Qualitative shape from shading, highlights, and mirror reflections

Arthur Faisman

School of Computer Science, McGill University, Montreal, QC, Canada

Michael S. Langer

School of Computer Science, McGill University, Montreal, QC, Canada

The human visual system has a remarkable ability to perceive three-dimensional (3-D) surface shape from shading and specular reflections. This paper presents two experiments that examined the perception of local qualitative shape under various conditions. Surfaces were rendered using standard computer graphics models of matte, glossy, and mirror reflectance and were viewed from a small oblique angle to avoid occluding contour shape cues. The subjects' task was to judge whether a marked point on each surface lay on a local hill or valley. In the first experiment, performance was lower for glossy surfaces than matte surfaces, which is contrary to findings in previous studies of quantitative shape. For mirror surfaces, performance was high despite the absence of occluding contours, and performance was increased when the environment map was brighter in the upper hemisphere as in a natural environment. The second experiment examined how subjects resolve a depth-reversal shape ambiguity where surfaces can be either upward or downward facing. An upward-facing surface prior that is known to exist for matte surfaces was also found to exist for glossy and mirror surfaces. Subjects relied entirely on this prior to resolve the depth-reversal ambiguity for matte and glossy surfaces, but relied on perspective cues as well for mirror surfaces.

Introduction

The shape we perceive when looking at a surface depends on many factors other than the shape itself, including the surface reflectance, the illumination, the surface motion, and the viewer's position. In this paper, we address how perceived shape depends on specular reflections as well as their interaction with diffuse Lambertian shading. Specular reflections—also called specularities—include mirror reflections of the surrounding environment (Fleming, Torralba, & Adelson, 2004), as well as specular highlights, which are shiny mirror reflections of light sources such as the sun or a lamp. Specularities occur on many types of surfaces including plastic, polished wood, glossy paper, metal, and water. While both diffuse and specular reflections depend fundamentally on the local orientation of surfaces, specular reflections depend in a more complex way on the illumination, surface shape, and viewing position. As such, the “inverse optics” problem of estimating shape from specular reflections is typically more difficult than the problem of estimating shape from diffuse shading.

Most studies of the perception of shape from specularities have used compact solid shapes such as smoothly perturbed spheres, rendered with specular highlights. For example, early experiments used convex cylinders (Todd & Mingolla, 1983) and ellipsoids (Mingolla & Todd, 1986) and addressed perceived surface curvature or orientation. Subsequent studies used more general solid shapes that also contained concave and saddle regions and used an interactive graphical shape probe called a gauge figure (Koenderink, van Doorn, & Kappers, 1992) to examine local perceived shape. These studies were mainly concerned with quantitative shape perception—namely they measured the accuracy and precision of gauge figure settings.

Several important questions have been addressed in such studies, including whether specular highlights increase or decrease the accuracy of shape probe settings, and how highlights interact with other cues such as binocular disparity and motion. For example, Norman, Todd, and Phillips (1995) found lower errors in orientation settings when stereo or motion cues were present, but no strong differences between the diffuse shading (matte) versus diffuse shading with highlights (glossy) conditions. Todd, Norman, Koenderink, and
Kappers (1997) examined stereoscopically presented static solid shapes and found that local surface orientation was more accurately perceived when the surfaces were rendered with diffuse shading plus highlights than with diffuse shading alone. Norman, Todd, and Orban (2004) also found that diffuse shading plus highlights gave better shape discrimina-
tibility than diffuse shading alone, and that highlights alone gave good discriminairebility. Nefs, Koenderink, and Kappers (2006) used static monocular presentations and found no difference in shape discriminairebility when highlights were added to diffuse shading. Taken together, these findings indicate that highlights sometimes help and, in the worst case, do not impede quantitative shape perception, both when viewed from a single point of view and when viewed in stereo and motion.

In studies that have considered motion or stereo, it has been argued that there is a simple qualitative relationship between the shape and the location of specular highlights (Blake & Brelstaff, 1988; Zisserman, Giblin, & Blake, 1989), namely that highlights produce virtual images that lie beyond a convex surface and in front of a concave surface. However, the quantitative relationship between the shape and location of highlights can be complex in the case of general smooth surfaces (Blake & Bulthoff, 1991; Oren & Nayar, 1996). For this reason, it is surprising that previous studies of quantitative shape perception consistently found that adding highlights to diffuse shading did not hinder shape perception. In a pilot study for this paper (Langer & Faisman, 2012), we observed the opposite trend, namely that performance in a shape task indeed decreased when highlights were added to existing diffuse shading. The present paper re-addresses this question.

We also review studies of shape perception from another kind of specular reflection, namely mirror reflections of the environment. For example, Savarese, Fei-Fei, and Perona (2004) examined perception of qualitative shape from mirror reflections using small patches of convex-, cylindrical-, or saddle-shaped surfaces that did not contain smooth occluding contours. A regular texture was used for the illumination environment. Subjects were asked to choose between the three qualitative shape categories (convex, concave, or saddle), and performed above chance. Fleming et al. (2004, 2009) addressed quantitative shape perception from mirror reflections of complex natural environments, using simple solid shape stimuli and gauge figures. They found that under monocular viewing, subjects could reliably estimate local surface orientation from these mirror reflections. They also presented a novel theory of how shape cues from specular reflections differ from shape cues from texture. Whereas texture produces local image compressions that depend on the first derivative of surface depth with respect to image position (foreshortening), specular reflections produce compressions that depend on the second derivatives of depth, which in turn depend on both the first derivative of depth and on the intrinsic surface curvature. In more recent work (Murty, Welchman, Blake, & Fleming, 2013), it was shown that, under binocular viewing, these shape from specularity cues are often ignored in favor of stereo disparity cues which match a virtual rather than physical surface.

