On The Extraction Of Shape Information From Shading

Alex Pentland

Vision Science Group, The Media Lab,
Massachusetts Institute of Technology Room
E15-410, 20 Ames St., Cambridge MA 02138

We present a closed-form solution to the problem extracting shape information from image shading, given standard assumptions and oblique illumination. Neither integration nor iterative propagation of information is required. An improved method for estimating the illuminant direction is also presented.[1]

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: Local algorithms, which attempt to estimate shape from local variations in image intensity, and information propagation algorithms, which attempt to propagate contour information across a shaded surface.

Local algorithms, such as Pentland [1] or Ferrie and Levine [2], use shading information within a small region to estimate surface orientation. Thus a subsequent integration step is required to obtain surface shape. These local methods of estimating surface orientation have been shown [2,3] to produce accurate estimates whenever $z(x,y)$, the imaged surface, has derivatives that obey:

$$\frac{z_u^2 - z_v^2}{z_u z_v} = \frac{z_{uu} - z_{vv}}{z_{uv}} \qquad (1)$$

Examples of surfaces which satisfy this condition everywhere are surfaces of revolution whose axis is parallel to the z axis and cylinders whose axis lays in the image plane [4], however on more general surfaces this condition holds only at infrequent, isolated points

The global algorithms, principally developed by Horn and his students [5], makes use a smoothness assumption to relate adjoining points. This enables the strong information available at smooth occluding contours to be iteratively propagated across the surface. The assumption of smoothness is perhaps the primary limitation to the applicability of these algorithms; for instance, the smoothness constraint adopted in [5] implies that the algorithm will converge to the correct surface only when the condition in Eqn.(1) holds [6]. Integration is normally required to obtain the surface shape.

In this paper we develop a novel formulation of the shape-from-shading problem, one that permits a direct, closed-form solution for the surface. Neither integration nor iterative propagation of information is required to solve for the height field, however low-frequency shape information cannot be recovered. This formulation also permits us to develop an improved estimator of illuminant direction.

2 The Imaging of Surfaces

Let $z = z(x,y)$ be a surface, and let us assume that:

(1) the surface is Lambertian,

(2) the surface is illuminated by (possibly several) distant point sources,

(3) the surface is not self-shadowing.

We will also take $z < 0$ within the region of interest, and assume orthographic projection onto the x, y plane.

We will let $\mathbf{L} = (x_L, y_L, z_L) = (\cos\tau \sin\sigma, \sin\tau \sin\sigma, \cos\sigma$ 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).

[1]This research was made possible by National Science Foundation, Grant No. DCR-85-19283. I wish to thank Berthold Horn, Graham Smith, and Yvan Lelerc for their comments and insights.

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

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

where

$$p = \frac{\partial}{\partial x} z(x,y) \quad (3)$$

$$q = \frac{\partial}{\partial y} z(x,y) \quad (4)$$

2.1 Linear Approximation

If we then take the Taylor series expansion of $I(x,y)$ about $p,q = 0$ up through the quadratic terms, we obtain

$$I(x,y) \approx \cos\sigma + p\cos\tau\sin\sigma + q\sin\tau\sin\sigma - \frac{\cos\sigma}{2}(p^2+q^2) \quad (5)$$

This expression gives an excellent approximation when $|p|, |q| < 1$.

Under the condition $|p|, |q| < 1$ the linear terms of Eqn. (5) will dominate the power spectrum except when the average illuminant is within roughly $\pm 30°$ of the viewers' position. When either $p,q \ll 1$ or the illumination direction is roughly perpendicular to the line of sight the quadratic terms will be negligible.

We will assume that such is the case in the following analysis[2]. Thus we will approximate the intensity function by:

$$I(x,y) = \cos\sigma + p\cos\tau\sin\sigma + q\sin\tau\sin\sigma \quad (6)$$

Note that this is exactly the lunar reflectance function [4].

