Shape Information From Shading: A Theory About Human Perception

Alex Pentland

Vision Sciences Group, E15-410
The Media Lab, M.I.T.
20 Ames St., Cambridge MA 02138

ABSTRACT[1]

I show that people assume a simple, linear reflectance function when interpreting shading information. Using this reflectance function I derive a closed-form solution to the problem extracting shape information from image shading. The solution does not employ a assumptions about surface smoothness and so is directly applicable to complex natural surfaces such as hair or cloth. A simple biological mechanism is proposed to implement this recovery of shape. It is shown that this simple mechanism can also extract significant shape information from line drawings.

Shading appears to me to be of supreme importance in perspective, because, without it opaque and solid bodies will be ill-defined.... Leonardo Da Vinci, Notebooks

1 Introduction

People's ability to extract shape from shading is well known [1,2,3], however relatively little research has gone into determining what simplifying assumptions people might be employing. There is good reason to expect that such assumptions are critical to this task, as without them the problem is vastly underconstrained [4,5,6]. The physics of image shading indicates that there are only three types of simplifying assumption that might be useful [4]. 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).

Assumptions about surface shape have received the most attention in recent research. There have been two types of simplifying assumption that have been found useful: (1) smoothness assumptions, first employed by Horn [4], and (2) assumptions about local surface curvature, first employed by Pentland [6]. The resulting shape-from-shading techniques have estimated surface orientation, so that integration is required to recover depth. Techniques employing smoothness assumptions have the disadvantage of requiring dozens of iterations to converge to an acceptable answer, however these techniques are generally thought to be more accurate (but see [7,8]). Despite considerable effort neither type of surface assumption has resulted in the sort of reliable, locally accurate recovery of shape typical of human vision.

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

With respect to the distribution of illumination, most research has accepted the hypothesis that people assume a single, distant illuminant within "fairly large" image regions. I have been able to show, for instance, that a model based on this assumption and an assumption of surface isotropy is able to accurately predict human perceptions of illumination direction [9]. This type of estimator has proven robust and accurate in a variety of real-world situations.

Research concerning the final type of simplifying assumption — assumptions about the reflectance function — has mostly focused on the role of highlights and specularities [2,10,11,12]. Highlights and specularities, however, account for a very small part of the total image area. Virtually no attention has been given to the diffusely reflecting areas which constitute most of the image.

In this paper I will first explore human assumptions about surface reflectance functions, and then use our conclusions to develop a new type of shape-from-shading theory, one based on an assumption about the reflectance function rather than about the surface shape. I will then describe a simple biological mechanism that implements this shape-from-shading process. Finally, I will show that this mechanism can extract a significant amount of shape information from line drawings.

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:

(1) the surface is Lambertian,

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

(3) the surface is not self-shadowing.

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

I 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).

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

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

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

$$p = \frac{\partial}{\partial x} z(x,y) \tag{2}$$

$$q = \frac{\partial}{\partial y} z(x,y) \tag{3}$$

2.1 Transform Properties

Equation (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

$$I(x,y) \approx \cos\sigma + p\cos\tau\sin\sigma + q\sin\tau\sin\sigma - \frac{\cos\sigma}{2}(p^2+q^2) \tag{4}$$

This expression yields an excellent approximation of the image intensities under all commonly-occurring situations.

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

$$F_z(f,\theta) = m_z(f,\theta)e^{i\phi_z(f,\theta)} \tag{5}$$

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, e.g.,

$$F_p(f,\theta) = 2\pi\cos\theta f m_z(f,\theta)e^{i(\phi_z(f,\theta)+\pi/2)} \tag{6}$$

$$F_q(f,\theta) = 2\pi\sin\theta f m_z(f,\theta)e^{i(\phi_z(f,\theta)+\pi/2)} \tag{7}$$

Under the condition $|p|, |q| < 1$ the linear terms of Equation (4) will dominate the image intensity function 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 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):

$$F_I(f,\theta) = 2\pi\sin\sigma f m_z(f,\theta)e^{i(\phi_z(f,\theta)+\pi/2)}[\cos\theta\cos\tau + \sin\theta\sin\tau] \tag{8}$$

That is, *the image intensity surface $I(x,y)$ is a linear function of the height surface $z(x,y)$*[2]

[2]This linear relation has a long history in image understanding work. It is a good approximation of the moon's reflectance function, and as such was the basis for early work on obtaining shape information from shading. More recently this equation was analyzed in 1986 by Kube and Pentland [18] in order to obtain a constructive proof of Pentland's 1983 work concerning the imaging of fractal surfaces [19], and still later by Simchony and Chellappa 1987 [20] while investigating possible applications of direct solution methods to vision problems involving Poisson equations.

When the quadratic terms of Equation (4) 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 $\mathbf{L} = (0,0,1)$ will be (very 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.