In the above studies, the surfaces were typically compact solid shapes. Such surfaces contain cues beyond shading, texture, and specular reflections that contribute to the perception of shape. For example, smooth occluding contours provide boundary conditions that strongly constrain the perceived shape interior (Beusmans, Hoffman, & Bennett, 1987; Hoffman & Richards, 1984; Huggins & Zucker, 2001; Koenderink, 1984; Koenderink & van Doorn, 1976; Tse, 2002) to be globally convex (Hill & Bruce, 1994). Additionally, for mirror surfaces, the choice of a nonhomogeneous environment map can provide supplemental cues that may be exploited by the visual system. In particular, Fleming and colleagues (Fleming et al., 2004; Murty et al., 2013) used a natural environment map containing an outdoor scene, without consideration of these supplemental cues. Natural environments tend to be brighter in the upper hemisphere, and this causes image intensity to be correlated with surface orientation, namely upward-facing normals yield greater image intensities than downward-facing normals. The ability of the visual system to exploit this cue is investigated in the present study.

We present two experiments that address qualitative shape perception from specular reflections. The task in both experiments is to decide whether marked points on a surface lie on locally convex or concave regions. The experiments use smooth, randomly generated terrain surfaces that consist of many convex, concave, and saddle regions. We choose a viewing position such that occluding contours do not occur in the surface interior and hence cannot be used as a shape cue. This allows us to better isolate how shape is perceived from shading and specular reflections cues.

The first experiment compares qualitative shape judgments for several surface rendering conditions including texture alone, several combinations of diffuse shading and specular highlights, and mirror reflections of an environment. There are several goals including those mentioned above. The first is to confirm and elaborate on our earlier finding of our pilot study (Langer & Faisman, 2012) that adding highlights to diffuse shading decreases performance in a qualitative shape task. Secondly, we examine whether qualitative shape perception from mirrors depends on the envi...
environment map, in particular, on whether performance improves when the map is dominated by light from above. We also examine whether motion cues improve performance.

The second experiment addresses a different issue that arises in qualitative shape perception, namely the depth-reversal ambiguity. The classical version of the ambiguity states that a surface with diffuse reflectance that is viewed under orthographic projection produces the same image when it is reversed in depth and illuminated from a mirror-symmetric direction. This ambiguity is the basis of the hollow mask illusion (Hill & Bruce, 1994), for example, and the ambiguity has been studied in local shape perception as well (Langer & Bulthoff, 2001; Liu & Todd, 2004; Mamassian & Landy, 1998; Reichel & Todd, 1990). We show that, in theory, an identical ambiguity holds for specular reflections as well (see Appendix). It is known that the visual system resolves this ambiguity for diffuse reflection by relying on several prior assumptions such as light-from-above and upward-facing surface, also known as viewpoint-from-above. Experiment 2 investigates whether such priors also play a role in disambiguating the perception of qualitative shape from glossy and mirror reflections.

**Experiment 1**

**Method**

**Stimuli**

Each surface was determined by a mesh terrain composed of a 350 × 350 grid of vertices. Terrain heights were specified by band-pass filtered noise, from five to nine cycles per surface width. For each surface, a probe point was selected that was either on a convex or concave region (hill or valley). The subject’s task was to determine which type of region a given probe point lay on. The probe points were selected randomly by choosing from a set of points whose principal curvatures were above some threshold and of the same sign. The probe points were restricted to the middle 42% of the surface width and height so that each probe point was several hills and valleys away from the terrain boundary.

Each surface was rendered in perspective with the virtual viewer located at 53 cm from the center. Each surface was rotated back from frontoparallel so that it was upward facing with a slant of 30° (see Figures 1 through 3). We chose a 30° slant because it was large enough to produce perspective effects, but small enough that it does not produce occluding contours that would have provided strong cues to surface shape, as discussed in the Introduction.

The stimuli were displayed on a 24-in. Apple monitor at 1920 × 1200 resolution and a gamma of 2.2 and viewed in a darkened room. Subjects viewed the stimuli monocularly with an eye patch over the nondominant eye, and head motion was restrained using a chin rest. The viewing distance was 53 cm from the screen, which matched the rendering conditions. The resulting stimulus subtended a viewing angle of about 17° × 13°.

Several different surface reflectance and lighting models were used: texture only, mirror reflectance with two types of environment map, and a Phong (1975) model with four different combinations of diffuse and specular components.

For the texture condition, each surface mesh element was either white or brown and was illuminated by ambient light only. The texture was generated by randomly assigning each grid rectangle to be either white or brown with uniform probability. A sample stimulus is shown in Figure 1 as well as in a closeup in the top row of Figure 9.

Mirror surfaces were rendered using OpenGL spherical environment mapping. This assigns the red-green-blue (RGB) value of an image pixel to the RGB value of a pixel in an environment map image. The pixel coordinates of the environment map correspond to the directions on a sphere on which the environment

![Figure 1. Example of rendered texture surface. Here the surface has spatially varying reflectance and the lighting is assumed to be uniform ambient. See the top row of Figure 9 for a close-up.](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932811/ on 11/28/2018)
illumination is defined. To render each image pixel $p$, OpenGL simulates a light ray sourced at the pixel position and directed outward from the camera. It then reflects this ray off the mirror surface point that is visible at that pixel. This reflected ray, $R(p)$, defines the index into the environment map with its direction.

