Perception, 2000, volume 29, pages 649 – 660

DOI:10.1068/p3060

Depth discrimination from shading under diffuse lighting

Michael S Langer¶, Heinrich H Bülthoff
Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, Tübingen 72076, Germany;
¶ also at NEC Research Institute, 4 Independence Way, Princeton, NJ 08540, USA;
e-mail: michael.langer@tuebingen.mpg.de
Received 10 September 1999, in revised form 18 February 2000

Abstract. The human visual system has a remarkable ability to interpret smooth patterns of light on a surface in terms of 3-D surface geometry. Classical studies of shape-from-shading perception have assumed that surface irradiance varies with the angle between the local surface normal and a collimated light source. This model holds, for example, on a sunny day. One common situation in which this model fails to hold, however, is under diffuse lighting such as on a cloudy day. Here we report on the first psychophysical experiments that address shape-from-shading under a uniform diffuse-lighting condition. Our hypothesis was that shape perception can be explained with a perceptual model that "dark means deep". We tested this hypothesis by comparing performance in a depth-discrimination task to performance in a brightness-discrimination task, using identical stimuli. We found a significant correlation between responses in the two tasks, supporting a dark-means-deep model. However, overall performance in the depth-discrimination task was superior to that predicted by a dark-means-deep model. This implies that humans use a more accurate model than dark-means-deep to perceive shape-from-shading under diffuse lighting.

1 Introduction

For centuries, artists have studied how smooth patterns of light on a surface can provide visual cues to surface shape. Shading is now considered to be a fundamental visual cue along with binocular disparity, texture, and contour. While it is known that the visual system uses shading to perceive surface shape, it is not yet known how the visual system does so. Previous studies of shape-from-shading perception have concentrated mainly on a sunny day lighting condition, in which a surface is illuminated by a collimated light source (Todd and Mingolla 1983). In this paper, we address an alternative lighting condition that is very common in nature but that has received relatively little attention, namely diffuse lighting such as on a cloudy day. We begin by reviewing how shading differs under collimated lighting versus diffuse lighting. For simplicity, we restrict the discussion to a Lambertian surface of uniform reflectance.

A simple model of shading under collimated lighting is as follows. Let the image coordinates be (x, y) and let $\boldsymbol{N}(x, y)$ be the unit surface normal corresponding to position (x, y). Let the light source be collimated and represented by a vector $\boldsymbol{L}$ and let $v(x, y, \boldsymbol{L})$ be a function that takes the value 0 or 1 depending on whether the light source is visible from the surface point at (x, y), that is whether the corresponding surface point lies in shadow or not. The surface irradiance at (x, y) may be modeled as

$$I(x, y) = v(x, y, \boldsymbol{L})\boldsymbol{N}(x, y)\cdot\boldsymbol{L}. \tag{1}$$

Standard qualifiers to this model are as follows. First, since no real surface has exactly Lambertian reflectance, the reflected radiance typically depends on the viewing position (Horn 1975, 1977). By assuming Lambertian surfaces, we ignore this dependence and consider the surface irradiance only. Second, although a scene may indeed have a well-defined light-source direction, there may be a diffuse component to the illumination as well. For example, on a sunny day, one cannot ignore the illumination from the blue sky. Third, surface elements are illuminated not only by light sources but also by each other via interreflections. Given these qualifiers, one must regard the above model as a 'rough and ready' approximation to shading on a sunny day.

While the above model provides a good approximation to shading if there is indeed a dominant light-source direction, it is a poor model of shading under diffuse lighting. Under diffuse lighting, surface irradiance depends primarily on the angle of the diffuse source that is visible from each point on the surface, rather than on the local surface normal (Langer and Zucker 1994). Intuitively, the hills of a surface tend to be brighter than the valleys under diffuse lighting, since a greater fraction of the diffuse source is visible from the hills than from the valleys.

We can formally model shading under diffuse lighting by adapting the model of equation (1) as follows. [For a more complete treatment, see for example Moon and Spencer (1981).] Again we assume Lambertian reflectance and ignore interreflections. Let the diffuse source be uniform over all directions on the unit sphere—a Ganzfeld with unit radiance. Let $\theta(x, y)$ be the solid angle of the diffuse source that is visible from position (x, y) and let $\mathrm{d}\Omega$ denote an infinitesimal angle centered at a unit direction vector $\boldsymbol{u}$. Surface irradiance $I(x, y)$ may be modeled by integrating over directions $\boldsymbol{u} \in \theta(x, y)$,

$$I(x, y) = \frac{1}{\pi} \int_{\theta(x, y)} \boldsymbol{N}(x, y) \cdot \boldsymbol{u} \, \mathrm{d}\Omega \, . \tag{2}$$

Recall that in the collimated-lighting model, shadowing is a binary phenomenon—a point either lies in shadow or it does not. In the diffuse-lighting model, shadowing is a continuous phenomenon—many points lie partly in shadow.