2.2 Spectral Properties

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

$$F_z(f,\theta) = m_z(f,\theta)e^{i\phi_z(f,\theta)} \quad (7)$$

where $m_z(f,\theta)$ is the magnitude at position (f,θ) on the Fourier plane, and ϕ_z is the phase.

Now since p and q are partial derivatives of $z(x,y)$, their transforms F_p and F_q are related to F_z in an elementary fashion. We can write

$$F_p(f,\theta) = 2\pi\cos\theta f m_z(f,\theta)e^{i(\phi_z(f,\theta)+\pi/2)} \quad (8)$$

$$F_q(f,\theta) = 2\pi\sin\theta f m_z(f,\theta)e^{i(\phi_z(f,\theta)+\pi/2)} \quad (9)$$

In this case, the Fourier transform of the image I is (ignoring the DC term):

$$F_I(f,\theta) = \frac{2\pi\sin\sigma f m_z(f,\theta)e^{i(\phi_z(f,\theta)+\pi/2)}}{[\cos\theta\cos\tau + \sin\theta\sin\tau]} \quad (10)$$

2.3 Recovery of Shape

This spectrum depends, as expected, upon the illuminant direction and the surface $z(x,y)$. What is remarkable is that given the illuminant direction we can recover the surfaces' Fourier transform directly, except for an overall scale factor and the low-frequency terms which are lost in going from Eqn.(7) to Eqns.(8) and (9). That is, if we let

$$F_I(f,\theta) = m_I(f,\theta)e^{i\phi_I(f,\theta)} \quad (11)$$

then the Fourier transform of the z surface is simply

$$F_z(f,\theta) = \frac{m_I(f,\theta)e^{i(\phi_I(f,\theta)-\pi/2)}}{2\pi\sin\sigma f[\cos\theta\cos\tau + \sin\theta\sin\tau]}. \quad (12)$$

The ability to directly recover surface shape from the Fourier components of the image suggests a theory of human shape perception. It is known that the visual systems' 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, it is clear that one could produce a coarse estimate of shape by (1) phase-shifting the cells' responses by $\pi/2$, (2) scale the cells activity by $1/f$, where f is the spatial frequency that the cell is tuned for, and (3) biasing the cells' activity to remove coarse variation in the distribution of activity versus orientation, i.e., to remove the effects of the illumination direction.

[2]Note that when these conditions are *not* true, then the image gradient direction is approximately parallel to the image-plane component of the surface orientation (tilt), and the magnitude of the image intensity gradient is proportional to the remaining component of orientation (slant). Thus when our assumptions are seriously violated, the recovery of surface orientation may be accomplished by local analysis of the image gradient field.

2.4 Estimating the Illuminant Direction

Pentland [7] 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:

$$(x_L^*, y_L^*) = (\beta^T \beta)^{-1} \beta^T (dI_1, dI_2, ..., dI_n) \qquad (13)$$

where (x_L^*, y_L^*) are the unnormalized x and y components of the illuminant direction, β is a 2 x 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^*) we may then find the complete illuminant direction, which is simply:

$$x_L = x_L^*/k \qquad y_L = y_L^*/k \qquad z_L = \sqrt{1 - x_L^2 - y_L^2} \qquad (14)$$

where

$$k = \sqrt{E(dI^2) - E(dI)^2} \qquad (15)$$

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 [2,3,7], however the assumption of uniformly distributed surface orientations is disagreeably strong. We can improve this method substantially by observing that in Eqn.(10) the illuminant produces a similar effect *in each frequency band*. Thus if we make the much weaker assumption that the power in a particular spatial frequency band is uniformly distributed over orientation[3] then we 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, Eqn.(13) becomes

$$(x_L^*, y_L^*) = (\beta^T \beta)^{-1} \beta^T (m_1, m_2, ..., m_n). \qquad (16)$$

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

3 Surface Recovery Results

We have applied Eqn.(12) to both synthetic images of complex surfaces, such as is shown in Figure 1(a) (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 2(a) and 3(a). 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 1(b) shows the distribution of intensity values obtained when the surface of Figure 1(a) is illuminated from $\mathbf{L} = (1,1,1)/\sqrt{3}$. Figure 1(c) shows the distribution of errors 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 (i.e., $\max(p,q) \approx 5.0$).