Two types of environment maps were used, which we refer to as homogenous and nonhomogeneous. The homogeneous environment maps were two-dimensional (2-D) image textures that were chosen to be approximately isotropic and homogenous. Each was radially deformed such that it mapped a coordinate system needed by OpenGL for spherical environment mapping. The nonhomogeneous environment maps were based on the homogeneous ones, but were modified to be relatively brighter above the horizon, similarly to natural environments. The nonhomogeneous environment maps were made by modulating the RGB values of the homogeneous environment maps as a function of the elevation in the environment. If the vertical elevation angle of $R(p)$ is $\phi(p)$ degrees from the zenith, then the color of the nonhomogeneous environment map $I(p)$ was determined from the color of a homogenous environment map $H(p)$ by

$$I(p) = H(p) \frac{0.3 + \max(0, \cos(\phi(p)))}{1.3}.$$  

This intensity modulation results in environment maps that are uniformly darkened in the lower hemisphere, and with brightness increasing in the upper hemisphere from horizon to zenith, similarly to naturally occurring environments (Sobolev, 1975). Sample renderings for the homogeneous and nonhomogeneous cases are shown in Figure 3a and b, respectively, and closeups are shown in Figures 9 and 10. The renderings using the nonhomogeneous environment map seem to give a more compelling sense of 3-D than the homogeneous one—possibly because they are visually similar to a light-from-above shading pattern.
The remaining rendering conditions used the Phong (1975) reflection model. A single directional white light source was set at an infinite distance above and behind the virtual viewer at an angle of 40° degrees above the horizontal plane. The surfaces were achromatic, so the color at every pixel was fully described by a scalar intensity corresponding to a sum of diffuse and specular components:

\[ I = k_d \max(0, \mathbf{L} \cdot \mathbf{N}) + k_s (\max(0, \mathbf{L}_R \cdot \mathbf{V}))^x \]  

(2)

where \( \mathbf{L} \) is the direction from a point on the surface to the distant white light source, \( \mathbf{N} \) is the surface normal at that point, \( \mathbf{L}_R \) is the direction that a ray from the light source would travel after being mirror reflected at the surface point, and \( \mathbf{V} \) is the direction from the surface point to the virtual viewpoint. The exponent \( x \) is a highlight width parameter which OpenGL allows to hold values from 0 to 128, where larger values of \( x \) correspond to narrower highlights. We used \( x = 51 \).

The parameters \( k_d \geq 0 \) and \( k_s \geq 0 \) are shown in Table 1. Each surface was rendered against a black background. We used four combinations of diffuse and specular components (see Figure 2): white diffuse, gray diffuse, gray diffuse + highlight, and black diffuse + highlight. The purpose of differentiating between the white versus gray conditions was to distinguish the effect of “adding” specular highlights to a diffuse reflectance versus using a different proportion of diffuse and specular highlights, a distinction we were concerned with following a preliminary study (Langer & Faisman, 2012). This is further discussed in the Results section.

For each rendering model, a surface was presented either in the static orientation shown in Figures 1 through 3 or else rotating. In the latter case, the axis of rotation was vertical but slanted back such that it was contained in the plane of the rendered surface and passed through the center of the surface. The rotation angle was sinusoidal with a period of 1.7 s and an amplitude of 10°. For the rotating surfaces, the probe point was fixed to the surface and so moved with it.

Subjects

Twenty-five subjects participated. Each was between 18 and 40 years of age and had normal or corrected-to-normal vision. Each subject was compensated $10. Results from all subjects are included.

Procedure

In each trial, a new randomly perturbed surface was computed, and a hill or a valley probe point was chosen such that it lay either at the top of a hill or at the bottom of a valley. For the initial 0.34 s of surface display time, a large red sphere subtending a visual angle of 3.2 degrees was rendered at that probe point. This sphere indicated the location of the probe point and allowed the subject to make an eye movement to that location. After 350 ms, this large sphere was replaced by a small sphere with a diameter of 0.2 visual angle, which remained attached to the surface. The red spheres were rendered using the ambient component but no diffuse or specular component (\( k_d = k_s = 0 \)). Thus the spheres appeared as a small flat frontoparallel disk attached to the surface. Sample test surfaces including the small red sphere are shown in Figures 1 through 3.

The subject’s task in each trial was to indicate whether the probe point appeared to lie on a hill or in a valley. Each surface was presented for 3.5 s total during which the subject had to press one of two buttons on the keyboard indicating hill or valley. The limited
display time discouraged subjects from trying to count hills and valleys from the side of the surface. In addition, we instructed subjects not to count in this manner. For the rotating surface condition, this time allowed for about two rotation cycles. If the subject did not respond by the end of the presentation interval, the screen turned red to indicate that the time limit was reached. Otherwise, the screen turned gray. The screen remained at that constant color for two seconds while the next mesh surface and probe point were computed. No other feedback was provided.

Before the experiments, subjects were briefed on the types of surfaces that they would be looking at. The surfaces were described as “largely flat perturbed by small hills and valleys” and several samples were shown to demonstrate the task and the types of surfaces involved. In addition, each subject was shown an example of a surface under the rotating homogenous mirror condition and was explicitly told that this was a curved mirror. No feedback was provided to the subjects about whether particular points were on hills or valley. Subjects were then given a set of practice trials, and they began the task when they indicated they were ready.

**Design**

The experiment was divided into blocks. Each block contained 28 trials: one hill and one valley for each of the 14 different conditions: two motion conditions (static and rotating) × seven rendering conditions (texture only, two mirror + environment map conditions, and four Phong conditions). These 28 trials were presented in random order within each block. If the subject did not respond within the 3.5 s time limit on one of the trials, a new trial with the same condition was presented again later in the block. Typical response times were 2 s, well within this time limit. Each subject ran nine blocks in a row, then was given an optional few minutes break, then ran eight more blocks. Each subject therefore had 476 trials = 17 blocks × 28 trials per block, or 34 trials per experimental condition. The resulting percent correct scores per condition were analyzed using a within-subjects ANOVA for N = 25 subjects.