Since shading differs so dramatically between collimated lighting and diffuse lighting, it is fair to ask whether these differences are manifest in the perception of shape from shading. On the one hand, one might expect no perceptual differences at all. Since both lighting conditions are common in nature, one might expect the visual system to have evolved mechanisms in each case (possibly separate ones) and that these mechanisms would be sufficient to perceive shape up to a given level of accuracy for a given task. On the other hand, the relationship between shape and shading is very different under the two lighting conditions, and so one might expect that certain properties of shape are inherently more salient under one condition than the other. For example, a sphere illuminated by a Ganzfeld has constant irradiance and thus is indistinguishable from a disk, and yet the same sphere illuminated under collimated lighting contains shading and thus appears curved.

To clarify how differences in perceived shape-from-shading can depend on the type of lighting condition, we decided to carry out a controlled study. In this paper we present three experiments. In the first, we examine qualitative shape perception, namely how well observers can distinguish hills from valleys under collimated lighting versus diffuse lighting. The second experiment tests a specific hypothesis that observers perceive shape-from-shading under diffuse lighting by identifying darker points as deeper. The third experiment is a control on the second.

2 Experiment 1: hill versus valley

The aim of the first experiment was to compare local qualitative shape perception under collimated lighting versus diffuse lighting. Observers were asked to judge whether isolated marked points on a surface were "on a hill" or "in a valley" (see figure 1).

2.1 *Method*

2.1.1 *Stimuli.* Surface shapes were defined by modulating the radius of a 1024×1024 square-mesh cylinder with low-pass-filtered white noise. The cutoff frequency of the noise was 60 cycles. The modulated cylinder was sliced into eight disjoint disks (short

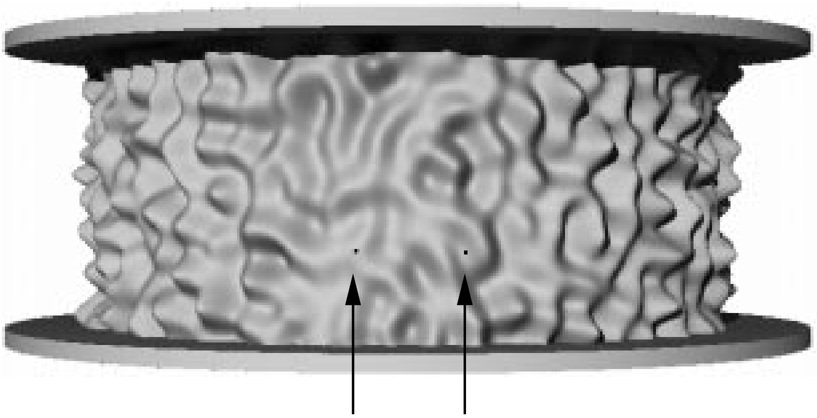

Figure 1. A rendered image is shown in which the light source is collimated (sunny day). The direction of the source is from the above-left. Two examples of single probe points are marked on the surface. The surface point on the left lies on a hill and the surface point on the right lies in a valley.

cylinders), each defined by a 128×1024 mesh. Each of these eight disks was rendered from four different virtual viewpoints, by perspective projection. The viewpoints were defined as follows. Let (r, x, ϕ) be the coordinate system of the original cylinder, with radius r, height x, and rotation angle ϕ. The radial distance r to the viewpoint was always equal to the height of the original cylinder, and the viewing direction was always radial inward. The height x of each viewing position was at the center of the corresponding disk. As a result, each disk subtended a visual angle of about 20 deg $\times$ 10 deg with respect to the virtual viewpoint. Four viewing directions were used for each disk, separated by 90° increments of ϕ. This yielded 32 viewing positions in total, ie four positions for each disk. Finally, each disk was framed by two thin flat disks, which provided perspective cues.

The surfaces were rendered with the use of RADIANCE software (Ward Larson and Shakespeare 1998). Surfaces were Lambertian with reflectance of 30% and interreflections were computed to two bounces. Two different types of lighting conditions were used: a diffuse condition in which the source was a uniform sphere surrounding the object, and four different collimated-source conditions. The directions of the collimated sources were always 11° from the viewing direction, and in one of four diagonal directions with respect to horizontal and vertical. Specifically, when the viewing direction was (1, 0, 0), the four collimated-source directions were the four combinations of $(5, \pm 1, \pm 1)$. We refer to these four source directions as above-left, above-right, below-left, and below-right. Each source direction was close enough to the line of sight that cast shadows did not appear in the central region of the surface. A weak diffuse-source component (85% collimated + 15% diffuse) was added in each collimated-source condition to simulate secondary illumination in the scene.