Figure 1(d) shows the surface recovered by use of Eqn.(12). Because the low-frequency terms and the overall amplitude cannot be recovered, it was necessary to scale the recovered surface to have the same standard deviation as the original surface before we could compare the two surfaces. Figure 1(e) 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 Taylor expansion approximation breaking down for steeply-sloped regions of the surface, i.e., those with $|p|, |q| >> 1$.

Figure 2(a) shows a high-altitude image of a mountainous region outside of Phoenix, Arizona. This area has been the subject of intensive study, so that we are able to compare our shape-from-shading algorithm to, for instance, results obtained using stereopsis. In particular, the Defense Mapping Agency has created a depth map of this region using their interactive stereo system. The stereo depth map they recovered is shown in Figure 2(b). Such maps are hard to interpret, so we created an synthetic image from this stereo depth map using standard computer graphics techniques. The image created from the stereo depth map is shown in Figure 2(b). In addition, Figure 2(d) shows a perspective view of this stereo depth map.

Figure 2(e) shows the depth map recovered from the shading information in Figure 2(a), by use of Eqn. (12). As part of the recovery process, the illuminant direction was estimated from the Fourier transform of the image by use of Eqn.(16). To aid in the evaluation of this shading-derived depth map, we also created an

[3] Or, more precisely, is not distributed in a way that is correlated with the illuminant effects

intensity image (Figure 2(f)), and a perspective view (Figure 2(g)).

The accuracy of shape recovery for this image can be assessed by either comparing the original image (Figure 2(a)) with the synthetic image created from the recovered surface (Figure 2(f)), or (better) by comparing perspective views of the stereo-derived surface (Figure 2(d)) with the shading-derived surface (Figure 2(g)). These comparisons demonstrate that the recovery of shape from shading in this example is quite accurate. The only major defect of the recovered depth map is a slight bowing of the entire surface, which appears to stem from inaccuracies in the estimation of the illuminant direction.

It is interesting to compare these results to those achieved using the most sophisticated iterative algorithms. In [8], for instance, reasonably accurate recovery of surface shape was achieved for complex synthetic images, such as shown in Figure 1, after 50 iterations of the shape-from-shading interleaved with 50 iterations of the integrability algorithm. Although no quantitative error statistics were given, the graphs shown indicate about 10% average error in the recovered height surface. In similar situations (where $|p|, |q| < 1$, oblique illumination) we have found that our algorithm typically has significantly less average error — and, of course, no iteration or integration is required.

A final example of shape recovery is shown in Figure 3. Figure 3(a) shows a complex image widely used in image compression research. Figure 3(b) shows the shape estimated for this image. Although the surface appears jumbled, because of large changes in surface albedo, it is in fact a correct interpretation of Figure 3(a) — i.e., illuminating the surface shown in Figure 3(b) will result in the image shown in Figure 3(a). Moreover, the outline and correct general shape of the woman is present in this recovered surface. Figure 3(c) shows a close-up of the recovered surface in the neighborhood of the face. The eyes, cheek, lips, the nose arch and nostrils can all be clearly seen in the recovered surface. The accuracy of recovery is illustrated more strikingly in Figure 3(d), which shows a shaded, oblique view of the recovered surface. The somewhat "smeared" appearance of the facial features is due to the highly foreshortened view in the original image.

4 Summary

By approximating the Lambertian image formation process using the linear terms of the Taylor series, we can achieve a simple closed-form expression that relates surface shape to image intensity. This new result may be used in two ways.

First, the image-surface equation gives us the basis for a new method of estimating the illuminant direction that makes weaker assumptions about the viewed scene.

Second, the image-surface equation is invertible in closed form, except for overall scale and low spatial frequency terms, given the illuminant direction. Thus we may recover high-frequency shape information directly from image shading without the necessity of integration or iterative propagation of shape information.