**Results**

The results are shown in Figure 4. The first striking result is that subjects performed significantly worse on the gray diffuse + highlights condition than on the white diffuse condition for both the static and rotating cases, $F(1, 24) = 25, p < 0.0001; F(1, 24) = 24.7, p < 0.0001$, which confirms our preliminary finding (Langer & Faisman, 2012). This may be contrasted with previous studies that found that highlights enhance the perception of quantitative shape (Norman et al., 2004; Todd et al., 1997) or at least do not impede it (Nefs et al., 2006; Norman et al., 1995). The percent correct scores were also significantly lower for the gray diffuse + highlights condition than for the white diffuse + highlights condition for both the static and rotating cases, $F(1, 24) = 30, p < 0.0001; F(1, 24) = 26, p < 0.0001$. Moreover, there was no significant difference between the white diffuse and gray diffuse condition in either the static or rotating cases, $F(1, 24) = 0.06, p = 0.8; F(1, 24) = 0.16, p = 0.69$. 

Figure 4. Percent correct scores for Experiment 1.
To interpret the above results, we note that there are two steps involved in changing a typical diffuse rendering scheme into a glossy one: Firstly the diffuse component is reduced in magnitude, and secondly the remaining contrast is filled by a specular component. Since there are no differences between the white diffuse and gray diffuse percent correct scores, we can conclude that for our surfaces, the first of these two steps has no effect on the shape percepts relevant to the task. Therefore, the reduced percent correct scores for the gray diffuse + highlight surfaces are due strictly to the addition of highlights onto a diffuse surface. We return to this point in the Discussion section and propose a possible reason for the reduction in performance.

Subjects performed worst in the black diffuse + highlights condition. This condition provided barely enough shape information to allow for above chance performance in both the static and rotating conditions, $t$ test against 50%; $p \approx 0.03$ for both, with percent correct scores of 54.5% and 56.5%, respectively. This was largely unsurprising since rendering a surface using only highlights produces small isolated white regions, separated by large black regions that are the same intensity as the background. Interestingly, rotating the highlight surfaces also failed to significantly increase the percent correct scores, $F(1, 24) = 0.75, p \approx 0.4$. Since the red probe was fixed to the surface, the movement of the highlights relative to the surface created a parallax cue that could have been exploited to produce significantly higher performance on this condition (Blake & Bulthoff, 1991). It appears, however, as if this parallax cue was largely ignored by the subjects.

In the mirror conditions, subjects scored well above chance for both environment map conditions. Performance was higher in the nonhomogenous condition than in the homogenous condition for both the static and rotating cases, $F(1, 24) = 25, p < 0.0001; F(1, 24) = 29, p < 0.0001$. This implies that the inhomogeneities within a natural outdoor environment map carry shape cues that can aid in the perception of surface shape. However, it’s unclear whether all possible environment map inhomogeneities can be directly exploited by the visual system to increase performance or whether specifically ones with a brighter lit upper hemisphere are especially beneficial. The latter case appears likely since when the environment map is brighter in the upper hemisphere, locally convex regions tend to become brighter above the probe point, and concave regions tend to become brighter below the probe point, creating a visual cue that is very similar to diffuse shading with light from above (see Figures 3b and 10b). This is further discussed in the Discussion section.

Performance in the homogenous condition was still quite high (over 80%), which is somewhat surprising since images in this condition produce a much less vivid sense of 3-D shape. Thus, even without the possible benefit of a light from above cue in the environment map and nearby occluding contour cues, observers were able to judge qualitative shape relatively well. Note that while it has been suggested previously (Fleming et al., 2004) that occluding contours are not strictly necessary for shape perception on mirror surfaces; this had been demonstrated only informally and for renderings that used natural environment maps that are dominated by light from above. The homogenous condition in our experiment shows more definitively that neither occluding contours nor environment maps with light from above are necessary for qualitative shape perception.

Finally, an interesting general trend was that scores in the rotation condition were typically not significantly better than in the static condition, contrary to findings such as Norman et al. (1995) that motion improves shape perception. Table 2 shows the results of a within-subjects ANOVA comparing static and rotating for each rendering condition. The only rendering condition for which motion had a significant effect was for the texture. A possible reason for the advantage of motion in this condition is that the motion cue is most reliable when the luminance of each surface point is constant over time (Norman et al., 1995). Among our experimental conditions, this is only true for the texture case. In the diffuse and specular cases, the surface normals rotate relative to the illumination, which changes the intensities of each point over time.

We have established that different kinds of specular reflections can produce significant differences in qualitative shape perception, but questions still remain about why these differences occur. In particular, what information are subjects using to achieve better performance in some of the conditions than others? Some of the information to perform the task must have been contained in the pattern of image intensities in a neighborhood of the probe—for example, in the positions and shapes of the highlights or the deformations of the reflected environment in the case of mirrors. For specular reflections in general, these intensity patterns depend to some extent on the viewing position. Thus, some consideration of the relative position of the viewer with respect to the surface may play a role in explains performance. We return to these questions in the Discussion section. Before we do, we present a second experiment which addresses the question of how subjects were able to achieve such high performance in any of the conditions due to the presence of a fundamental perceptual ambiguity.
Experiment 2

Another general and important issue in qualitative shape perception is that a given image often can correspond to two very different but equally valid 3-D shapes. In particular, when a surface is reflected about any plane parallel to the view plane, and the illumination environment is reflected across the optical axis, the resulting rendered image remains unchanged. Strictly speaking, this “depth reversal” ambiguity holds only under orthographic projection and with directional light sources, but it holds approximately under perspective projection and with distant light sources as well. The depth reversal ambiguity has been known for centuries (Brewster, 1826; Rittenhouse, 1786) for the case of diffuse or ambient lighting, and it is the basis of the popular hollow mask illusion (Hill & Bruce, 1994). As we show in the Appendix, the depth reversal ambiguity applies to general specular reflections as well, including both highlights and mirrors.