In the diffuse-lighting condition, the renderings contained pixel noise which was due to the stochastic sampling of the diffuse source by the RADIANCE software. We estimated the magnitude of this noise by separately rendering spherical concavities of various depths, and comparing the irradiances of the rendered spherical concavities to ground truth which is known for the spherical concavity geometry (Moon and Spencer 1981; Langer 1999). For the rendering parameters we used, the irradiance noise was roughly 1% (RMS) of the maximum irradiance of the surface, and this noise was approximately independent of the depth of the concavity. Noise of roughly the same magnitude was introduced by interreflections.

The rendered images were presented achromatically on a CRT monitor that was calibrated so that screen luminance was linearly related to rendered surface irradiance. Images were normalized so that all surfaces had the same maximum luminance. This maximum luminance was always 10% less than that of a uniform white background against which the surfaces were viewed.

Observers wore an eye patch over the non-dominant eye and viewed the stimuli in a dimmed room at a distance of 2 m. Each surface subtended a viewing angle of 10 deg $\times$ 4 deg. This viewing distance slightly exaggerated the perspective. The distortion was greatest near the flanking regions of the stimulus and was zero at the center of the image. We used the larger viewing distance to reduce the visibility of noise in the rendering.

Probe points were chosen from the central 2 deg $\times$ 2 deg region of the 10 deg $\times$ 5 deg image. In this region, no occlusion contours were present nearby (Howard 1983; Todd and Reichel 1989) and the surface was roughly frontoparallel (Reichel and Todd 1990). Hence, the only local cue to shape was the shading. Probe points were chosen according to the following conditions. The surface normal at each probe was required to be less than 12° from the line of sight, and the principal curvatures at each probe were required to be either both positive (hill) or both negative (valley). We set the threshold on curvature to be as large as possible, subject to the constraint that at least five hill probes and five valley probes were obtained for each of the 32 viewpoints.

2.1.2 *Observers.* Twenty observers participated (age 18 – 30 years) and were paid at the rate of 15 DM per hour. All observers had normal or corrected-to-normal vision.

2.1.3 *Procedure.* Each trial consisted of a grey silhouette presented for 0.2 s, followed by a probe that was superimposed on the silhouette for 0.8 s during which time the observer made an eye movement to the probe. A rendered image then replaced the silhouette and the probe remained superimposed on the image. The area of the probe was reduced by a factor of two when the rendered image appeared, so that as little of the local shading would be covered as possible. The size of the probe as it appeared on the image is shown in figure 1.

For each probe, the observer's task was to judge whether the probe was "on a hill" or "in a valley" (see figure 1). The observer had a maximum of 1.2 s to respond by pressing on a left or right response key. No feedback was given. Responses longer than 1.2 s were discarded and the same item was repeated at the end of the same block. The duration of 1.2 s was chosen because it was long enough that most observers in a pilot study were comfortable with the task, and short enough that observers could not solve the task 'cognitively', for example by counting hills and valleys from an occluding contour at the flanking region of the image.

Each observer ran 320 trials, corresponding to the 10 probe points for each of the 32 viewing positions. The 320 trials were divided into ten blocks of 32 trials each. The five lighting conditions (one diffuse + four collimated) were balanced over all trials. Half of the observers ran a mixed condition in which the 320 trials were ordered randomly, and half ran a blocked condition in which the lighting condition was constant within each block, but varied in a cycle between blocks. No effect was found for mixed versus blocked conditions and the data were subsequently pooled. We present only the pooled data.

Prior to the experiment, each observer ran a practice session of a single block. No feedback was given. The experiment lasted roughly 15 min.

2.2 *Results and discussion*

In a debriefing session, observers reported that the allotted response time had been too short for them to attend to the global properties of the stimulus, and that they had attended to a local neighborhood of each probe only. They also reported that they were comfortable with the task. Indeed, observers were rarely unable to respond within the allotted time. Finally, observers consistently reported that they had been unaware that lighting conditions had been manipulated. Some observers reported that certain images had lower contrast than others, but they were unaware of the cause of this difference. Images in the diffuse-source condition did have lower contrast than those in the collimated-source conditions.

The percentages of correct scores for each of the five lighting conditions are shown in figure 2. The scores were significantly higher when the collimated source was above the line of sight than when it was below the line of sight ($F_{1,18} = 86$, $p < 0.0001$). They were also higher when the collimated source was from the left than from the right (Sun and Perona 1998) although this effect was not significant ($F_{1,18} = 2.2$, $p = 0.15$). In the diffuse-lighting condition, the percentage of correct scores was significantly above chance ($t_{19} = 11.6$, $p < 0.0001$) and was as high as under the best collimated-source condition tested (above – left).