2.2 Two Experiments

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 being an accurate description of the Lambertian reflectance function under a wide variety of common situations.

The hypothesis to be tested, then, is whether or not people assume a linear reflectance function when interpreting image shading. This hypothesis can be tested in two ways: One, by asking people to describe surface shape directly, and two, by asking people to make in front/in back judgments about pairs of surface points.

Experiment 1: The "egg crate" surface shown in Figure 1(a) ($z(x,y) = \cos(kx)\cos(ky)$) was illuminated from a variety of directions to produce the images shown in Figure 1(b). The actual images were full grey-scale images, halftone was used only for reproduction in this paper. For some illuminant positions ($\mathbf{L} = (1,1,1)$, $\mathbf{L} = (0.1,0.1,1.0)$) there is an approximately linear relationship between image intensity and surface shape. For other illuminant directions ($\mathbf{L} = (0.01,0.01,1.0)$, $\mathbf{L} = (0,0,1.0)$) the quadratic terms dominate, so that "frequency doubling" occurs.

Subjects ($N = 10$) were presented (free viewing on a CRT screen, random ordering) with the images in Figure 1(b) and asked to trace the height profile of the imaged 3-D surface along the black line drawn through each of these images. Tracing was done using paper and pencil, with the paper held against the screen just below the black line. Subjects were told about the direction of illumination, and that the image was a view looking straight down on a synthetically-generated surface.

The results were that for all subjects the fundamental spatial frequency of the traced surface profile matched the fundamental spatial frequency of the image intensity surface to within 5% error. That is, there was a (first order) linear relationship between the image intensity surface and the perceived surface shape.

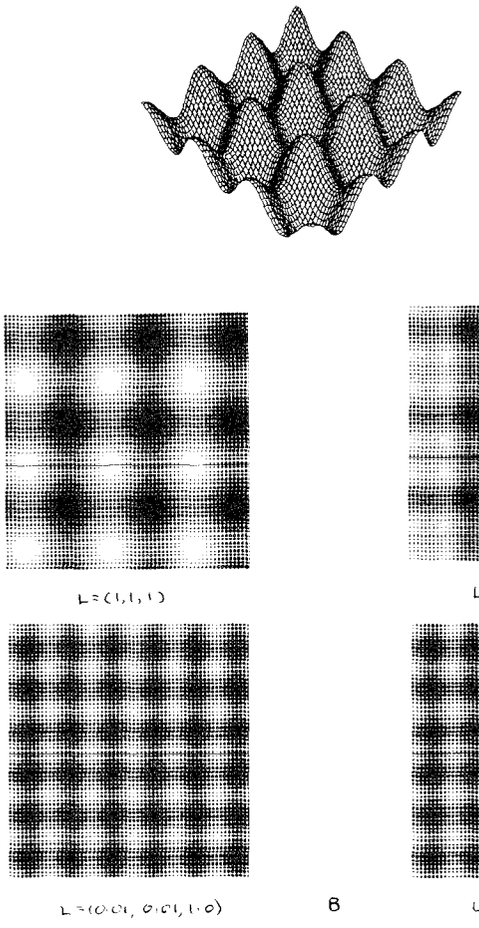

Figure 1: (a) An "egg crate" surface, (b) images of this surface produced using various illumination directions: $\mathbf{L} = (1,1,1)$, $\mathbf{L} = (0.1,0.1,1.0)$, ($\mathbf{L} = (0.01,0.01,1.0)$, and $\mathbf{L} = (0,0,1.0)$. The actual images were full grey-scale images, half-tone is used only for reproduction in this paper.

