to a real nickel the reader can determine that the surface recovery is in fact quite

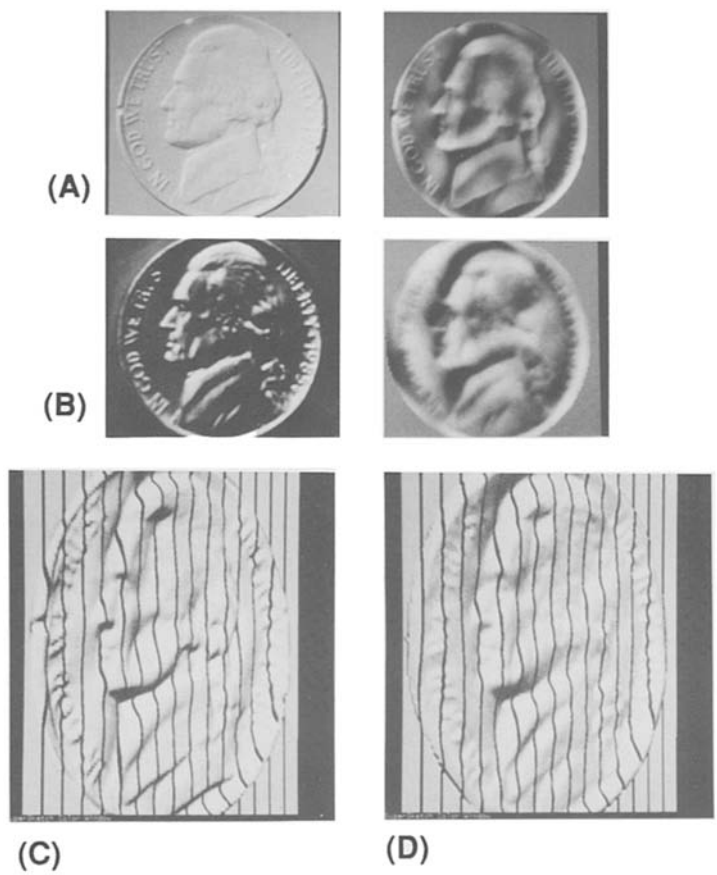

Fig. 4. (a) Image of a plaster cast of a nickel and the extraced range image. (b) Image of a shiny new nickel and the extracted range image. (c) A shaded perspective view of the surface extracted from the plaster cast image. (d) A shaded perspective view of the surface extracted from the shiny nickel image.

accurate. The main defect is that the areas surrounding the head are not sufficiently flat.

Figure 4b shows a second example of recovering surface shape, this time using an image of the very shiny metal surface of a new nickel. In this case we can expect surface recovery to be somewhat less accurate, as the surface's reflectance function is quite nonlinear. Figure 4d shows the surface recovered from figure 4b. As expected the surface recovery is somewhat less accurate than when a diffusely reflecting plaster nickel was used, however, the differences between the two examples are surprisingly small. This example shows the ability of this linear shape-from-shading mechanism to deal with a wide range of reflectance functions, given only that the range of (p, q) is small or that the illumination is sufficiently oblique.

A third example of shape recovery is shown in figure 5. Figure 5a shows a bas-relief sculpture from the New York Metropolitan Museum of Art. This example was chosen to illustrate the effect of variable surface albedo (average reflectance) on the recovery process, as there is significant darkening of the surface in the lower left and upper right corners due to surface dirt. Because the shape-from-shading mechanism does not have input from color or other reflectance mechanisms, it will misinterpret these changes in surface reflectance as changes in shape. Figures 5b and 5c show two shaded perspective views of the recovered surface. It can be seen that although the surface shape recovery is generally accurate, in the lower left and upper right corners there are bulges that are due solely to changes in surface albedo.