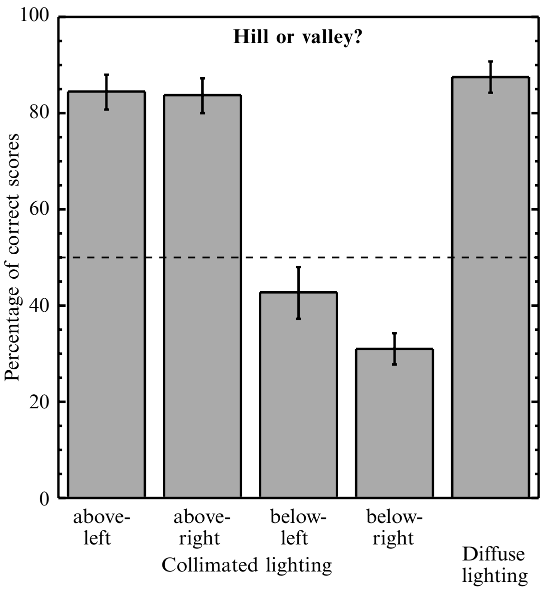

Figure 2. Percentage of correct scores in experiment 1, in which observers judged whether a single marked point was on a hill or in a valley. Five lighting conditions are compared. Error bars indicate the standard error of the mean.

The results in the collimated-source conditions are consistent with previous studies, that the visual system prefers a light from above over a light from below (Rittenhouse 1786; Brewster 1826; Berbaum et al 1983, 1984; Ramachandran 1988; Howard et al 1990). The results in the diffuse-lighting condition, however, cannot be explained by a preference for light-from-above. Under diffuse lighting, an equal amount of light arrives from above and from below. An alternative model is thus needed to account for the results in the diffuse-lighting condition.

One possible model is suggested by a heuristic used in artistic rendering of drapery surfaces. To create an impression of a concave fold in a drapery, it is suggested that the artist should render deeper points as darker (Nicolaides 1941). This perceptual phenomenon that dark appears deep has been known for centuries. For example, Leonardo da Vinci wrote that "among bodies equal in size and distance, that which shines the more brightly seems to the eye nearer" (MacCurdy 1938). A tendency to see dark as deep has been found in formal studies of shape perception in line drawings (Dosher et al 1986), and in sunny-day shape-from-shading (Christou and Koenderink 1997). It has even been hypothesized that dark-means-deep is used as a default for shape-from-shading under diffuse lighting (Langer and Zucker 1994; Tyler 1998), although this hypothesis has not been formally tested. Our second experiment was designed to test this hypothesis directly.

3 Experiment 2: depth discrimination

A dark-means-deep model only partly captures the physics of shading under diffuse lighting. An example of how the dark-means-deep model fails is that valleys of surfaces often contain local irradiance maxima at their deepest points (see figure 3). Such maxima occur for several reasons. First, the surface normal may turn toward the visible part of the source, yielding a classical $\boldsymbol{N} \cdot \boldsymbol{L}$ shading effect. Second, there may be a local maximum in the visible angle of the diffuse source. A third possibility, which can occur for non-Lambertian surfaces (hence not for the surfaces we use), is that there may be a specular reflection of the diffuse source from the valley. Fourth, local maxima can occur in a valley because of interreflections.

