Diffuse Illumination as a Default Assumption for Shape-From-Shading in the Absence of Shadows

Christopher W. Tyler

Smith-Kettlewell Eye Research Institute, San Francisco, California 94115 USA

Sinusoidal luminance patterns appear dramatically saturated toward the brighter regions. The saturation is not perceptually logarithmic but exhibits a hyperbolic (Naka–Rushton) compression behavior at normal indoor luminance levels. The object interpretation of the spoke patterns is not consistent with the default assumption of any unidirectional light source but implies a diffuse illumination source (as if the object were looming out of a fog). The depth interpretation is, however, consistent with the hypothesis that the compressed brightness profile provided the neural signal for perceived shape, as an approximation to computing the diffuse Lambertian illumination function for this surface. The surface material of the images is perceived as non-Lambertian to varying degrees, ranging from a chalky matte to a lustrous metallic.

Journal of Imaging Science and Technology 42: 319–325 (1998)

Introduction

It is common to assume that the perception of the shape of an object from its shading image follows a few simple principles based on default assumptions about the light source and surface properties. For example, much of the computer vision literature makes the assumption of a spatially limited (or approximately point) source of light and surfaces of Lambertian (or uniform matte) reflectance properties. Such assumptions are commonly supposed to provide reasonable approximations to the typical interpretations of the human perceptual system (at least in the absence of explicit highlight features). In fact, however, the present analysis will show that there is wide variation in the interpreted surface quality depending on minor variations in the luminance profile of the shading image. Human observers do not seem to make a default assumption about reflectance properties, but to impute them for the particular shading image. Moreover, their interpretation of simple shading images is not consistent with the point-source assumption. The theoretical expectations for a variety of illuminant assumptions is examined in an attempt to determine what default assumption is made by human observers.

Prior work in diffuse illumination includes extensive analysis of the properties of diffuse illumination by Langer and Zucker[1] (considered in detail below) and a study of fogging of non-Lambertian objects by Barun.[2] Although these studies consider the inverse problem of estimating the shape of the surface from the resultant luminance information, they do not address the specific ambiguities and distortions that are the topic of the present work. It is well known that the surface depth corresponding to a particular (submaximal) luminance value is indeterminate for point-source illumination, because its luminance is controlled by its angle to the surface normal. For the one-dimensional surface, this ambiguity reduces to two possible values. It is less clear from the cited studies that the same ambiguity pertains to the diffuse illumination case, where surfaces become darker as they lie deeper in "holes." This kind of issue and the processes that the human brain may use to decode the surface shape are the topic of the present analysis.

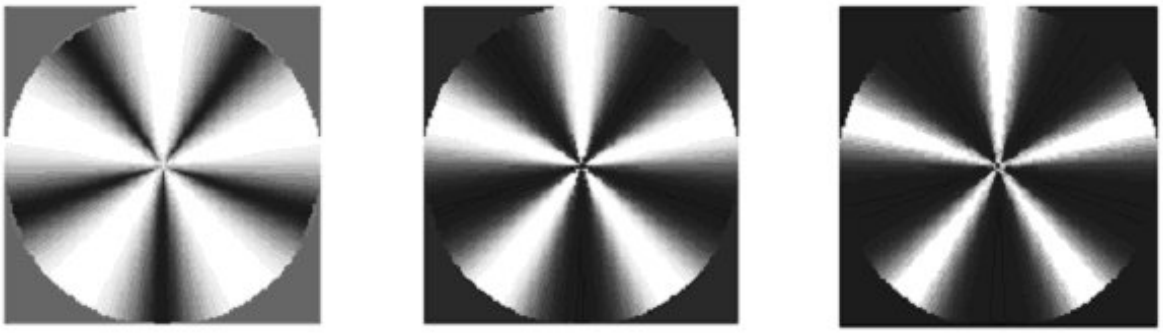
The focus will be on shading images based on sinusoidal and related luminance functions. As an initial demonstration of the shapes perceived from sinusoidal shading images, Fig. 1 depicts three spoke patterns in which there is repetitive modulation as a function of radial angle. The first pattern has a linear sinusoidal profile, the second is predistorted so as to have an approximately sinusoidal appearance to most observers, and the third is further distorted so as to appear as an accelerating function with wider dark bars than light bars. Note that, in this radial format, there is a strong tendency to perceive these luminance profiles as deriving from three-dimensional surfaces.