The fourth example, shown in figure 6a, also has variable surface albedo. It is of a mountainous region outside of Phoenix, Arizona, that has been the subject of intensive study so that it is possible to compare our shape-from-shading algorithm to results obtained using stereopsis. In particular, the Defense Mapping Agency

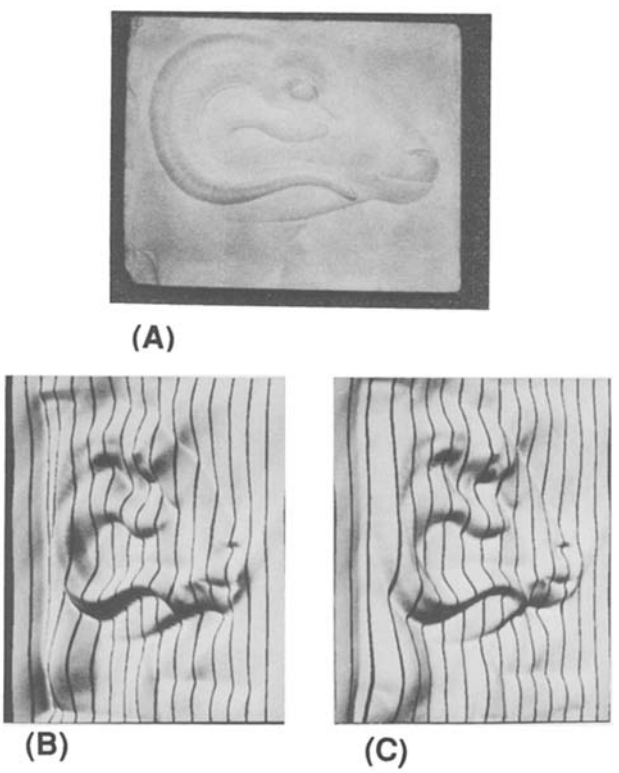

Fig. 5. (a) An image of a bas-relief sculpture of a ram's head, from the New York Metropolitan Museum of Art. (b), (c) Shaded perspective views of the recovered surface.