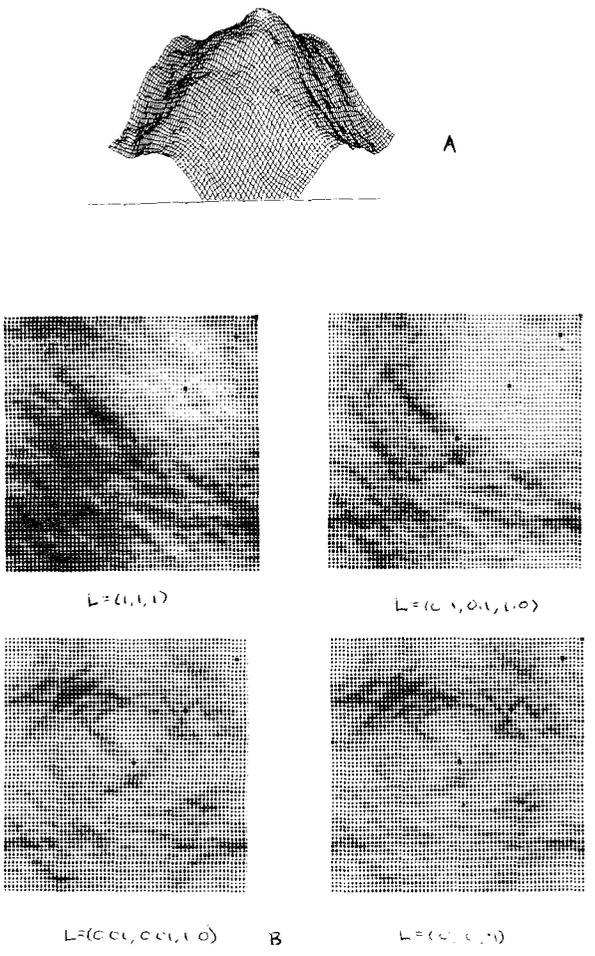

Figure 2: (a) A fractal Brownian surface, (b) images of this surface produced using various illumination directions: $\mathbf{L} = (1,1,1)$, $\mathbf{L} = (0.1,0.1,1.0)$, ($\mathbf{L} = (0.01,0.01,1.0)$, and $\mathbf{L} = (0,0,1.0)$. The actual images were full grey-scale images, half-tone was used only for reproduction in this paper.

Experiment 2: The complex Fractal surface shown in Figure 2(a) was illuminated from a variety of positions, as shown in Figure 2(b). The actual images were full grey-scale images, half-tone was used only for reproduction in this paper. As before, for illuminant positions $\mathbf{L} = (1,1,1)$ and $\mathbf{L} = (0.1,0.1,1.0)$ there is an approximately linear relationship between the image intensity and the surface shape. For illuminant directions $\mathbf{L} = (0.01,0.01,1.0)$ and $\mathbf{L} = (0,0,1.0)$ the quadratic terms dominate so that "frequency doubling" occurs. Three dark points were superimposed on each image, as shown.

Subjects ($N = 10$) were presented (free viewing on a CRT screen, random ordering) with these images and asked to decide which of the three points on the surface was closest, and which was farthest. Note that in each case the actual surface was the same, only the illumination direction varied. Subjects were told about the direction of illumination, and that the image was a view looking straight down on a synthetic, terrain-like surface.

The results were that the depth ordering of the dots was consistent with that predicted by the linear reflectance function hypothesis (rank correlation, $p < 0.01$) when the linear terms dominated, and not consistent for those illumination conditions in which the quadratic terms dominate and fre-

quency doubling occurs. In the frequency doubling cases, the results were consistent (rank correlation, $p < 0.01$) with the ordering obtained assuming (incorrectly) a linear reflectance function.

The results of these two experiments are clear: It appears that people assume a linear relationship between the shape of the surface and that of the image, even when that is objectively not the case. That is, they seem to assume a linear reflectance function for the imaged surface, such as the following linear approximation to the Lambertian reflectance function:

$$I(x,y) \approx \cos\sigma + p\cos\tau\sin\sigma + q\sin\tau\sin\sigma \qquad (9)$$

so that the relationship between the Fourier transform of the image and that of the surface is given by Equation (8).

2.3 Recovery of Shape

Examining Equation(8) 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

$$F_I(f,\theta) = m_I(f,\theta)e^{i\phi_I(f,\theta)} \quad , \qquad (10)$$

then the Fourier transform of the z surface is simply

$$F_z(f,\theta) = \frac{m_I(f,\theta)e^{i(\phi_I(f,\theta)-\pi/2)}}{2\pi\sin\sigma f[\cos\theta\cos\tau + \sin\theta\sin\tau]}. \qquad (11)$$

Note that Fourier transforms are used here simply for convenience; any orthogonal linear basis set could have been used.

Thus the assumption of a linear reflectance function results in a computationally attractive scheme for recovering surface shape, one that does not — unlike previous schemes — require iteration or integration. Further, no strong assumptions about the surface are required, so that the the scheme is applicable to complex surfaces such as hair or cloth.

Nonetheless, it is important to remember that this computational theory does assume a constant average surface reflectivity, and a constant, known illumination direction. Thus a full theory of shape-from-shading requires segmenting the image into regions of homogeneous reflectivity, a topic that is beyond the scope of this paper. The estimation of illuminant direction is discussed in the Appendix; the scheme used here is an improvement on the original scheme suggested by Pentland [7].

The next section develops a biological mechanism for computing shape from shading. The following sections show results obtained by this computational theory to a variety of real-world images.

3 A Biological Mechanism

The ability to recover surface shape by use of Equation (11) suggests a biological 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, it is clear that one could produce an estimate of shape by summing the output of these cells in a selective manner.