What are the properties of the perceived surfaces? Although the generator function is one-dimensional, we are able to estimate simultaneously the surface shape, the reflectance properties, and something about the illuminant distribution. We thus parse the one-dimensional luminance function at a particular radius in the image into three distinct functions. Such parsing can occur only if the visual system makes default assumptions about two of the functions. The question to be addressed is what default assumptions are made?

Because the patterns are radially symmetric, the illuminant distribution must itself be symmetric (or the

Original manuscript received November 10, 1997

† Email: cwt@skivs.ski.org; Web: www.ski.org/cwt; Fax:415-561-1610



(a) (b) (c)

Figure 1. Depictions of sinusoidal spoke patterns with various levels of brightness distortion; (a) linear sinusoid; (b) perceptually sinusoidal compensation, accelerating hyperbolic distortion to provide sinusoidal appearance; (c) overcompensation for perceptual inversion, extreme hyperbolic distortion to appear as an accelerating distortion. Best approximation to intended appearance will be obtained if viewed from a distance so that pixellation is not visible.

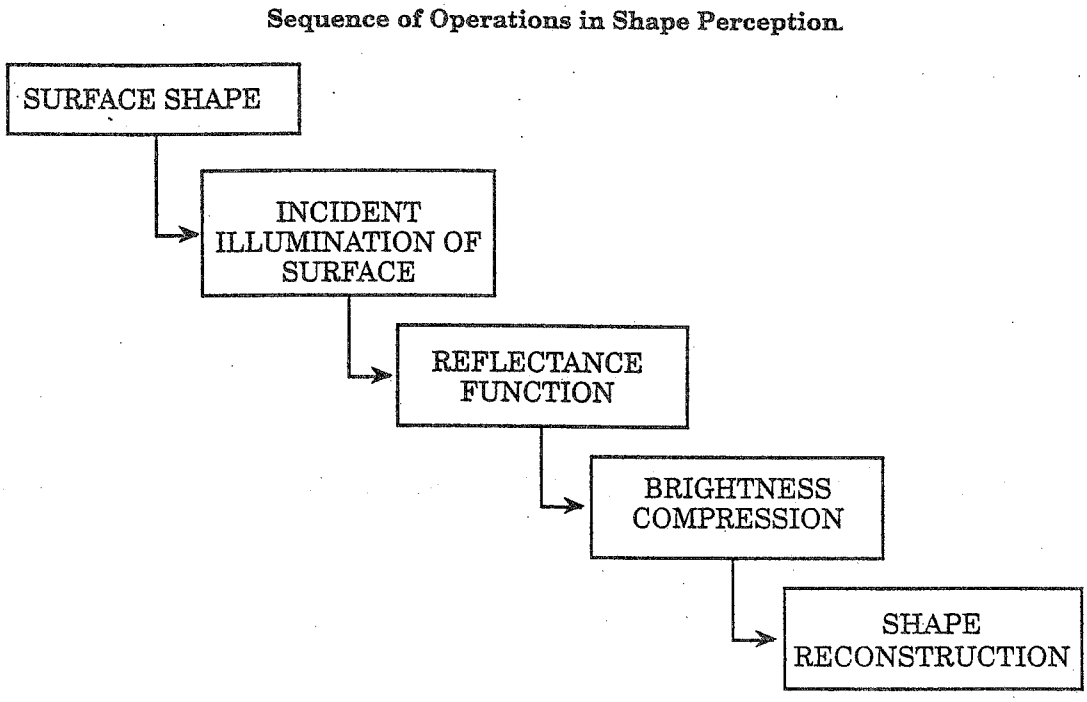


Figure 2. The sequence of operations involved in the perception of the shape of a viewed object from the luminance shading information. Does the visual system reconstruct the full sequence or use the simplifying assumption that the output approximates the input?

shading on different spokes would vary with the orientation of the spokes relative to the direction of the illuminant). Thus the only possible variation of illuminant properties is the degree of diffusion of the illuminant from a point source (positioned above the center of the surface). To most observers, the surface appears to be of matte (or Lambertian) material in Fig. 1(a) and to become progressively more lustrous in Figs. 1(b) and 1(c). Somehow, the human visual system partitions the single function in each image into separate shape, reflectance and illumination functions. This study is an initial attempt to explore the rules by which such partitioning takes place.

Compressive Brightness Distortion. Before proceeding with the analysis of surface properties, first consider the simple compressive distortion of the brightness image. If the surface properties are ignored for the moment, the direct brightness profile of Fig. 1 does not appear to be sinusoidal: the dark bars look much narrower than the bright bars (based on the perceived transition through mid-gray). This narrowing effect is far more pronounced in high-contrast images on a linearized CRT screen than in this printed example, which has a contrast of about 95%. For several reasons, it is probable that the perceived distortion arises at the first layer of visual processing, the output of the retinal cone receptors (Macleod et al.,[3] Hamer and Tyler.[4] However, the focus here is on the distortion's perceptual characteristics, not its neural origin.