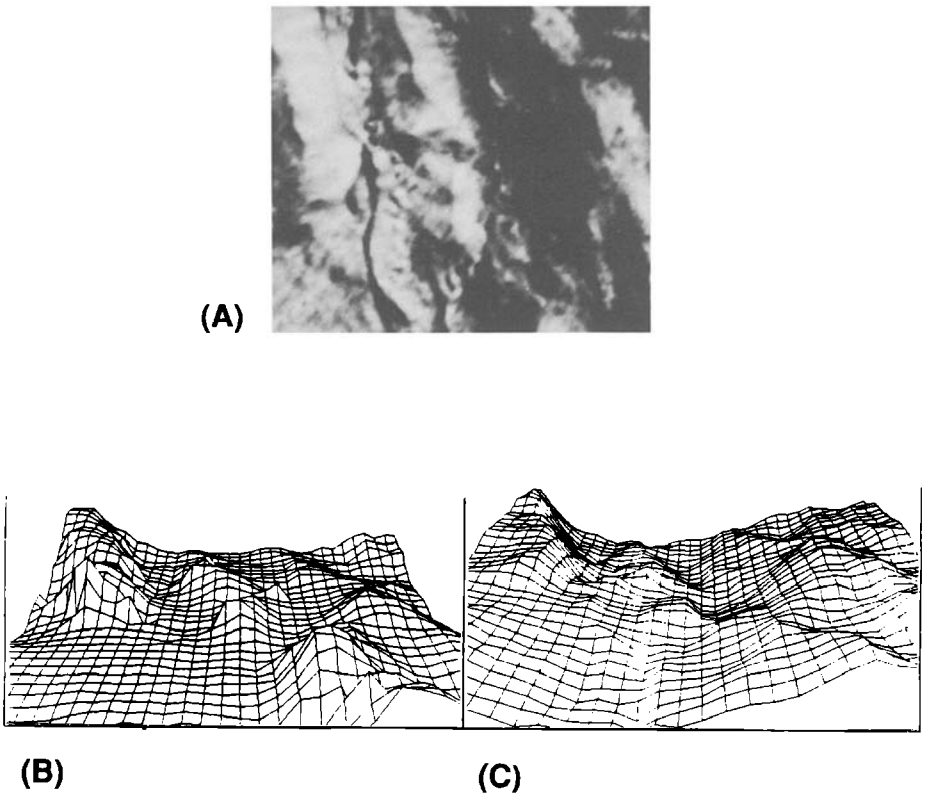

Fig. 6. (a) An image of a mountainous region outside of Phoenix, Arizona. (b) A perspective view of a digital elevation map of this region, obtained from a stereo pair by the Defense Mapping Agency. (c) A perspective view of the elevation map recovered from shading information alone, by use of equation [11].

has created a digital elevation map of this region using their interactive stereo system. A perspective view of this stereo elevation map is shown in figure 6b.

Figure 6c shows a perspective view of the elevation map recovered from the shading information in figure 6a, by use of equation (11). The accuracy of shape recovery for this image can be assessed by comparing the perspective view of the stereo-derived surface (figure 6b) with that of the shading-derived surface (figure 6c). It can be seen that although the shading-derived elevation surface displays a pronounced low-frequency distortion (presumably due to variations in the surface albedo), the details of the recovered surface are still fairly accurate.

A final example of shape recovery is shown in figure 7. Figure 7a is a complex image widely used in image compression research. Figures 7b and c show two shaded perspective views of the recovered surface in the neighborhood of the face. The eyes, cheek, chin, lips, nose, and nostrils can all be clearly seen in the recovered surface, and are generally correct. The wavy, dark area in the lower right is a small portion of the woman's hair.

5 Summary

Often the true reflectance function within an image region can be accurately approximated by a linear function of p and q. In such cases there is a simple closed-form expression that relates surface shape to image intensity. This result may be especially useful in applications where near-real-time performance is required. Further, because the technique can be implemented by use of linear filters similar to those thought to exist in biological visual systems, it may serve as a model for human perception. Experimental results indicate that the recovery process is stable and can be quite accurate.

Special aspects of this approach are that it makes no assumption about surface smoothness or shape, and that it does not require (but can make use of) boundary conditions to obtain an estimate of shape. To avoid requiring known boundary conditions I have used the assumption of general viewing position to fill in missing boundary conditions with default values. The use of these default boundary conditions seems to produce the most accurate shape estimates for complex, highly textured surfaces. When boundary conditions are