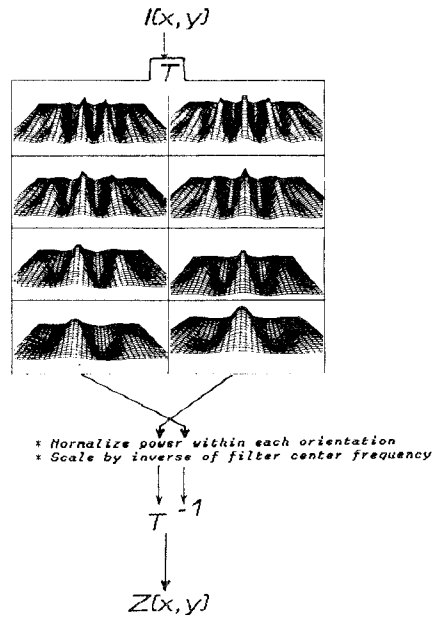

Figure 3: A Shape-from-shading mechanism: A transformation **T** produces localized measurements of sine and cosine phase frequency content, and then the inverse transformation is applied, switching sine and cosine phase amplitudes and scaling the filter amplitude in proportion to the central frequency. The output of this process is the recovered surface shape.

Figure 3 illustrates this mechanism. I assume that some transformation **T** of the image produces a set of filters that are localized in 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 [13,14,15].

Although the details of *exactly* what these filters look like is unimportant for recovering shape estimates I will for purposes of discursive clarity assume that these filters are sine and cosine phase Gabor filters. 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, so that

the equations developed above are directly applicable.

In order to recover surface shape from this filter set, the transformations indicated in Equation (11) 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 $\mathbf{T}^{-1}$ in Figure 3; it is the inverse of what is believed to happen between the retina and striate cortex[3].

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 filters 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 (the "receptive field" of 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 [16]. 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.

4 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 examples of using this approach to recover shape from shading.

Figure 4(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 I can compare this shape-from-shading algorithm to, for instance, results obtained using stereopsis. In particular, the Defense Mapping Agency has created a elevation map of this region using their interactive stereo system. Figure 4(b) shows a perspective view of this stereo-recovered surface. Note that this surface contains serious errors at several isolated points, a result typical of stereo processing.

I estimated the surface shape in Figure 4(a) by use of our shape-from-shading mechanism. As part of the recovery process, the illuminant direction was estimated from the Fourier transform of the image, as described in the Appendix. Figure 4(c) shows an image created by illuminating this recovered surface using standard computer graphics techniques. Figure 4(d) shows a perspective view of the shading-derived surface shape.

The accuracy of shape recovery for this image can be assessed by either comparing the original image (Figure 4(a)) with the image created from the recovered surface (Figure 4(c)), or by comparing the perspective views of the stereo-derived surface (Figure 4(b)) and the shading-derived surface (Figure 4(d)). These comparisons demonstrate that the recovery of shape from shading in this example is quite accurate. The only 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 Figure 5. Figure 5(a) shows a bas relief sculpture from the Metropolitan Museum of Art. The surface of this sculpture has very low relief and roughly constant average reflectance except for the upper right hand corner. The surface was illuminated from all directions on a photographer's light table to produce this image. Figure 5(b) shows several shaded views of the recovered surface, from various viewing angles. The surface shape recovery is impressively accurate, except for low-frequency variations due to variations in illumination and surface reflectance, and a small region above the eye whose convexity was reversed due to inadvertent shadowing.

A third example of shape recovery is shown in Figure 6. Figure 6(a) shows a complex image widely used in image compression research. Figure 6(b) shows the shape estimated for this image. Although the surface appears jumbled, because of large changes in surface reflectance, the outline and correct general shape of the woman is present in this recovered surface. Figure 6(c) shows a perspective view 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 6(d), which shows several shaded views of the recovered surface. The somewhat "smeared" appearance of the recovered facial features is due to the highly foreshortened view in the original image.

It is interesting to compare these results to those achieved using the most sophisticated iterative algorithms. One difficulty in such a comparison this that iterative algorithms have been tested almost exclusively on simple, synthetic images. However in [17] results were presented for synthetic imagery

[3] This filters used in the reconstruction can be identical to the original filters; this is not true for Gabor filters but is true for the orthogonal filter sets that are currently attracting attention