It is reasonable to be skeptical of the linearity of the reproduction of Fig. 1. A simple test of the accuracy of its linearity is to view the figure in (very) low illumination, after dark-adapting the eyes for a few minutes. In such conditions, the visual system defaults to an approximately linear range, and it can be seen that Fig. 1(a) now appears to have roughly equal widths of the bright and dark bars.

In terms of shape-from-shading issues, the question arises whether the depth interpretation mechanism of the visual system "knows" that it is being fed a distorted input. The most adaptive strategy, for either genetic specification or developmental interaction with the environment, would be for the brightness distortion to be compensated

in the depth interpretation process so that the perceived brightness distortion does not distort the depth interpretation (Fig. 2).

However, the observed depth interpretation from these patterns seems to follow closely the waveform of the perceived brightness profile; when the brightness is perceived as sinusoidal [Fig. 1(b)], the surface is perceived as a roughly sinusoidal "rosette." When the brightness pattern is perceived as having narrow dark bars [Fig. 1(a)], the surface is perceived more like a ring of cones with narrow valleys between them. Fig. 1(c) continues this trend, although a second principle of change in surface properties now appears. The question to be addressed is: what principles is the visual system using in deriving its surface interpretation from the luminance profile?

The direct relationship between perceived brightness and surface depth that is the typical perception of the patterns of Fig. 1 is surprising in relation to the luminance profiles that should be expected from geometric reflectance considerations. For example, in Fig. 1(b) the surface appears approximately sinusoidal and peaks in phase with the peaks of the luminance image. As the following illumination analysis will show, this interpretation is completely incompatible with point-source illumination in any position. This incompatibility is surprising in view of the widespread use of the point-source assumption in the field of computer vision. Development of a diffuse illumination analysis then provides an explanatory basis for the observed perceptual interpretations. An additional benefit of the diffuse illumination analysis is that it shows how the direct relationship between perceived brightness and surface depth perception is compatible with the operation of a compensation for early brightness compression in the perceived brightness function.

Properties of Diffuse Illumination. Although much of the computer vision literature has concentrated on illumination by point sources, Langer and Zucker[1a,b] have laid the groundwork for the analysis of the luminance properties arising from diffuse illumination. The basic assumption is that the illuminance of any point on a surface is the integral of the incident light at that point. This amounts to the cross-section of the generalized cone of rays reaching that point through the aperture formed by the rest of the surface. Its properties are described by the sum of a direct illumination term, a (first-order) self-illumination term of reflections from other surface points and a residual ε encompassing higher order self-illumination terms.

$$R(x) = \frac{\rho}{\pi} \int_{\nu(x)} R_{src} N(x) \cdot u d\Omega + \frac{\rho}{\pi} \int_{\eta(x)/\nu(x)} R\left(\prod(x,u)\right) N(x) \cdot u d\Omega + \varepsilon, \quad (1)$$

where x is a surface point, $N(x)$ is the surface normal, $\eta(x) = \{u: N(x) \cdot u > 0\}$ is the hemisphere of outgoing unit vectors, $\eta(x)$ is the set of directions in which the diffuse source is visible from x, $d\Omega$ is an infinitesimal solid angle, and $\Pi(x,u)$ is the self-projection to the surface from point x in direction u.

The properties of self-illumination by reflection to a point from nearby surfaces have been treated for diffuse illumination by Stewart and Langer.[5] Although some special cases deviate in detail, they show that for complex surfaces the self-illumination component tends to operate as a multiplicative copy of the direct term, so that the whole equation for $R(x)$ may be approximated by the first term multiplied by a constant close to 1.

$$R(x) \approx (1+k) \frac{\rho}{\pi} \int_{\nu(x)} R_{src} N(x) \cdot u d\Omega. \quad (2)$$