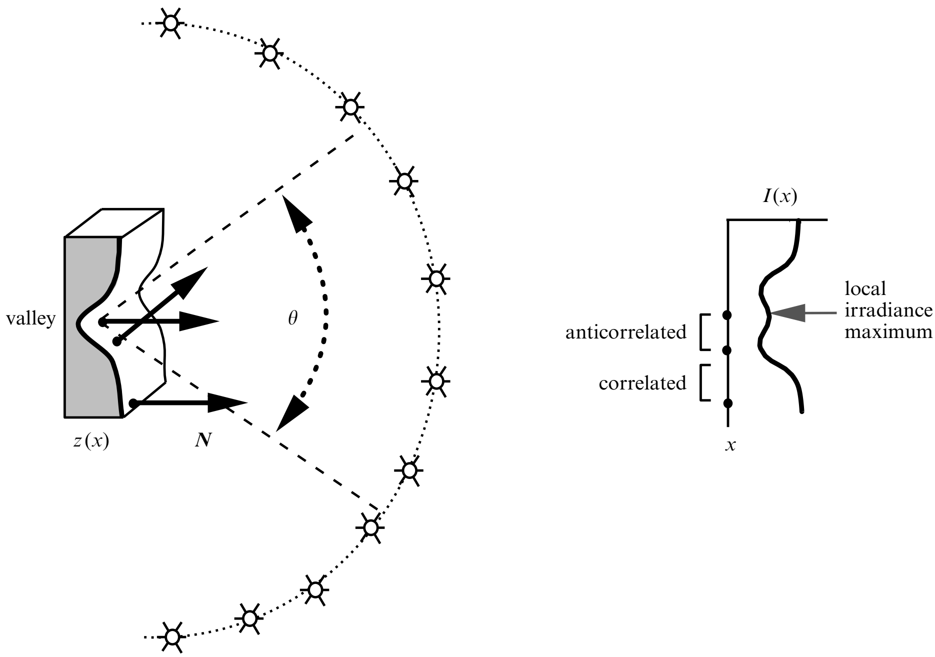

Figure 3. A surface with a height function $z(x)$ is shown. Three surface points are marked: a point at the bottom of the valley, a point on the side of the hill, and a point on top of the hill. The point at the bottom of the valley has the same surface normal $\boldsymbol{N}$ as the point on the top of the hill, yet the valley point has a smaller visibility angle θ and hence is darker. A local irradiance maximum occurs at the bottom of the valley since there the surface normal turns directly towards the visible source. The point on the bottom of the valley and the point on the side of the hill form an anticorrelated pair. The points on the side and on top of the hill form a correlated pair.

If the visual system were to rely on a dark-means-deep model to perceive shape-from-shading under diffuse lighting, then local irradiance maxima in valleys should yield illusory local hills. Our second experiment tested for this illusion. The experiment used the same stimuli as the first, but now pairs of points were marked and for each pair the task was to judge which of the two points was higher, that is which was closer to the observer along the line of sight (Todd and Reichel 1989; Koenderink et al 1996b). The idea of the design was to create two conditions, one of which was consistent with a dark-means-deep model, and the other not. In the *correlated* condition, the darker point is deeper. In the *anticorrelated* condition, the brighter point is deeper (see figure 4). If the visual system were to rely on a dark-means-deep model for discriminating depth, then observers should score above chance in the correlated condition and below chance in the anticorrelated condition, and performance should be at chance overall. Further details of the method now follow.

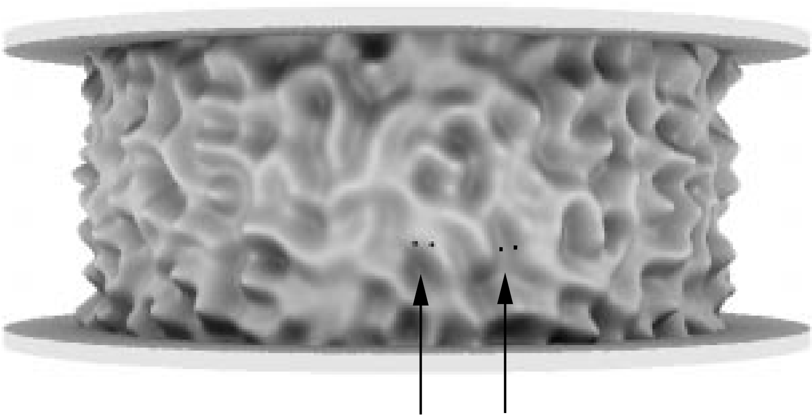

Figure 4. The same surface as in figure 1 but now shown under the diffuse-lighting condition (cloudy day). Two probe pairs are marked (see experiments 2 and 3). In the correlated pair, the surface point on the left is darker and deeper than the point on the right. In the anticorrelated pair, the surface point on the left is brighter and deeper than the point on the right.

3.1 *Method*

3.1.1 *Stimuli.* Three lighting conditions were used: the diffuse-source condition and two collimated-source conditions (above-left and below-right). For each of the 32 viewing positions and for each lighting condition, four pairs of probe points were chosen, again from the central square region of the image. The pairs were chosen as follows. We required that surface height varied monotonically between the two points (Todd and Reichel 1989) and that both the irradiances and the heights of the two points differed by a given amount. Let I_1 and I_2 be the irradiances and let z_1 and z_2 be the heights of two points, respectively, where height is defined by the distance between the surface point and the viewpoint. Each probe pair was required to have an irradiance contrast $|(I_1 - I_2)/(I_1 + I_2)|$ and a height difference $|z_1 - z_2|$ that was within 60% to 80% of the standard deviation over the central test square of the image. This range of values was sufficiently broad to allow two correlated and two anticorrelated pairs for each viewpoint and for each lighting condition. This yielded 384 trials in total.

3.1.2 *Procedure.* The procedure was the same as in the first experiment, except that there were twelve rather than ten blocks.

3.1.3 *Observers.* Seventeen new naive observers participated, eight in a mixed condition and nine in a blocked condition. Again no effect was found for mixed versus blocked conditions and we present only the pooled data.

3.2 *Results and discussion*

The results are shown in figure 5. In the collimated-source conditions, the percentage of correct scores was higher when the source was above-left than below-right ($F_{1,15} = 88.7$, $p < 0.0001$), as expected. In the diffuse-lighting condition, the score was higher in the correlated condition than in the anticorrelated condition, and the overall score in the diffuse-lighting condition was 63%. This is significantly above chance (paired t-test, $t_{16} = 8.1$, $p < 0.0001$) and suggests that observers were not simply relying on a dark-means-deep model.