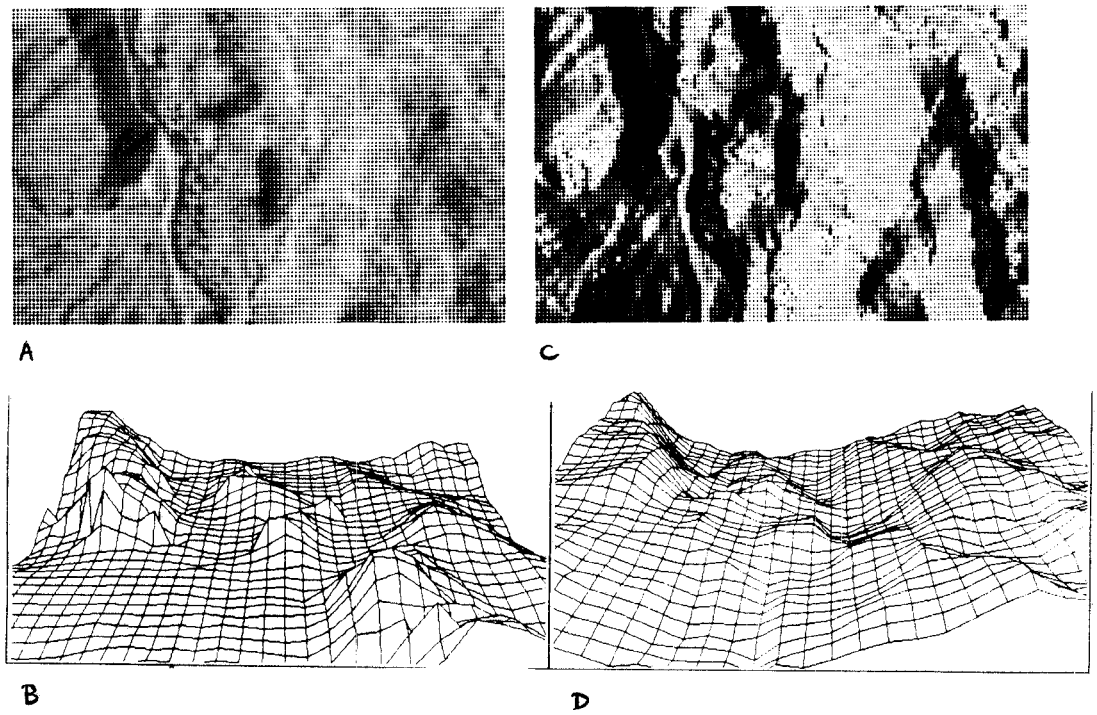

Figure 4: (a) An image of a mountainous region outside of Phoenix, Arizona, (b) a perspective view of a stereo-derived surface shape, (c) an image created from the shading-derived surface shape, and (d) a perspective view of the surface shape derived from image shading.

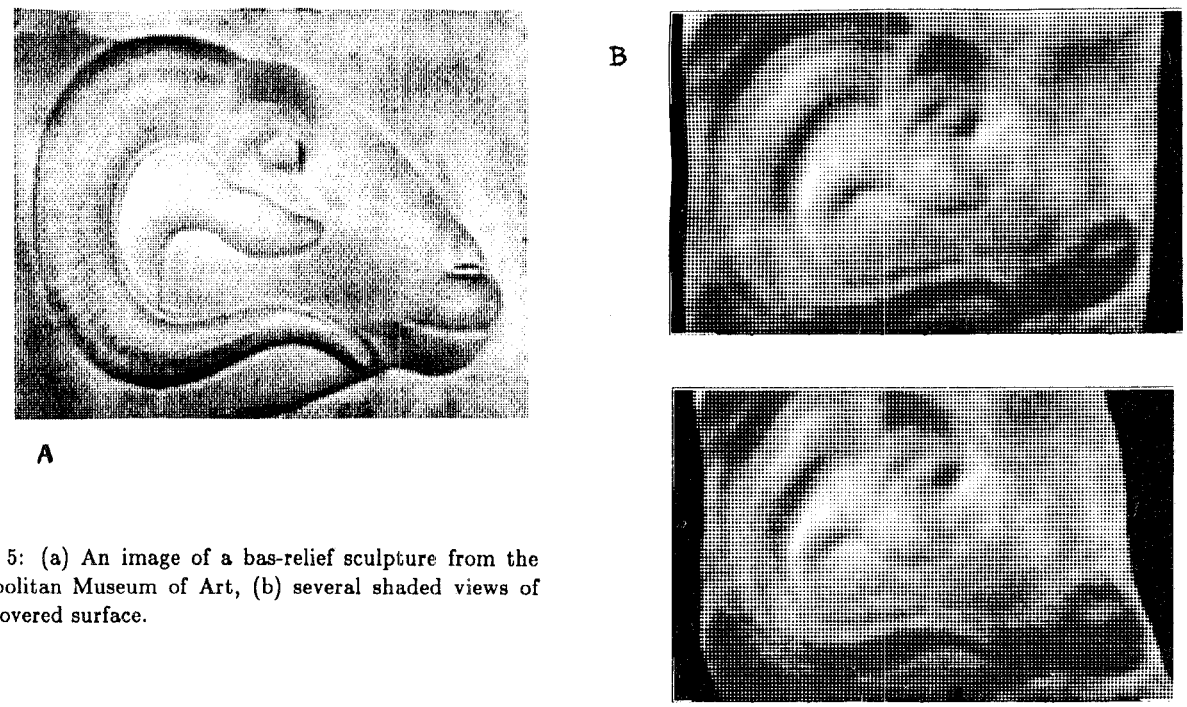

Figure 5: (a) An image of a bas-relief sculpture from the Metropolitan Museum of Art, (b) several shaded views of the recovered surface.