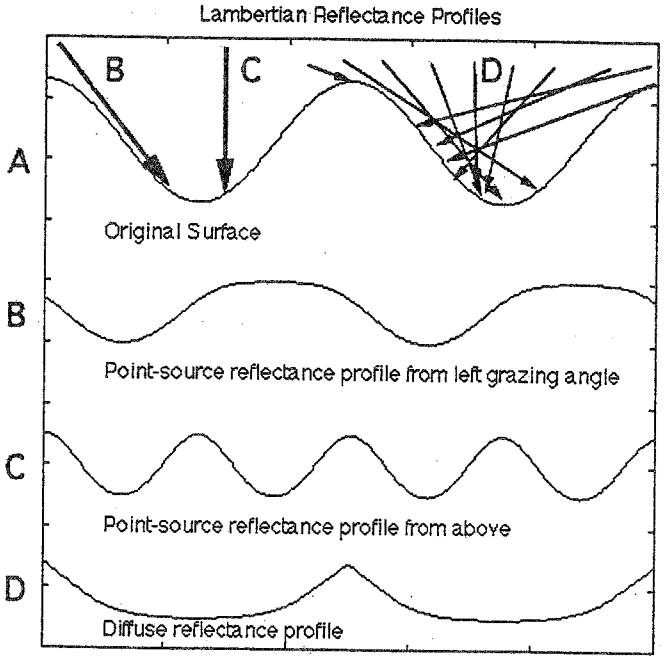


Figure 3. Lambertian reflectance profiles for a sinusoidal surface (a) under three illumination conditions; (b) point-source illumination from infinity at a grazing angle to the left-hand slopes; (c) point-source illumination from infinity directly above the surface; and (d) diffuse illumination from all directions.

Intuitively, this simplification occurs because the maximum self-illumination generally arises from surfaces of similar luminance to the point under consideration. This result is particularly clear for surfaces that are symmetric with respect to the average surface normal (such as a V-shaped valley), where the closest points across the valley are those at the same height as a chosen point. Stewart and Langer show that even extreme departures from this symmetry (such as an overhanging cliff) introduce only relatively mild distortions into the net diffuse illumination function.

Illumination Analysis. The general principles of luminance profiles based on Lambertian objects are well known, but it is instructive to consider the variety of luminance patterns that may arise from a simple object such as a sinusoidal surface under different illumination conditions, for comparison with human perceptual performance in the reconstruction of shape from shading when the light sources is unknown. For point sources at infinity, the angle of incidence is a critical variable. For the alternative assumption of diffuse illumination, the principal factor is the acceptance angle outside which the diffuse illumination is blocked from reaching a particular point on the surface. The assumptions for the following analysis are:

1. The surface has constant albedo (inherent reflectance).
2. The surface has Lambertian reflectance properties.
3. Secondary reflections from one part of the surface are negligible for point-source illumination and as described in Eq. 2 for diffuse illumination.

The Lambertian reflectance assumption is that the surface illumination is proportional to the sine of the angle of incidence at the surface and that the reflectance is uniform at all angles. Hence, the reflected light is assumed to follow the cosine rule of proportionality to the cosine of the angle of incidence relative to the surface normal.

Figure 3 shows (top) the profile of a sinusoidal surface, below which are three luminance profiles for selected illumination conditions designed to illustrate the variety of outputs. Because the surface is assumed Lambertian, the reflected luminance is proportional to the incident illumination and hence proportional to the cosine of the angle of the surface to the viewer.

The first luminance profile is derived from a point source at infinity whose angle grazes (is tangential to) the left-hand descending slopes of the sinusoidal surface. Hence, the reflected luminance is lowest at the position of the grazing slope and highest along the opposite slope, as shown by Fig. 3(b). Note that, in this position, the luminance profile has the same number of cycles as the original surface (though distorted rather than being a strict derivative).

The second luminance profile is derived from a point source at infinity directly above (normal to) the surface [Fig. 3(c)]. Because the peaks and troughs of the surface waveform are at the same angle, they have the same Lambertian reflectance and hence produce a frequency-doubled luminance profile. For 100% luminance modulation, this profile is close to sinusoidal as described in the following section. Here the point is that a quantitative shift in the angle of incidence of the point source produces a qualitative change in the resulting luminance profile of the same object.

The third luminance profile [Fig. 3(d)] is derived from the assumption of a diffuse illumination source rather than point source. The resulting luminance profile is again very different from the other two based on point sources. These examples are chosen to illustrate the complexity of the interpretation of shape from shading, because a given shape can give rise to qualitatively different shading profiles depending on the assumed source of illumination. When confronted with a luminance profile that is actually sinusoidal, does the human observer assume that it is a frequency-doubled reflection of a underlying surface of half that frequency, the diffusely illuminated profile of a nonsinusoidal surface, or a non-Lambertian surface, etc.?