Figure 5. Depth-reversal ambiguity. Schematic illustration of cross sections of an upward-facing Surface A (green) and its downward-facing reflection, Surface B (blue), along with their corresponding light source positions. Because of the depth-reversal ambiguity, the two surfaces appear identical when viewed through orthographic projection and rendered under the Phong model using their corresponding light sources. An analogous ambiguity exists for mirror reflections as well, with the environment map reflected about the optical axis instead of the light sources (see Appendix). A given probe point on an image therefore could either be interpreted as a concave point on an upward-facing surface or a convex point on a downward-facing surface.

For ambient lighting as well as for point source lighting with diffuse reflectance, the visual system resolves the depth-reversal ambiguity using prior assumptions about the surface structure and lighting that are consistent with typical natural viewing conditions. That is, when it is ambiguous which of two possible 3-D surfaces and lighting patterns is present, subjects favor certain prior assumptions, namely that the surface is illuminated from above (Rittenhouse, 1786), it is globally convex (Hill & Bruce, 1994; Liu & Todd, 2004), or is slanted upwards relative to the line of sight like a floor rather than downwards like a ceiling (Langer & Bülthoff, 2001; Mamassian & Landy, 1998; Reichel & Todd, 1990). While one would expect that these surface-shape priors are used to disambiguate the qualitative shape of glossy surfaces and mirrors as well, to our knowledge this has not been directly examined before. The goal of Experiment 2 is to investigate whether subjects were relying entirely on these priors to resolve the depth-reversal ambiguity for specular reflections in Experiment 1, or whether they were also relying to some extent on the perspective cues which were present.

To clarify further, the depth reversal ambiguity formally requires that surfaces are projected under orthographic rather than perspective projection (see Appendix). However, if subjects were to ignore the distortions that are due to perspective projection, then each of the surfaces used for Experiment 1 would be subject to the depth-reversal ambiguity. That is, each of

Figure 6. Percent correct scores for Experiment 2. Vertical black bars indicate means in each of the four rendering conditions.

Downloaded From: https://jov.arvojournals.org/pdfaccess.ashx?url=data/journals/jov/932811/ on 11/28/2018
the upward-facing surfaces could be incorrectly perceived as a downward-facing surface with all of the convex and concave sections inverted and lighting from below rather than from above the line of sight (see Figure 5). Since subjects performed well above chance in most conditions of Experiment 1, we know that they typically correctly perceived the qualitative shape of both the convex and concave portions, and therefore they must have correctly resolved the depth-reversal ambiguity in favor of upward-facing surfaces. However, there are two general ways in which the subjects they could have done so. They could have relied on a prior assumption that the surface was facing upwards (combined with a prior for light from above in the Phong conditions), or they could have used the perspective cues that indicated each surface was facing upwards. The most obvious perspective cue to this effect was the global shape of the surface on the image that appears roughly as a trapezoid with the bottom wider than the top because the bottom is closer to the viewer.

Our second experiment investigates the extent to which the perceived qualitative shape is affected by the prior for upward-facing surfaces versus the available perspective cues. The conditions include several of the Phong and mirror surfaces used in Experiment 1 that were facing upward with light source from above, and in addition we include surfaces that are facing downward and for which the light source is from below. If subjects relied only on perspective cues to resolve the depth-reversal ambiguity for a particular rendering condition, then they should perform similarly for upward- and downward-facing surfaces in Experiment 2, as they will resolve the depth-reversal ambiguity for both cases with similar competence. On the other hand, if they relied only on their priors for upward-facing surfaces and light from above, then they should perform worse than chance in the downward-facing condition. In previous studies with diffuse reflectance, strong perspective cues, and globally curved surfaces (bumpy cylinders), subjects ignored the perspective cues in resolving the depth-reversal ambiguity and instead relied entirely on priors, namely for light from above, upward-facing surface orientation, and global convexity (Langer & Bülthoff, 2001). In Experiment 2, we pit the priors for upward-facing surfaces (and light from above, when relevant) against perspective cues for surfaces that contain specular reflections.

### Method

#### Stimuli

The stimuli were similar to those of Experiment 1 except for the following changes. Firstly, only four rendering models were used: white diffuse, gray diffuse + highlights, black diffuse + highlights, and mirror reflectance with a homogeneous environment map. Secondly, each rendering condition had two versions, each rotated 30° degrees from frontoparallel, one to produce a surface that was facing upward identically to those used in Experiment 1 and the other symmetrically facing downward. When the downward-facing surfaces were rendered using the Phong model, the distant light source was placed symmetrically at 40° below instead of above the horizontal plane. This ensures that an upward-facing surface was identical to a corresponding depth-inverted downward-facing surface, except for the global perspective cue (see Figure 5 and corresponding discussion concerning the depth-reversal ambiguity). For the mirror condition, we did not need to invert the environment maps since we only used the homogeneous ones. To view sample downward-facing surfaces used for Experiment 2 one may simply rotate the corresponding figures from Experiment 1 by 180°.

#### Design

The experiment was divided into 10 blocks, each containing 32 trials: one hill and one valley for each of the 16 different rendering conditions, namely four rendering models (white diffuse, gray diffuse + highlights, black diffuse + highlights, homogenous mirrors) × two motion conditions (static and rotating) × two surface orientation conditions (upward and downward facing). The 32 trials within each block were presented in random order. Subjects were allowed a few minutes break after the first five blocks. Each subject therefore had a total of 320 trials = 10 blocks × 32 conditions per block, or 20 trials per experimental condition.