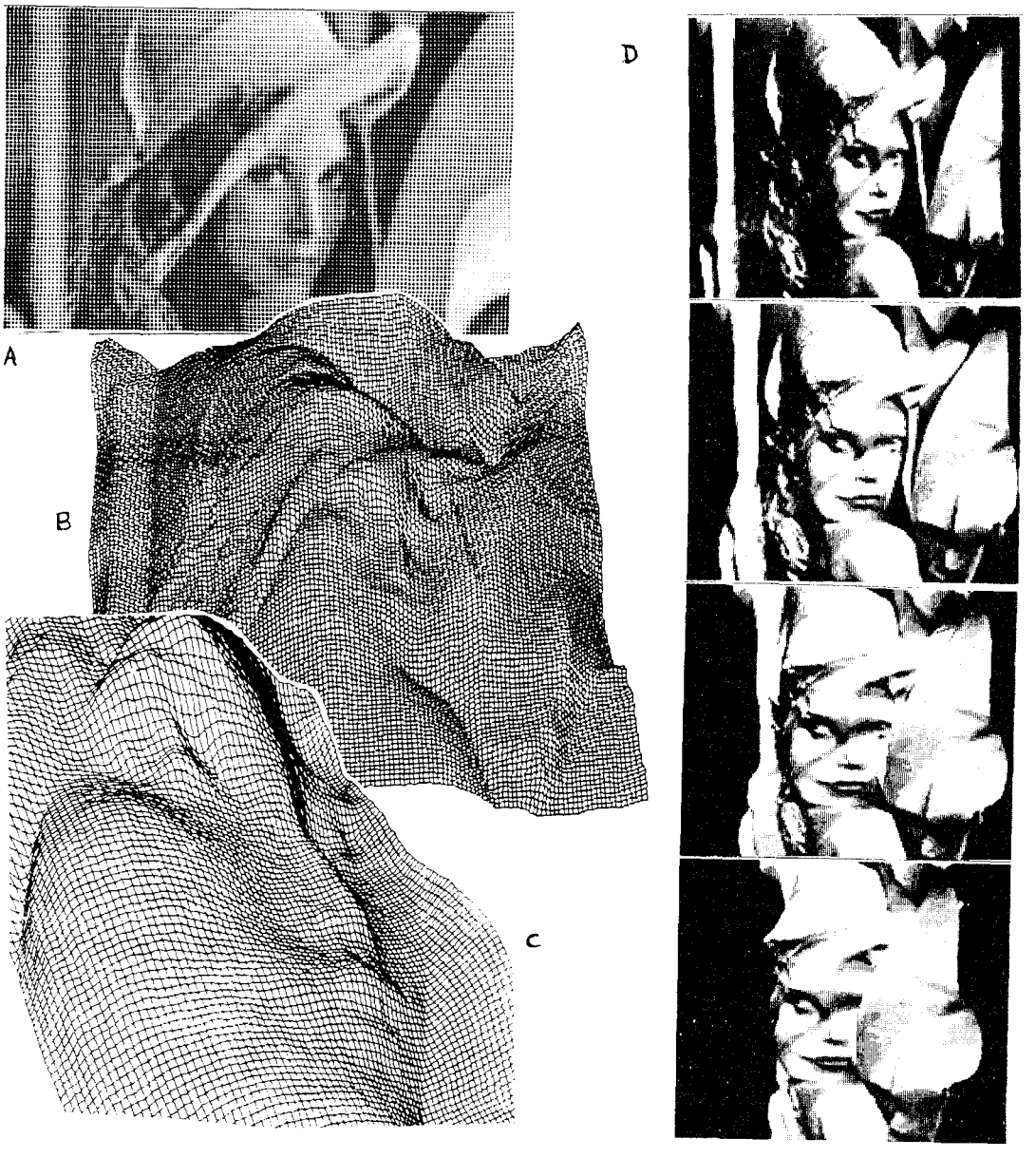

Figure 6: (a) An image widely used in image compression research, (b) a perspective view of the surface shape recovered from the image shading, (c) a perspective view of the recovered surface in the neighborhood of the womans' face; note the presence of eyes, cheek, lips, nostrils, and nose arch, (d) several shaded views of the recovered surface.

of a complex surface with uniform reflectance. Reasonably accurate recovery of surface shape was achieved after 50 iterations of their shape-from-shading technique interleaved with 50 iterations of a integrability algorithm. Although no quantitative error statistics were given, the graphs shown indicate about 10% average error in the recovered height surface. With similar synthetic imagery (where $|p|, |q| < 1$, oblique illumination) I have found that our shape-from-shading mechanism typically has roughly half that amount of error — and, of course, no iteration or integration is required.