Geometric Derivation. To develop the theoretical reflectance functions of Fig. 3 required two stages: computation of the angle-of-incidence functions for the selected illumination conditions according to Eq. 2 and conversion to reflectance functions through the Lambertian reflectance assumption. The sinusoidal surface profile is shown again for reference in Fig. 4, below which are plots of the angle of incidence for three different illumination conditions.

The first angle-of-incidence function [Fig. 4(b)] is derived from a point source at infinity whose angle grazes (is tangential to) the left-hand descending slopes of the sinusoidal surface. Hence, the angle of incidence is zero at the position of the grazing slope and highest along the opposite slope, as shown by Fig. 4(b). This curve will itself be sinusoidal if (and only if) the amplitude of the surface sinusoid (top curve) is such that opposite flanks are at a 90º angle to each other. Note that, in this position, the angle-of-incidence function has the same number of cycles as the original surface function (though shifted in phase in the direction of the angle of the incident light). The second angle-of-incidence function [Fig. 4(c)] is derived from a point source directly above (normal to the mean orientation of) the surface. Because the peaks and troughs of the surface waveform are at the same angle, they produce a frequency-doubled luminance profile that is asymmetric with respect to its peaks and troughs. The third angle-of-incidence function [Fig. 4(d)] is derived from the assumption of a diffuse illumination source rather than a point

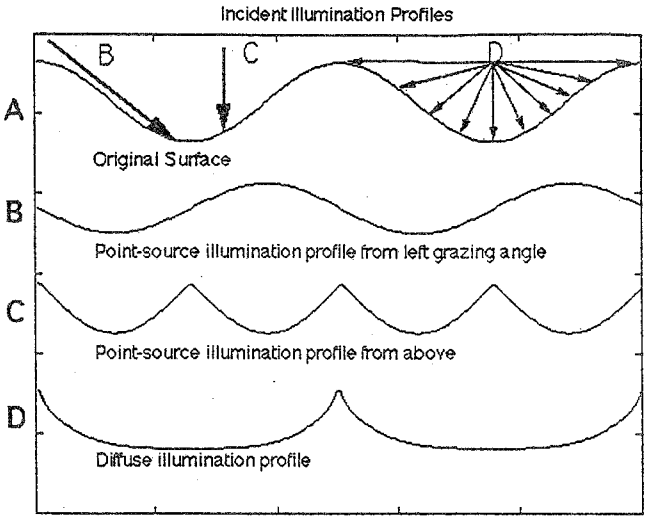


Figure 4. Net angle-of-incidence profiles for a sinusoidal surface (a) under three illumination conditions; (b) point-source illumination at a grazing angle to the left-hand slopes; (c) point-source illumination directly above the surface; and (d) diffuse illumination from all directions.

source. The light is assumed to be coming equally from all directions but to be occluded if any part of the surface lies in its path according to Eq. 2. The resulting luminance profile is again very different from the other two based on point sources.

The derivation of the diffuse illumination profile depicted in Fig. 4(d) is depicted in Fig. 5. For a particular point on the upper trace of the surface being viewed, the acceptance angle for any point on the surface is the angle between the line passing through point p that is tangent to the surface on the left [Fig. 5(b)] and the one that is tangent to the surface on the right [Fig. 5(c)]. The sum of the two angles ϕ_L and ϕ_R defines the acceptance angle for each point on the surface. Within this acceptance angle, the light from all directions has to be integrated according to the Lambertian cosine rule for each direction of the diffuse illumination relative to the orientation of the surface, as specified in Eq. 2.

The net result of the diffuse illumination analysis is shown for the sinusoidal surface by the lowest curve of Figs. 3 and 4. Note that this curve peaks at a value of π at each peak of the waveform but drops to some lower (nonzero) value depending on the absolute depth of the sinusoidal modulation of the surface. Interestingly, the acceptance angle is not a well-known function such as a catenary but has marked shoulders between relative straight regions. Note that the flatness of the lower portion implies that the trough of the sinusoid approximates the shape of a circle, which has a constant acceptance angle relative to a gap in its surface (as was demonstrated by Euclid).