#### Subjects

This experiment was performed on 22 new subjects. Each was compensated $10. Each was between 18 and 45 years of age, with normal or corrected-to-normal vision. Results from all subjects are included.

### Table 2. Test of differences for rotating versus static condition.

<table>
<thead>
<tr>
<th>Rendering model</th>
<th>F(1, 24)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture</td>
<td>31.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>White diffuse</td>
<td>2.09</td>
<td>0.16</td>
</tr>
<tr>
<td>Gray diffuse</td>
<td>3.47</td>
<td>0.07</td>
</tr>
<tr>
<td>Gray diffuse + highlights</td>
<td>2.86</td>
<td>0.10</td>
</tr>
<tr>
<td>Black diffuse + highlights</td>
<td>0.75</td>
<td>0.40</td>
</tr>
<tr>
<td>Homogenous environment</td>
<td>0.24</td>
<td>0.63</td>
</tr>
<tr>
<td>Nonhomogeneous environment</td>
<td>0.01</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Downloaded From: https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932811/ on 11/28/2018
**Procedure**

The procedure was the same as Experiment 1, with the one alteration that the stimuli were described to the subjects in a slightly different manner than in the first experiment. Since all of the surfaces were facing upward in Experiment 1, it was possible to use the terms “hill” and “valley” consistently with the upward facing terrain. Since these terms might have been misleading for a downward facing terrain, we instead use the terms “bump” and “dent” to describe convex and concave probe locations, respectively.

**Results**

Percent correct scores are plotted in Figure 6. For the white diffuse, gray diffuse + highlight, and mirror conditions, scores were far greater in the upward-facing than downward-facing conditions. In the case of the diffuse condition, this is consistent with the well-known prior for upward-facing surfaces (also known as “viewpoint from above” prior, i.e., the assumption that surfaces are typically floors rather than ceilings). This prior clearly applies to mirror and glossy surfaces as well since performance was significantly higher in the upward-facing than downward-facing surface orientation conditions for these rendering conditions.

To what extent do subjects use perspective cues in addition to this prior to resolve the depth-reversal ambiguity? If subjects relied entirely on their prior assumption that surfaces are upward-facing, then they should have the same scores as in Experiment 1 for the upward-facing surfaces, and the scores for the downward-facing surfaces should be symmetrically below chance, with overall scores in rendering condition being at chance. If, however, subjects use the available perspective cues to some extent to disambiguate global surface orientation, then this should raise performance in the downward-facing surface condition, so the average of the upward and downward conditions should be above chance.

Following this line of thought, we consider the overall means for each of the four rendering conditions, which are shown by vertical black bars in Figure 6. Although these overall scores were above chance in each case, it was not significantly so for the white diffuse, the gray diffuse + highlights, and the black diffuse + highlights conditions (t-test $p > 0.05$ for all three conditions, against 50% baseline). We conclude that for these conditions, subjects did not use perspective cues to resolve the depth-reversal ambiguity but rather relied on the prior for upward-facing surfaces. This is consistent with previous findings for diffuse reflectance (Langer & Bülthoff, 2001).

For the mirror condition the upward-facing surface gave a much higher score than the downward-facing surface, implying that the prior for upward-facing surfaces was also used. A more surprising observation is that the overall percent correct score for the mirror surfaces was significantly above chance ($p < 0.0001$ using a t test). This implies that subjects were using perspective cues in addition to the upward-facing surface prior. It is unclear why subjects relied on the perspective cue more strongly for the mirror condition. One possibility is that because this surface type is not commonly encountered, the strength of the corresponding prior is weaker, and therefore perspective cues are more strongly taken into account. Another possibility is that in contrast with the other rendering conditions, there was no light-from-above cue in the homogeneous mirror condition, decreasing the overall strength of the priors, which may have led subjects to rely more on the available perspective information.

**Discussion**

One of the surprising findings of Experiment 1 was that adding highlights to diffuse reflectance impeded the perception of qualitative shape, which was inconsistent with previous studies (see Introduction). There are several differences between our study and previous studies that may account for this inconsistency. One difference is the task—namely previous studies examined the ability to discriminate quantitative shape for surfaces that had similar qualitative shapes (Norman et
al., 2004), whereas our study required subjects to discriminate quite different qualitative shapes (hill vs. valley). Since highlights are relatively sensitive to small changes in the surface normal, it is possible that highlights are more helpful for the former task than the latter. A second difference is that in several of the previous studies, different colors were used for the diffuse and highlight regions—for example, blue or yellow diffuse surface reflectance and white highlights (Nefs et al., 2006; Norman et al., 2004; Todd et al., 1997). Using different colors may have allowed subjects to perceptually decompose the image into diffuse versus highlight layers, which could have allowed subjects to more easily perceive shape from the diffuse component.

We now examine the differences in perceptual cues that may have caused subjects to perform better in the diffuse than in the diffuse + highlight conditions. Consider the case that the surface is upward facing with light from above. Recall that the light source was 40° above the line of sight and the surface was slanted back by 30°, which results in a light source direction that is 10° above the overall surface normal. For a local hill, this produces a typical peak intensity of the diffuse shading component above the probe point on an oblique part of the surface, whereas the highlights typically peak below the probe point on a more frontally oriented portion of the surface. The opposite trend is typical for valleys (see Figure 7). Since the peak of the diffuse component tends to cover the oblique portion of the surface, the diffuse shading distribution is typically relatively compressed around the contour of the hill or valley due to foreshortening, more so than if it covered a frontoparallel portion of the surface. This foreshortening effect presumably aids in the perception of qualitative shape. By contrast, since highlights tend to occur in the more frontoparallel surface regions, their shape is much less affected by foreshortening, and therefore, their intensity distribution tends to be relatively symmetric about the peak. The addition of highlights onto a surface illuminated with diffuse shading tends to displace the lit portion of the surface from the oblique to the frontoparallel section. This results in a diffuse + highlights shading distribution that contains weaker foreshortening cues than a diffuse shading distribution, a factor that is arguably responsible for the subjects’ lower performance in this condition.