An alternative interpretation, however, is that observers did rely on the dark-means-deep model but that they could not discriminate the surface brightness equally well in the correlated and anticorrelated conditions. This is plausible since anticorrelated points tended to have lower luminance than correlated points. (Recall that anticorrelated

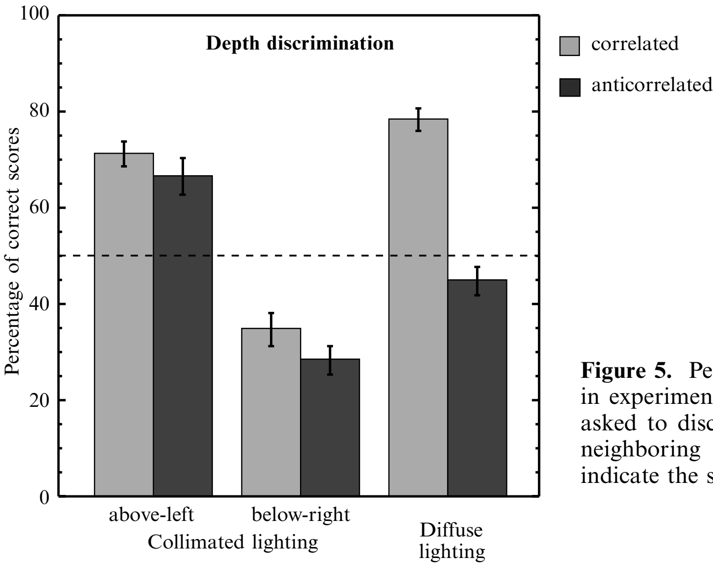

Figure 5. Percentage of correct scores in experiment 2 in which observers were asked to discriminate the depths of two neighboring surface points. Error bars indicate the standard error of the mean.

points tended to occur in valleys and correlated points tended to occur on hills—see figure 3.) Since the signal-to-noise ratio in the rendering was lower in the anticorrelated condition than in the correlated condition, observers may have guessed at their response more often in the anticorrelated condition and this could have led to above-chance performance overall. To rule out this possibility, a third experiment was needed.

4 Experiment 3: brightness discrimination

The third experiment was identical to the second, but now the task was to discriminate the brightness (perceived luminance) of each pair of points, rather than the depth.

4.1 *Method*

Eleven new naive observers participated. Each ran the mixed condition in which the lighting condition varied randomly from trial to trial.

4.2 *Results and discussion*

If observers used a dark-means-deep model in the depth-discrimination task (experiment 2), then the responses in that task should be similar to the responses in the brightness task. Certain trials in the depth task should have yielded a correct response and others should have yielded an incorrect response, and this tendency should have been the same as in the brightness task.

In the diffuse-lighting and anticorrelated condition, we indeed found a significant trial-by-trial correlation between responses in the depth and brightness tasks ($r = 0.544$, $p < 0.0001$). Despite this correlation, however, observers were 77% correct in the brightness task (see figure 6) but only 55% ‘correct’ in the depth task (see figure 5), where the quotation marks indicate the hypothesis that observers used a dark-means-deep model in the depth task. This difference is significant ($F_{1,26} = 27$, $p < 0.0001$) and implies that observers in the depth task perceived the brighter point to be deeper more often than if they had relied entirely on their brightness percepts, that is on a dark-means-deep model. Thus, the dark-means-deep model is too weak to fully account for observers’ responses in the depth-discrimination task.

The role of the dark-means-deep model thus remains unclear. One subtle issue is the brightness task itself was non-trivial—observers were only 80% correct in this task. We have already discussed one of the sources of this difficulty, namely that there was noise in the rendering. Two other sources of difficulty are worth considering as well.

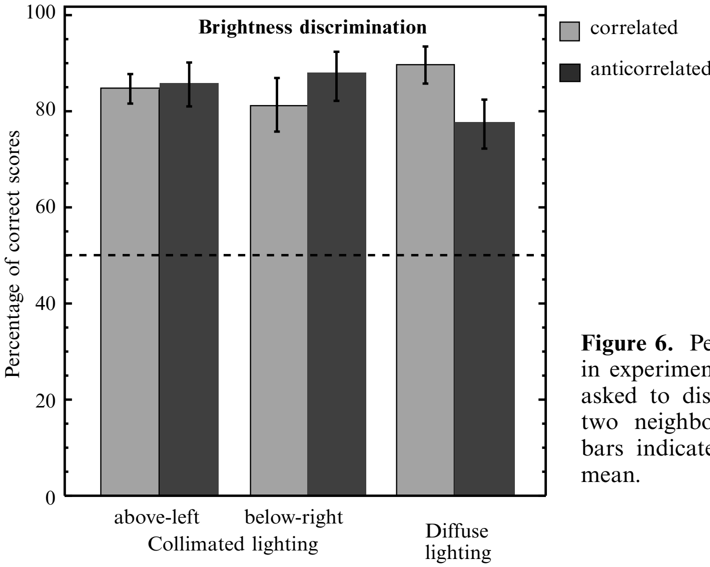

Figure 6. Percentage of correct scores in experiment 3 in which observers were asked to discriminate the brightness of two neighboring surface points. Error bars indicate the standard error of the mean.