Discussion

The conclusion from the analysis of the three paradigm cases in Fig. 3 is that, contrary to the appearance of images in Fig. 1, there is no point-source illumination assumption of a sinusoidal Lambertian surface form that would give rise to a periodic luminance profile matching the frequency and phase of the surface waveform (as is perceived by the human observer). The only luminance function that has the observed frequency and phase relative to the peaks of the surface is the diffuse one, and even

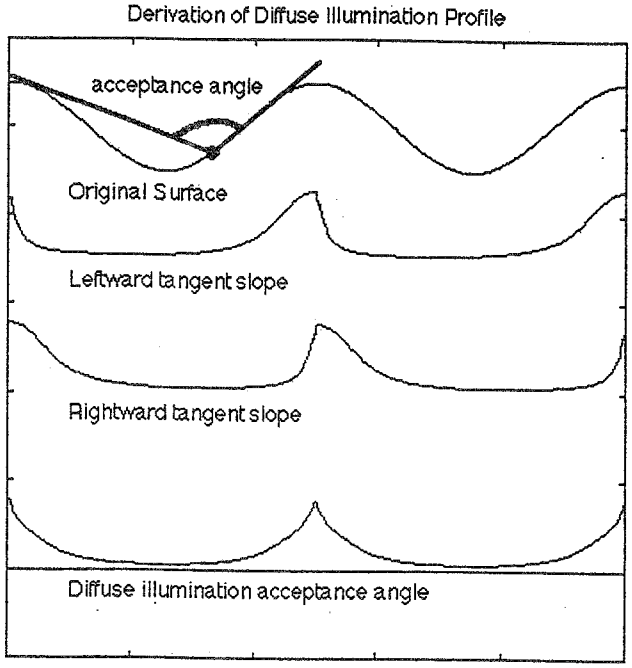


Figure 5. Derivation of diffuse illumination profile for the sinusoidal surface (a); (b) surface tangent to the left of each point along surface; (c) surface tangent to the right of each point along surface; and (d) net acceptance angle at each point.

it is much more cuspy than a sinusoid. It therefore seems clear that the human observer is defaulting to a diffuse illumination assumption, in contrast to the point source typically assumed for computer graphic displays.

Perception of Sinusoidal Patterns. With the analysis in hand, we may now analyze the perception of the patterns of Fig. 1. The most important result is that these patterns do give pronounced depth perceptions, even though they are qualitatively incompatible with any position of point-source illumination. These reports correspond most closely to the diffuse reflectance profile of Fig. 2 (bottom curve), as looking like a surface with peaks at the positions of the luminance peaks. However, the case where the brightness profile [Fig. 1(b)] looks most sinusoidal corresponds to the case where the perceived surface has the most sinusoidal shape. This seems odd because a sinusoidal surface is predicted to have a much more peaked luminance distribution according to the diffuse illumination assumption [Figs. 3(d) and 4(d)].

Note that typical deviations from the Lambertian and the diffuse assumptions will both enhance the discrepancy. If the surface had a reflectance function that is more focused than the Lambertian, it would tend to increase the luminance in the direction of the observer and hence make the peaks of the assumed surface brighter relative to the rest. Similarly, if the illumination source were more focused than a pure diffuse source, it would introduce a second-harmonic component into the reflectance function similar to Fig. 3(c), which would again enhance the peaks and also introduce a bright band in the center of the dark strips. Hence, the diffuse illumination function at the bottom of Fig. 2 is the least peaked function to be expected from any single illumination source.

Role of Perceptual Response Compression. Human vision is, of course, not linear as a function of image luminance L but shows a saturating compression of the internal response R that seems to be most closely approximated by a hyperbolic function (like the Naka–Rushton equations for receptor response saturation), as described in Chan et al.[6,7] and Tyler and Liu.[8] The optimal equation was of the form

$$R=\frac{a}{L+\sigma}. \tag{3}$$

Figure 6 illustrates how such a brightness compression behavior can result in an output that approximates the original surface shape. For a sinusoidal surface [Fig. 6(a)] the diffuse reflectance function under Lambertian assumptions is the peaky function of Fig. 6(b). The effect of a hyperbolic compression on this waveform is shown in Fig. 6(c) to result in an approximately sinusoidal output waveform. For comparison, the effect of the same hyperbolic compression on a straightforward sinusoidal waveform is shown in Fig. 6(d), appearing strongly asymmetric in terms of the peak versus trough shapes. It is thus plausible that the shape-processing system could use the compressed brightness signal as a simple means of deriving the original surface shape from the diffuse reflectance profile.

If the visual system does indeed use its inbuilt brightness compression as a surrogate for a more elaborate reconstitution algorithm of the shape from shading under diffuse illumination assumptions, the approximation should work for other typical surface waveforms. One example to test this hypothesis is a cylindrical waveform corresponding to a one-dimensional version of the sphere that is used widely in computational vision (and which corresponds to the most-simplified form of an isolated object in the world).