In the examples presented here — which are real, rather than synthetic, images — the small-scale details of the surface were accurately recovered within each homogeneous area. However in each case there were also low-frequency errors due to a variety of factors, principally changes in the average surface reflectivity and variation in illuminant direction. This suggests that this type of shape-from-shading will be most useful in conjunction with some other cue to depth, such as stereo, which can reliably provide the coarse, low-frequency structure of the scene. Given such coarse shape information, it appears that an accurate, detailed estimate of surface shape can be produced by "blending" together the shading and stereo information in the frequency domain, giving the most weight to the stereo information in the low spatial frequencies and the most weight to the shading information in the high spatial frequencies.

5 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 7(a), 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 7(b). Figure 7(c) 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.

It is surprising that a shading analysis, without any high-level knowledge, can recover so much image structure from this line drawing. Part of what occurs in this image is that the artist has placed shading cues at places where large changes occur in the surface, and our technique interpolates between these points. Another important factor is that the technique acts to minimize the surface curvature in directions orthogonal to the illumination direction.

6 Summary

I have shown that people assume a linear model of the surface reflectance function when interpreting shading information. I have therefore approximated the Lambertian reflectance function using only the linear terms of the Taylor series around $p,q = 0$, and have been able to obtain a simple, closed-form expression that relates surface shape to image intensity. I am thus able to recover high-frequency shape information directly from image shading without the necessity of integration or iterative processing. Experimental results using real imagery indicate that the recovery process is both stable and (at least over small distances) quite accurate.

This shape-from-shading technique is easily stated in terms of filters that are believed to occur in biological visual systems. The resulting shape estimation mechanism is a simple modification of decomposition-reconstruction transformations previously proposed as a model for other types biological visual processing [13,14,15]. As a consequence of this similarity to other models, most of the details of this shape-estimation mechanism have already received wide acceptance.

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 much – and in some cases *most* – of the 3-D information in these drawings may be recovered by this simple shading analysis! These results indicate that conventional ideas about how shape is extracted from line drawings may require revision.

REFERENCES

[1] Pentland, A. (1982) The Perception Of Shape From Shading, *Proceedings of the OSA Annual Meeting*, Oct. 18-22, Tucson, AZ.

[2] Tod, J., and Mignolla, E., (1983) Perceptio of Surface Curvature and Direction of Illumination from Patterns of Shading, *Jour. Exper. Psych.: Human Perception and Perf.* Vol. 9, No. 4, pp. 583-595.

[3] Bulthoff, H., and Mallot, H., (1987) Interaction of Different Modules in Depth Perception, *IEEE/IAPR First Intl. Conf. on Computer Vision*, June 8-11, London, England, pp. 295-305

[4] Horn, B. K. P., (1975) Obtaining Shape from Shading Information, in *The Psychology of Computer Vision*, P. H. Winston (Ed.), McGraw-Hill