Experimental results indicate that the recovery process is both stable and (at least for the images so far examined) quite accurate. Further, because the technique is both simple and uses only biologically-available information, it may serve as a model for human perception.

One special aspect of this approach is that it makes no assumption of surface smoothness. In place of the smoothness assumption we substitute a somewhat simpler reflectance function, and the assumption general viewing position: Changes and discontinuities are inserted only where there is direct evidence for them in the image intensities. Because smoothness is not assumed this approach can be successfully applied to complex natural surfaces such as hair, mountains, or bushes. By the same token, however, on smooth-man made surfaces this approach may not perform as well as techniques that employ smoothness constraints.

5 REFERENCES

[1] Pentland, A. P. (1984) Local Analysis Of The Image, *IEEE Transactions on Pattern Analysis and Machine Recognition,* pp. 170-187, March 1984.

[2] Ferrie, F. P., and Levine, M. D., (1985) Piecing together the 3D shape of moving objects, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* San Francisco, CA. June 19-23., pp. 574-584.

[3] Smith, G. B., (1983) Shape from Shading: An Assessment, *SRI AI Center Tech. Note 287,* SRI International, Menlo Park, CA.

[4] Horn, B.K.P. *Personal Communication.*

[5] Brooks, M. J., and Horn, B.K.P., (1985) Shape and Source from Shading, *Proc. Int. Joint Conf. on Artificial Intelligence,* Los Angeles, pp. 932-936.

[6] Smith, G. B., *Personal Communication.*

[7] Pentland, A. P. (1982) Finding the illuminant direction *Optical Society of America,,* Vol. 72, No. 4, 448-455.

[8] Frankot, R.T., and Chellappa, R., (1987) A Method For Enforcing Integrability In Shape From Shading Algorithms, *Proc. First Intl. Conf. on Computer Vision,,* pp. 118-127, June 8-11, London, England

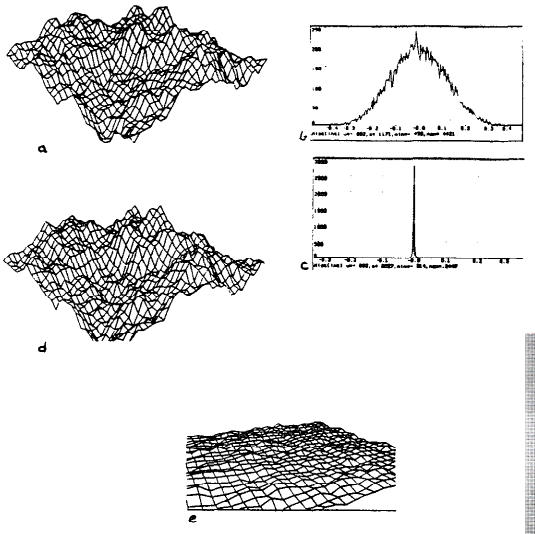

Figure 1: (a) A fractal Brownian surface, (b) the distribution of intensities within the image of the surface in (a), (c) the distribution of differences between the image and our linear-term-only Taylor series approximation, (d) the surface recovered from shading (compare to (a)), and (e) the errors in the recovery process.