A cylindrical waveform is depicted in one-dimensional cross-section in Fig. 7(a), although the vertical axis is extended relative to a purely circular cross-section. The subsequent panels, in the same format as Fig. 6, show the diffuse reflectance profile, the effect of brightness saturation on this profile, and a simple sinusoid with the same degree of compression. Notice that the saturated diffuse profile again looks similar to the surface waveform, supporting the idea that the brightness-compressed signal can generally act as a surrogate for the back-computation of the surface waveform. In this case, the compressed sinusoid looks somewhat similar to the surface waveform also, which may explain why the linear sinusoid of Fig. 1(a) resembles a ring of conical "dunce caps" (because a cone is a version of a cylinder with a converging diameter). If the visual system treats the brightness-compressed signal as an approximation to the depth profile of the object under diffuse illumination, any object that generates a similar signal after brightness compression should appear to have a similar shape.

Finally, some brief thoughts on the different qualities of surface material perceived in Fig. 1. Given that the image that appears Lambertian is the one that resembles the ring of dunce caps with circular cross-section, it may be that the visual system has a Bayesian constraint to prefer a solution that corresponds to such discrete objects rather than a continuously deformed surface. If so, shape reconstructions that deviated from such a circular cross-section (in the absence of explicit contour cues) may tend to be interpreted as deviations from the Lambertian assumption rather than deviations from the assumption of circular cross-section. It is not intended for the present work to provide an empirical analysis of this question but merely to frame the hypothesis.

Conclusion

The object interpretation of the spoke patterns of Fig. 1 is not consistent with the default assumption of any

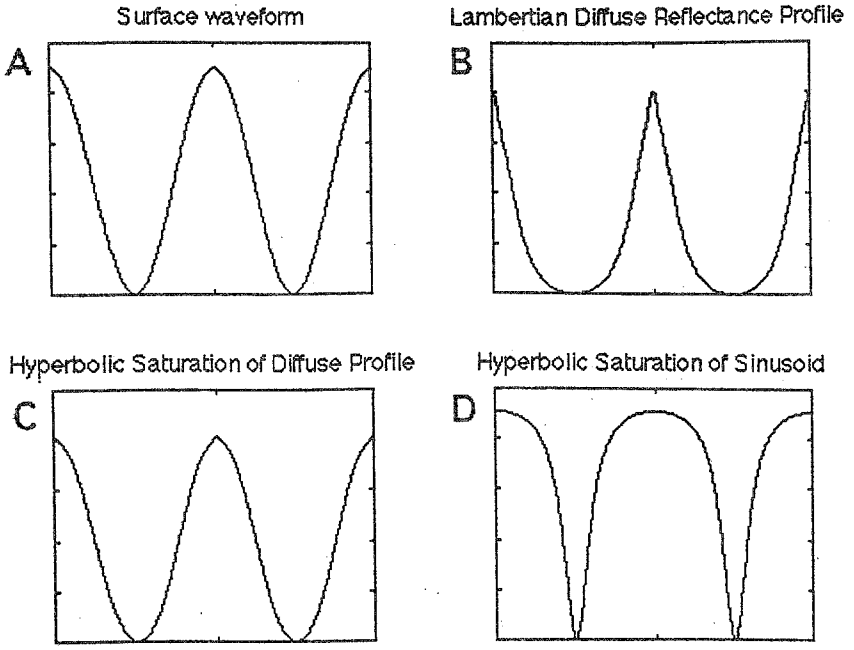


Figure 6. Role of response compression in the interpretation of depth from shading. (a) Sinusoidal surface shape; (b) net reflectance profile assuming diffuse illumination and Lambertian reflectance function; (c) Perceived brightness signal after hyperbolic saturation. Note similarity to original surface waveform; (d) same degree of hyperbolic saturation applied to a sinusoidal signal, to illustrate how much brightness distortion is perceived in Fig. 1(a) under high illumination.