First, contrast sensitivity in complex images depends on the local shading in a way that is still poorly understood (Peli 1990). We tried to equate the perceived contrast between the correlated and anticorrelated conditions by using Michelson contrast (see section 3.1.1), but this may not have been sufficient. Second, the dark-means-deep model may have played a role not only in discriminating depth, but also in discriminating brightness. For example, in trials in which the brightness gradient between two probes was difficult to resolve, observers may have relied on their depth percept instead when making the brightness judgment. This is plausible since depth percepts would incorporate shading from a larger neighborhood of the probes. Further experiments are thus needed to clarify the role played by the dark-means-deep model under diffuse lighting.

Finally, we note that in the collimated-source conditions we also found a significant trial-to-trial correlation between responses in the depth and brightness tasks, both for above-left ($r = 0.21$, $p = 0.017$) and for below-right ($r = 0.28$, $p = 0.011$). Again, this correlation cannot fully account for the observers' responses in the depth-discrimination task, since the percentage of correct scores was much higher when the light source was from above than from below. Nonetheless, the dark-means-deep model did appear to play a role here as well (see Koenderink et al 1996a; Christou and Koenderink 1997 for related findings).

5 Computational models

Experiments 2 and 3 showed that a dark-means-deep model is too weak to account for shape-from-shading perception under diffuse lighting, since observers scored higher in the depth-discrimination task than if they had relied on this model. What alternative model could explain observers' performance? To explore this question, we compared the responses of human observers in the depth-discrimination task to the responses of several computational models.

We first considered two computer-vision algorithms that have been proposed for the diffuse-lighting condition. The first algorithm (LZ) (Langer and Zucker 1994) is based on an approximation of equation (2) in which irradiance depends only on angle $\theta(x, y)$ of the visible source, and not on the surface normal. The second algorithm (SL) (Stewart and Langer 1997) accounts for the visible angle of the diffuse source and also for surface-normal effects and interreflections. It was shown by Stewart and Langer (1997) that SL typically recovers surface shape more accurately than LZ.

We ran each algorithm on a 2 deg × 2 deg square neighborhood of each probe pair, yielding a depth map for this neighborhood. Unlike the human observers, the algorithms did not 'see' the black probe points but rather the unobscured 2 deg × 2 deg square image. The response of each algorithm in a given trial of the depth-discrimination task was determined by comparing the recovered depths of the two points.

The percentages of correct scores for LZ and SL are shown in table 1 along with the trial-by-trial correlation of responses between human observers and each of the two algorithms. Several observations are in order. First, both algorithms were 100% correct in the correlated condition but made errors in the anticorrelated condition. The errors by SL are surprising, since SL uses the correct shading model to recover shape—that is it uses the same shading model as was used in the rendering (modulo the noise). The results suggest that the anticorrelated condition is inherently difficult, presumably in part owing to the lower signal-to-noise ratio in that condition (recall section 3.2).

Table 1. The percentage of correct scores for three computational models are shown, along with the correlation coefficient r between the models' responses and the human observers' responses in the depth-discrimination task. None of the models shows a significant correlation at the $p < 0.05$ level.

Model	Condition		r
	correlated	anticorrelated	
LZ	100	27	0.13
SL	100	77	−0.03
Blur	80	47	0.07

A second observation is that LZ was below chance in the anticorrelated condition. That is, LZ was systematically (though not always) fooled by the local irradiance maxima in valleys. This finding implies that the surface-normal effects which are not considered by LZ are important for judging shape from shading under diffuse lighting, especially in valleys that contain local irradiance maxima (recall the first paragraph of section 3).

A third observation is that, although the trend in the performance is similar between the human observers and the computational models (namely higher scores in the correlated condition than in the anticorrelated condition), there was no trial-to-trial correlation between the responses of the human observers and those of the two algorithms (third column in table 1). Thus, despite similar overall trends, the responses of SL and LZ do not yield a detailed account of the responses of the human observers.

A final model that we considered is similar to the dark-means-deep model but differs in that the visual system ignores the local luminance maxima in valleys. It does so by blurring the retinal image prior to applying the dark-means-deep model. We found that by using a Gaussian blurring kernel with a standard deviation equal to the distance between probes, such a blurred dark-means-deep model could achieve a percentage of correct responses that is nearly identical to that of the human observers (see table 1). However, as with the two computer-vision algorithms above (but unlike with the pure dark-means-deep model), we found no significant trial-to-trial correlation between the responses of this model and those of the human observers. So, despite the simplicity of this model, it does not seem to account for human performance.