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

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

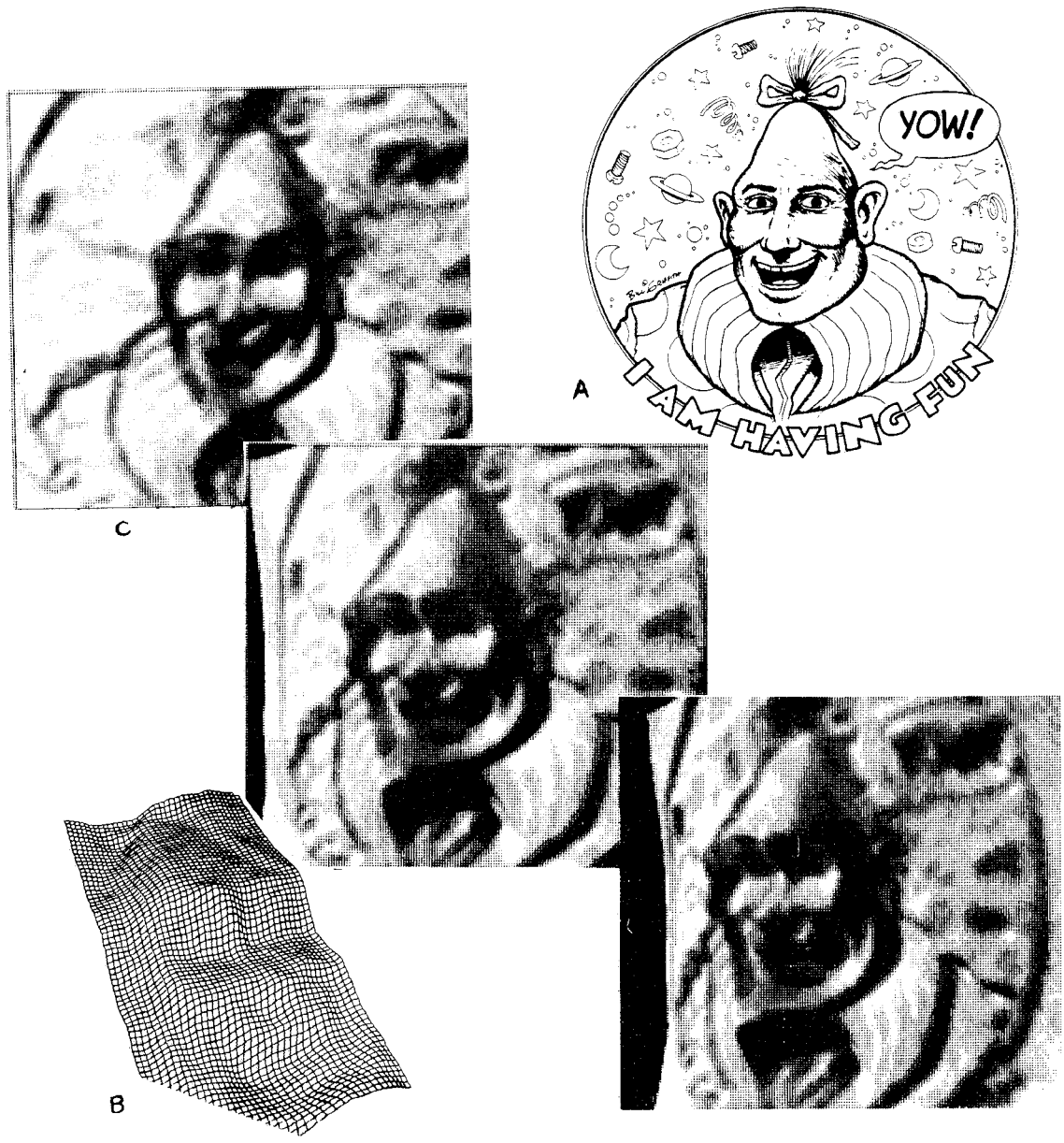

Figure 7: (a) A line drawing of a famous cartoon character, (b) a perspective view of the surface shape recovered by a shading analysis covering the face region, (c) several shaded views of the recovered surface.

Menlo Park, CA.

[8] 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.

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

[10] Lee, H.-C., (1986) Method for Computing the Scene-Illuminant Chromaticity from Specular Highlights, *Journal of the Optical Society of America A*, Vol. 3, No. 10, pp. 1694-1699.

[11] Beck, J. (1972) sl Surface Color Perception, Cornell U. Press, Ithaca, NY.

[12] Klinker, G., Schafer, S., Kanade, T., (1987) Using a Color Reflection Model to Separate Highlights from Object Color, Image Understanding Workshop, Feb. 22-23, Los Angeles, CA., pp. 614-619.

[13] Adelson, E., and Bergen, J., (1985) Spatiotemporal Energy Models for the Perception of Motion, *Journal of the Optical Society of America A* Vol. 2 No. 2, pp. 284-299.

[14] Daugman, J., (1980) Two-Dimensional Analysis of Cortical Receptive Field Profiles, *Vision Research*, Vol. 20, pp. 846-856.

[15] Watson, A., and Ahumada, A., (1985) Model of Human Visual-Motion Sensing,*Journal of the Optical Society of America A* Vol. 2 No. 2, pp. 322-342.

[16] Blaisdale, G., *Personal Communication.*

[17] 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

[18] Kube, P., and Pentland, A., (1986) On the Imaging of Fractal Surfaces, SRI AI Center Technical Note 390. Also: *IEEE Transactions on Pattern Analysis and Machine Vision, to appear Nov. 1988.*

[19] Pentland, A. (1984) Fractal-Based Description Of Natural Scenes, *IEEE Transactions on Pattern Analysis and Machine Recognition*, Vol. 6, No. 6, pp 661-674.

[20] Simchony, T. and Chellappa, R., (1987) Direct Analytical Methods for Solving Poisson Equations in Computer Vision Problems, *IEEE Computer Vision Workshop,* December, 1987, Miami Beach, FL.

Appendix: Estimating the Illuminant Direction

Pentland [9] 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:

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

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^*) I may then find the complete illuminant direction, which is simply:

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

where

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

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 [1,6,8,9], however the assumption of uniformly distributed surface orientations is disagreeably strong. This method can be substantially improved by observing that in Equation (8) the illuminant produces a similar effect *in each frequency band.* Thus using the much weaker assumption that the power in a particular spatial frequency band is uniformly distributed over orientation[4] then the same formula can be used to estimate the illuminant direction, by substituting the magnitude of the Fourier components for magnitude of the first derivatives. In particular, Equation (12) becomes

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

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

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