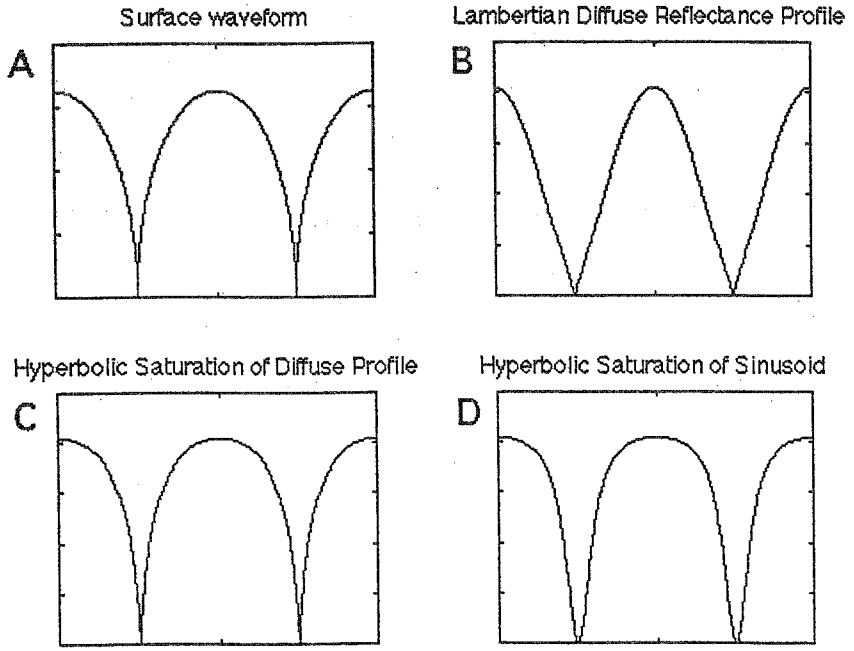


Figure 7. A second example of response compression in the interpretation of depth from shading. (A) Cyclic surface shape. (B) Net reflectance profile assuming diffuse illumination and Lambertian reflectance function. (C) Perceived brightness signal after hyperbolic saturation. Note similarity to original surface waveform. (D) Same degree of hyperbolic saturation applied to a sinusoidal signal, to illustrate similarity of result to (A) and (C).

unidirectional light source but implies a diffuse illumination (as if the object were looming out of a fog). The existence of such a default for human vision of shape from shading has not been previously described to our knowledge. Note that similar percepts are obtained for linear sinusoids of high contrast (such as a "stack of cigarettes"), although the sense of shape-from-shading is weaker initially. No-one ever seems to see a linear sinusoid in a rectangular aperture according to the predictions of Fig. 3(b) for a local illumination source, even though there is now no orientational symmetry to force a symmetric source illumination Thus, the default to diffuse illumination appears to be general unless specific cues imply an oriented source (e.g., Ramachandran).[9]

Given default diffusion, the depth interpretation is consistent with the hypothesis that the visual system uses the compressed brightness profile directly as the neural signal for perceived shape. It is shown that this equivalence is a reasonable approximation to computing the diffuse Lambertian illumination function for this surface. This match provides the visual system with a rough-and-ready algorithm for shape reconstruction without requiring elaborate back-calculation of the brightness compression and integral angle-of-acceptance functions through which the diffuse illumination image was built.

Acknowledgment. Supported by NEI grant No. 7890.

References

1. (a) M. S. Langer and S. W. Zucker, Casting light on illumination: a computational model and dimensional analysis of sources, *Comp. Vis. Image Understand.* **65**, 322–335 (1997); (b) M. S. Langer and S. W. Zucker, Shape from shading on a cloudy day, *J. Opt. Soc. Am.* **11**, 467–478 (1994).
2. V. V. Barun, Imaging simulation for non-Lambertian objects observed through a light-scattering medium. *J. Imaging Sci. Technol.* **41**, 143–149 (1997).
3. D. I. A. Macleod, D. R. Williams and W. Makous, A visual nonlinearity fed by single cones, *Vis. Res.* **32**, 347–363 (1992).
4. R. D. Hamer and C. W. Tyler, Phototransduction: Modeling the primate cone flash response, *Vis. Neurosci.* **12**, 1063–1082 (1995).
5. A. J. Stewart and M. S. Langer, Towards accurate recovery of shape from shading under diffuse lighting, *IEEE Trans. Patt. Anal. Mach. Intell.* **19**, 1020–1025 (1997).
6. H. Chan and C. W. Tyler, Increment and decrement asymmetries: Implications for pattern detection and appearance. *Soc. Inf. Displ. Tech. Dig.* **23**, 251–254 (1991).
7. H. Chan, C. W. Tyler, P. Wenderoth, and L. Liu, Appearance of bright and dark areas: An investigation into the nature of brightness saturation, *Investigative Ophthalmology and Visual Science, Suppl.* B, 1273 (1991).
8. C. W. Tyler and L. Liu, Saturation revealed by clamping the gain of the retinal light response, *Vis. Res.* **36**, 2553–2562 (1996).
9. V. S. Ramachandran, The perception of depth from shading, *Sci. Am.* **269**, 76–83 (1988).