Figure 8 provides a rough empirical evaluation of the shading distributions described above. This figure shows the mean intensity of 300 randomly generated hills and valleys for each of the three relevant rendering conditions. The images have been contrast enhanced to

![Figure 8. Average of 300 cropped images of hills and valleys for upward-facing surfaces with (a) white diffuse, (b) highlight only, (c) gray diffuse + highlight. The red square indicates the location of the probe point and the green square indicates the maximum intensity of the image. Note that this maximum intensity is sometimes off-center because of the finite number of samples. The images have been linearly contrast enhanced to contain intensities ranging from 0 to 1.](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932811/ on 11/28/2018)
better visualize the differences. We find as predicted that for hills, the peak intensities tend to lie above the probe point for diffuse surfaces and below the probe point for highlights, with the opposite trend for valleys.

While the shapes of individual shading distributions cannot be as clearly discerned from these averaged images, it is probable that the foreshortening effects of the oblique portion are responsible for the somewhat horizontally elongated intensity peak of the diffuse condition. By contrast, the highlight and diffuse + highlight peak distributions are relatively circular as they typically occupy a more frontoparallel region on the surface that is not subject to foreshortening effects.

What are the cues that subjects use for the texture and mirror renderings? Texture compression is a well-known cue for surface slant and tilt (Cutting & Millard, 1984; Gibson, 1950; Stevens, 1983), and more recent studies have generalized this idea to the texture compression that arises from specular reflections of the environment (Fleming et al., 2004) and more generally to spatial gradients in image orientations or “flow” (Fleming, Holtmann-Rice, & Bülthoff, 2011; Huggins & Zucker, 2001; Koenderink & van Doorn, 1980). Here
we briefly review some of these ideas and how they apply specifically to our stimuli.

When textured surfaces face upwards, image points that are just above hilltops have a larger vertical slant than points that are just below hilltops, so there is a greater vertical image texture compression above the hilltops than below them. The opposite pattern occurs for valleys (see first row of Figure 9). As can be seen in the other rows of that figure, similar compressions exist for the mirror conditions, although for different reasons as mirror compression depends on change in surface normal rather than the local surface slant. The reason for the compression in the mirror case is that, although both surface regions may have similar intrinsic surface curvatures, the region above the hilltop has a greater overall slant with respect to the viewing direction and so a given image step in the vertical direction above a hill typically spans a larger set of surface normals than does the same step below the hill (Fleming et al., 2004). A similar argument explains why, in valleys, the compression is below the valley bottom. We emphasize that knowing (or assuming) an upward-facing surface is critical for the correct interpretation of these compression patterns. Without such a constraint, the depth-reversal ambiguity would not allow one to distinguish hills from valleys.

We note additionally that although subjects use a prior for viewpoint from above to perform the local shape task, this does not imply that the subjects must have perceived downward-facing surfaces as globally upward slanting. The situation here is similar to examples such as the Escher staircase and the Penrose triangle, where global information about surface orientation conflicts with local information. In those examples, one has no trouble making local perceptual judgments, even though there is no possible consistent global percept. Similarly, when one views the images in Figures 1 and 3 upside down so that the surfaces are downward facing, the global orientation becomes perceptually very uncertain. This uncertainty is due to the conflict between the prior for viewpoint from above in the interior of the image versus the perspective cues, which are particularly apparent on the periphery. Whether a subject would rely more on the prior or on the perspective cues to perform the global orientation task cannot be predicted from the data for the local shape task.

There are additional questions concerning the perception of curved mirror surfaces that remain open for further research. The results of Experiment 1 indicated that the inhomogeneity of an environment map can strongly influence the resulting surface percept. Specifically, using an environment map that was brighter above the horizon aided the subjects in perceiving 3-D surface shape. But was it easier to judge overall surface shape in this case simply because the environment map was not homogenous, or was it because the inhomogeneity produces a shading distribution which is similar to a standard light-from-above shading cue (see Figure 3b)? In the latter case, performance might be lower if an environment map were used that was bright in the left or right rather than upper hemisphere. Further experiments are needed to address this point and to clarify similarities and differences in diffuse shading and specular reflections cues and how they are processed by the human visual system.

**Conclusion**

We have investigated how well subjects judge local qualitative shape from surfaces rendered using a variety of rendering models. Three main novel results were obtained. Firstly, we found that the addition of highlights to diffuse shading worsens qualitative shape judgments. This is contrary to previous findings in the literature that showed adding highlights to diffuse shading either improves shape perception or has no effect. A possible reason for this difference is that our task involves large differences in qualitative shape, whereas the previous tasks involved fine shape discriminations. Additionally, in previous studies highlights were rendered using a different color for the diffuse shading component, which reduced the obfuscation of the diffuse shading pattern by the highlight.

Secondly, we found that qualitative shape can be perceived quite well from mirror reflections alone even in a complex terrain environment. This result was somewhat surprising since, when homogeneous environment maps are used, the rendered mirror surfaces do not give a compelling sense of the layout of hills and valleys of the terrain. Although it had been shown before that shape can be perceived well from mirror reflections (Fleming et al., 2004), the corresponding stimuli had contained smooth occluding contours and also used real environment maps for the illumination, in which the upper hemisphere was brighter. Our experiments showed that complex curved mirrors without these additional cues also allow excellent judgments of qualitative shape. In addition, we showed in Experiment 1 that using an inhomogeneous environment map that is brighter above the horizon can improve shape percepts.