6 Summary

We summarise our main findings as follows. First, observers distinguish the local qualitative shape of surfaces as well under diffuse lighting as under the best collimated-source condition tested. Hence, the common claim that the visual system prefers a light source from above must be qualified. The visual system prefers light from above over light from below, but it does not prefer light from above over all other lighting models. Second, although a dark-means-deep model successfully predicts observer responses in a depth-discrimination task to some extent, observers scored better overall than if they were relying entirely on this model. Third, both human observers and computer-vision algorithms had more difficulty discriminating the depths of nearby points lying near a valley bottom than those lying near a hill top. This may be due to the fact that valleys tend to be darker than hills under diffuse lighting and hence the shading in a valley has a lower signal-to-noise ratio than the shading on a hill. It may also be due to the fact that the relation between shape and shading is more complicated near valley bottoms, because several confounding shape factors may give rise to similar shading effects. Further studies are needed to understand how well the visual system interprets shape from shading under different lighting conditions, what strategies the visual system uses under what lighting conditions, and how the visual system chooses these strategies from a given shading image.

Acknowledgements. The authors thank Isabelle Bülthoff, Volker Blanz, Volker Franz, Zili Liu, James Stewart, and Niko Troje for helpful comments and for technical assistance. This research was supported by the DAAD and by an Alexander von Humboldt Research Fellowship to Michael Langer.

References

Berbaum K, Bever T, Chung C, 1983 "Light source position in the perception of object shape" *Perception* **12** 411 – 416

Berbaum K, Bever T, Chung C S, 1984 "Extending the perception of shape from known to unknown shading" *Perception* **13** 479 – 488

Brewster D, 1826 "On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena" *Edinburgh Journal of Science* **4** 99 – 108

Christou C G, Koenderink J J, 1997 "Light source dependence in shape from shading" *Vision Research* **37** 1441 – 1449

Dosher B A, Sperling G, Wurst S A, 1986 "Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure" *Vision Research* **26** 973 – 990

Horn B K P, 1975 *The Psychology of Computer Vision* chapter "Obtaining shape from shading information" (New York: McGraw-Hill)

Horn B K P, 1977 "Understanding image intensities" *Artificial Intelligence* **8** 201 – 231

Howard I, 1983 "Occluding edges in apparent reversal of convexity and concavity" *Perception* **12** 85 – 86

Howard I, Bergstrom S, Ohmi M, 1990 "Shape from shading in different frames of reference" *Perception* **19** 523 – 530

Koenderink J J, Doorn A J van, Christou C, Lappin J S, 1996a "Perturbation study of shading in pictures" *Perception* **25** 1009 – 1026

Koenderink J J, Doorn A J van, Kappers A M L, 1996b "Pictorial surface attitude and local depth comparisons" *Perception & Psychophysics* **58** 163 – 173

Langer M S, 1999 "When shadows become interreflections" *International Journal of Computer Vision* **34** 193 – 204

Langer M, Zucker S, 1994 "Shape-from-shading on a cloudy day" *Journal of the Optical Society of America A* **11** 467 – 478

MacCurdy E (Ed.), 1938 *The Notebooks of Leonardo da Vinci* (London: Jonathan Cape) page 332

Moon P, Spencer D E, 1981 *The Photic Field* (Cambridge, MA: MIT Press)

Nicolaides K, 1941 *The Natural Way to Draw* (Boston, MA: Houghton Mifflin)

Peli E, 1990 "Contrast in complex images" *Journal of the Optical Society of America A* **7** 2032 – 2040

Ramachandran V, 1988 "Perception of shape from shading" *Nature (London)* **331** 163 – 165

Reichel F R, Todd J T, 1990 "Perceived depth inversion of smoothly curved surfaces due to image orientation" *Journal of Experimental Psychology: Human Perception and Performance* **16** 653 – 664

Rittenhouse D, 1786 "Explanation of an optical deception" *Transactions of the American Philosophical Society* **2** 37 – 43

Stewart A J, Langer M S, 1997 "Towards accurate recovery of shape from shading under diffuse lighting" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19** 1020 – 1025
Sun J, Perona P, 1998 "Where is the sun?" *Nature Neuroscience* **1** 183 – 184
Todd J, Mingolla E, 1983 "Perception of surface curvature and direction of illumination from patterns of shading" *Journal of Experimental Psychology: Human Perception and Performance* **9** 583 – 595
Todd J, Reichel F, 1989 "Ordinal structure in the visual perception and cognition of smoothly curved surface" *Psychological Review* **96** 643 – 657
Tyler C W, 1998 "Diffuse illumination as a default assumption for shape-from-shading in graded images" *Journal of Imaging Science and Technology* **42** 319 – 325
Ward Larson G, Shakespeare R, 1998 *Rendering with Radiance: the Art and Science of Lighting Visualization* (San Mateo, CA: Morgan Kaufmann)