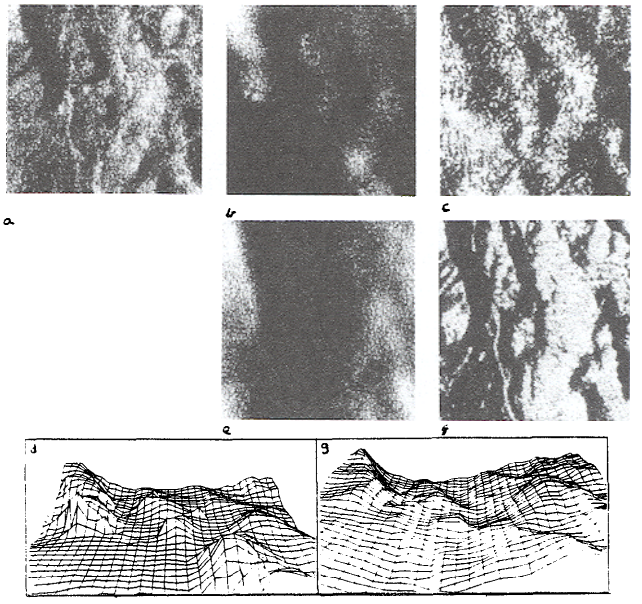

Figure 2: (a) An image of a mountainous region outside of Phoenix, Arizona, (b) a depth map of this region obtained from a stereo pair by the Defense Mapping Agency, (c) an image created from this stereo depth map, (d) a perspective view of the stereo depth map, (e) the depth map recovered from shading information alone, by use of Eqn. (12), (f) an image created from this shading depth map, and (g) a perspective view of the depth map derived from image shading.

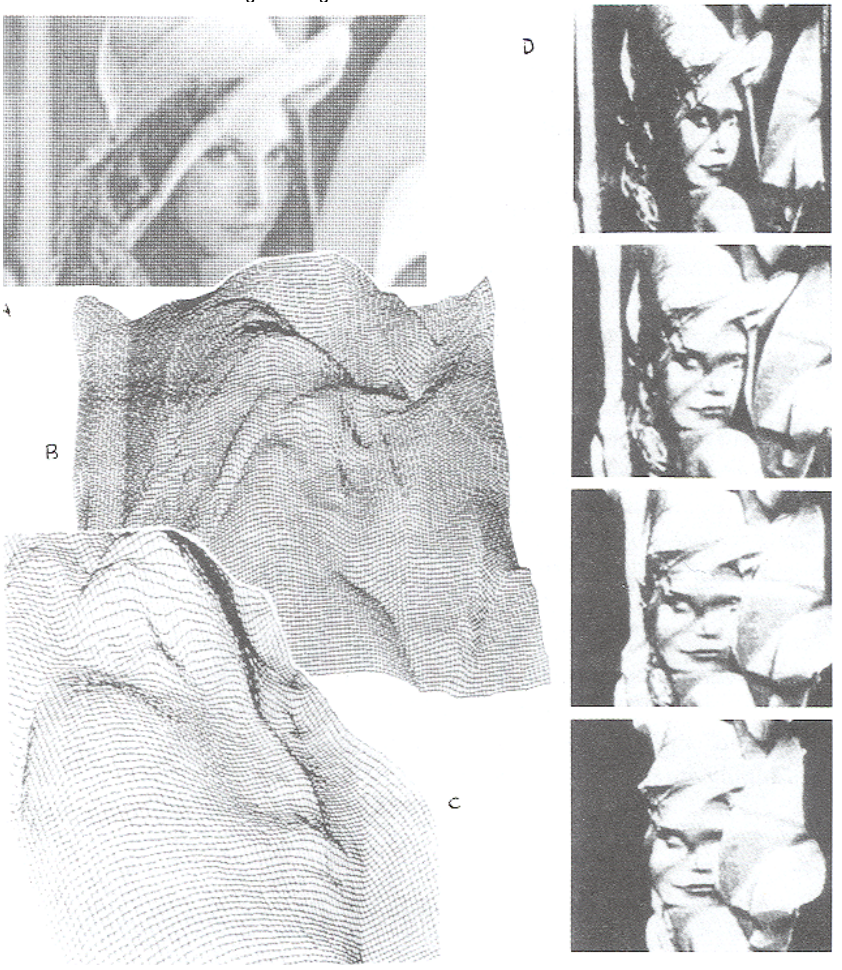

Figure 3: (a) An image of a woman used in image compression research, (b) a perspective view of the depth map recovered from shading information alone, by use of Eqn. (12), (c) a close-up of the recovered surface in the neighborhood of the womans' face; note the presence of eyes, cheek, lips, nostrils, and nose arch, (d) a shaded, oblique view of the recovered surface.