Thirdly, we observed that the classical depth-reversal ambiguity generalizes to specular reflections and that the visual system resolves this ambiguity using a prior for upward-facing surfaces for both glossy and mirror surfaces. Interestingly, whereas this prior completely dominates over perspective cues for the glossy render-
ing conditions, the prior did not completely dominate over perspective for the homogeneous mirror case. Further experiments are needed to explore why this is the case.

**Keywords:** shape, shading, highlights, mirrors, specular reflections, natural scenes, environment map

**Acknowledgments**

This research was supported by a Discovery Grant to MSL from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Commercial relationships: none.

Corresponding author: Michael S. Langer.

Email: langer@cim.mcgill.ca.

Address: School of Computer Science, McGill University, Montreal, QC, Canada.

**References**


of the National Academy of Sciences, USA, 110(6), 2413–2418.


### Appendix

Assume a surface rendered using orthographic projection such that the surface intensity at each point is determined either by a mirror environment map or by the Phong model as in Equation 2 with a distant light source. Furthermore, assume that every point on the surface is visible. In particular, there are no occluding contours. We show that in these rendering cases, an inversion of the rendered surface with respect to a plane $P$ parallel to the image plane followed by a simple reflection of the lighting over the optical axis produces an image identical to the original. Note that reflected lighting refers to either the distant Phong point light source or the mirror environment map. Additionally, note that the inversion of a slanted surface with respect to the plane $P$ leads to changing the upward/downward surface orientations used in Experiment 2 above, as well as changing each hill to a valley and vice-versa.

Consider such a surface rendered using orthographic projection with the $z$ direction pointing toward the viewer and the $z$-axis aligned such that the plane $z = 0$ corresponds to the inversion plane $P$. Let the surface depth as a function of the image plane as well as inversion plane coordinates be $f(x, y) = z$. Note that since by assumption every point on the surface is visible, the coordinates $(x, y)$ uniquely reference the same point, both before and after the inversion. The normal $\vec{N}$ to the surface at the point $(x, y, f(x, y))$ is

$$\vec{N}(x, y) = \frac{-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}} = C(x, y) \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\right)$$  \quad \text{(3)}$$

where $C(x, y) = (\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1)^{-1/2}$ and all of the partial derivatives are functions of $(x, y)$. Since we are considering orthographic projection, all of the light rays arriving at the camera are oriented in the direction of the viewer $\vec{V} = (0, 0, 1)$. Therefore, a mirror reflection $\vec{R}$ of a camera ray off the specular surface has direction:

$$\vec{R}(x, y) = 2 \left(\vec{V} \cdot \vec{N}(x, y)\right) \vec{N}(x, y) - \vec{V}$$  \quad \text{(4)}$$

$$= C(x, y)^2 \left(-2 \frac{\partial f}{\partial x} - 2 \frac{\partial f}{\partial y}, 1 - \left(\frac{\partial f}{\partial x}\right)^2 - \left(\frac{\partial f}{\partial y}\right)^2\right)$$  \quad \text{(5)}$$

A reflection of the surface in depth is equivalent to $f(x, y) \rightarrow -f(x, y)$ at the visible surface point corresponding to pixel $(x, y)$, which implies that $\frac{\partial f}{\partial x}(x, y) \rightarrow -\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y) \rightarrow -\frac{\partial f}{\partial y}(x, y)$. Since by
assumption every point on the surface is visible, an inversion in depth will not bring forward a previously occluded pixel, and these equations will hold uniquely for each pixel. Therefore, we can see directly from the equations above that both the surface normal vector $N(x, y) = [N_1(x, y), N_2(x, y), N_3(x, y)]$ and reflection vector $R(x, y) = [R_1(x, y), R_2(x, y), R_3(x, y)]$ change sign in the first two coordinates following a reflection in depth—i.e., become $-N(x, y)$ and $R(x, y)$, respectively.

Case 1: Mirror environment map. Since the mirror environment map determines a one-to-one correspondence between the reflected vector $R(x, y)$ and image intensity at the pixel $(x, y)$, we have shown that if we invert the surface in depth and reflect the environment map along the $x$- and $y$-axes, then the image will not change.

Case 2: Phong model. Consider Equation 2 for the surface corresponding to some particular pixel point $(x, y)$ and consider the changes incurred by the surface inversion. Since the light source is by assumption directional, we have that the light direction $\vec{L}$ does not depend on $(x, y)$, so the term $\max(0, \vec{L} \cdot \vec{N})$ changes into $\max\{0, \vec{L} \cdot [-N_1(x, y), -N_2(x, y), N_3(x, y)]\}$. Therefore, the diffuse component of the Phong model does not change if the surface is inverted and the first two components of the light source are negated—i.e., if the light source is reflected about the $x$- and $y$-axes. Similarly, consider the term $\max\{0, (\vec{L} \cdot \vec{V})^2\}$ corresponding to the specular component. Since $\vec{L}$ and $\vec{V}$ are the reflections of $\vec{R}$ and $\vec{L}$ at the surface with normal $\vec{N}$, the angle between them does not change after reflection, and therefore, $\vec{L}_R \cdot \vec{V} = \vec{R} \cdot \vec{L}$. But

$$\vec{R} \cdot \vec{L} = (-R_1(x, y), -R_2(x, y), R_3(x, y)) \cdot (-L_1(x, y), -L_2(x, y), L_3(x, y))$$  \hspace{1cm} (6)$$

Therefore, inverting the surface and negating the first two components of the light source does not affect the given image. Since there is no change in both the diffuse and specular components of Equation 2 under the surface inversion and a reflection of the light source along the $x$- and $y$-axes, we conclude that the depth-reversal ambiguity holds under specular reflection, similarly to the well-known ambiguity under diffuse reflection.