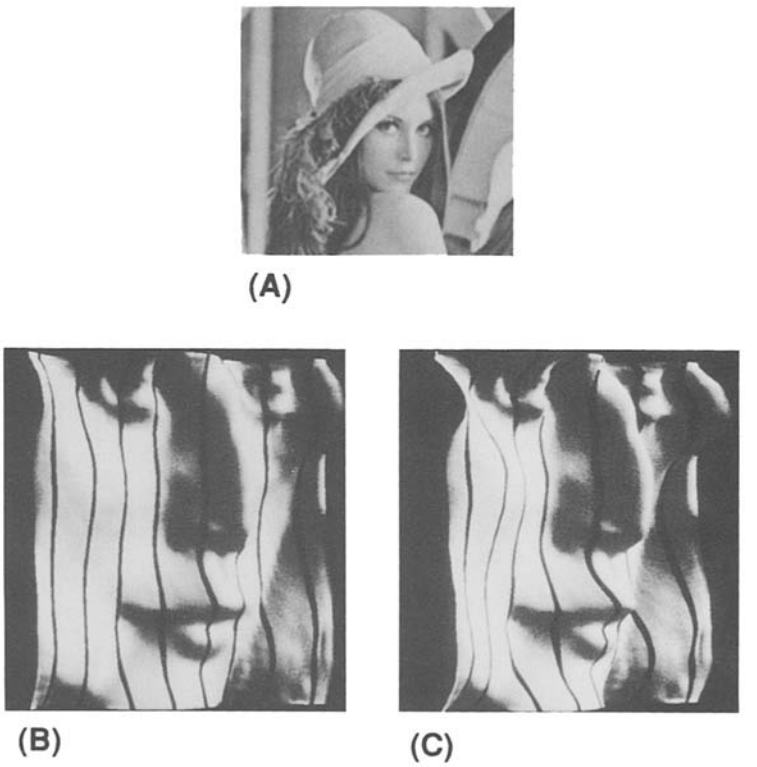

Fig. 7 (a) An image widely used in image compression research. (b), (c) Two shaded perspective views of the recovered surface.

available, of course, they can be directly incorporated into the shape estimate as described above.

I believe that this approach to shape-from-shading is ideally suited to the recovery of detailed surface shape within relatively small image regions, a task that is difficult to perform using other cues such as stereo or motion. Because there is no smoothness assumption, the technique can be directly applied to complex (but still continuous) natural surfaces such as hair, cloth, or mountains. One natural method of integrating coarse stereo (or motion) information with this shape-from-shading technique would be to combine their shape estimates in the frequency domain, weighting the stereo information most heavily in the low frequencies and the shading information most heavily in the higher frequencies.

Acknowledgments

This research was made possible by National Science Foundation, Grant No. IRI-87-19920. I wish to thank Berthold Horn, Ted Adelson, and David Heeger for their comments and insights.

References

1. A.P. Pentland, "Local analysis of the image," *IEEE Trans. Pattern Anal. Mach. Recog.* 6(2):170–187, 1984.
2. F.P. Ferrie, and M.D. Levine, "Where and why local shading works," *IEEE Trans. Pattern Anal. Mach. Recog.* 17(1), 1989.
3. D. Knill, and D. Kersten, "Learning a near-optimal estimator for surface shape from shading, *Comput. Vision, Graph., Image Process.* to appear, 1989.
4. G.B. Smith, "Shape from shading: An assessment," *SRI AI Center Tech. Note 287*, SRI International, Menlo Park, CA, 1983.
5. B.K.P. Horn, "Understanding image intensities," *Artificial Intelligence* 8(2): 210–231, 1977.
6. K. Ikeuchi and B.K.P. Horn, "Numerical shape from shading and occluding boundaries," *Artificial Intelligence* 17:141–185, 1981.
7. B.K.P. Horn, and M.J. Brooks, "The variational approach to shape from shading," *Comput. Vision, Graph., Image Process.* 33:174–208, 1986.

8. G.B. Smith, Personal communication.
9. T. Simchony and R. Chellappa, "Direct analytical methods for solving Poisson equations in computer vision problems," *IEEE Comput. Vision Workshop*, Miami Beach, FL, December 1987.
10. E. Adelson and A. Pentland, "Weiner filtering to improve shape from shading," In preparation, 1989.
11. A. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Pattern Anal. Mach. Recog.* 6(6):661–674, 1984.
12. P. Kube and A. Pentland, "On the imaging of fractal surfaces," *IEEE Trans. Pattern Anal. Mach. Vision* 10(5):704–707, 1988.
13. A.P. Pentland, "Finding the illuminant direction," *J. Opt. Soc. Amer.* 72(4):448–455, 1982.
14. A.P. Pentland, "Shape information from shading: A theory of human perception," *Proc. 2nd Intern. Conf. Comput. Vision*, pp. 404–413, 448–455, Tampa, FL, December 5–8, 1988.
15. J. Daugman, "Two-dimensional analysis of cortical receptive field profiles," *Vision Research* 20:846–856, 1980.
16. E. Adelson and J. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Amer. A* 2(2):284–299, 1985.
17. A. Watson and A. Ahumada, "Model of human visual-motion sensing," *J. Opt. Soc. Amer. A* 2(2):322–342, 1985.