Title
The perception of shape, lighting, and material properties in images

Author
O'Shea, James Patrick

Publication Date
2012

Peer reviewed|Thesis/dissertation

The perception of shape, lighting, and material properties in images

by

James Patrick O'Shea

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy
in

Vision Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Martin S. Banks, Chair
Professor Maneesh Agrawala
Professor James O'Brien

Fall 2012

The perception of shape, lighting, and material properties in images

Abstract

The perception of shape, lighting, and material properties in images

by

James Patrick O'Shea

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor Martin S. Banks, Chair

Three scene properties determine the image of a 3D object: the material reflectance, the illuminantion, and the object's shape. Because all three properties determine the image, one cannot solve for any one property without knowing the other two. Nevertheless, people often are able to perceive these properties consistently and relatively accurately. We explore the relationship between these scene properties, the sources of image information the visual system can use to recover these properties, and the assumptions the visual system tends to make. We first conducted a shape perception experiment in which we investigate whether the visual system assumes the angle between the lighting direction and the viewing direction. Observer errors were minimized when the light was 20-30° above the viewing direction, confirming the light-from-above prior. In a second study, we conducted two psychophysical experiments to determine how viewers use shape information to estimate the lighting direction from shaded images. We found that observers can accurately determine lighting direction when a host of shape cues specify the objects. When shading is the only cue, observers always set lighting direction to be from above. We modeled the results in a Bayesian framework that included a prior distribution describing the assumed lighting direction. Finally we explore how disparity and defocus information may be useful in material perception to distinguish glossy and matte surfaces. We describe the types of images needed to investigate this question, and introduce a method to render them.

List of Figures

Acknowledgements

The authors gratefully acknowledge support from NIH, NSF, Microsoft Research, The Okawa Foundation, and the Alfred P. Sloan Foundation. This work was supported by NIH Research Grant R01- EY012851 and NSF Research Grant BCS-0617701. We thank Hany Farid for discussions on solving for the lighting direction in an image, James OBrien for discussions on the global convexity analysis, Yaniv Morgenstern for discussing his work on lighting cues and spherical statistics, and David Hoffman for help with modeling. Some of the data were previously presented at the Vision Sciences Society Annual Meeting between 2008 and 2010.

Chapter 1

Introduction

An image is determined by three primary scene properties: 1) the 3D geometry of the objects in the image, the materials of which they are composed, and the illumination. One of the primary functions of the visual system is to perceive these different properties in order to interact with the our environment. For example, we rely on vision to identify objects, estimate their positions in our environment, identify what they are made of, etc.

Because all three properties interact to determine the image content, it is impossible to recover any one particular property without knowing the other two. For example, one cannot recover the 3D orientation of a surface without knowing its material properties and how it reflects light, as well as the lighting in the scene. Yet people typically have relatively accurate estimates of these properties in every day life. They must be using available sources of information to estimate these properties or making reliable assumptions about the properties. We describe three projects investigating how these properties form images, how we use different sources of information to estimate these properties, and the assumptions we make about them.

Shading is a fundamental shape cue and observers often have consistent 3D percepts when viewing images of shaded objects. Recovering 3D shape from shading is impossible without additional information about the scene. In the first project, we investigate the assumptions observers make about the illuminant when viewing shading images. We describe a shape perception experiment to identify how assumptions about the lighting direction affect estimates of 3D surface orientation.

Lighting is another principle property of image content, and many of our inferences about shape and material are based on accurately knowing the illumination of a scene. In the second project, we explore how different types of shape information affect an observers ability to estimate the lighting direction in a scene. We describe a series of perceptual studies in which we directed participants to match the perceived lighting direction in an image while we manipulated the reliability of the 3D shape informationn. Using a Bayesian framework, we model how observers make use of

the available shape information and their prior expectations when judging lighting direction.

The third project explores material perception and the different sources of information available to the visual system. Gloss is one of the primary characteristics of material reflectance, and it is often trivial to distinguish glossy and matte surfaces. Yet, as with 3D shape and illumination, solving for the material properties is ill-posed without additional information about the scene. We discuss how disparity and defocus provide geometrically identical information about the specular property of materials. We also describe a technique to independently manipulate these cues in an image without affecting other characteristics that may play a role in gloss perception.

Chapter 2

Shape

2.1 Introduction

Shading is a fundamental pictorial depth cue. In computer graphics, shading is typically employed to convey 3D shape on a flat display screen. This type of visualization is completely dependent on the visual system's ability to perceive shape from a shaded 2D image. In order to convey 3D information effectively, it is therefore important to understand how shading influences shape perception.

Shading conveys depth by varying the levels of brightness in an image to indicate the extent to which a surface is oriented towards a light source. In computer graphics, there are many methods to produce this effect with varying° of realism and complexity [60, 52, 9, 62, 48]. Consider the image of a surface with an albedo ρ. In a simple diffuse shading model, the intensity I of any particular point in the image is proportional to the cosine of the angle between the surface normal $\hat{n}$ and the vector pointing to the light $\hat{l}$ as follows:

$$I = \rho\big(\mathsf{max}(\hat{n} \cdot \hat{l}, 0)\big) \tag{2.1}$$

Computing the 3D shape from this image requires solving for the surface normals $\hat{n}$. There is no unique local solution for this formulation of the shape-from-shading problem because we have one equation with more than one unknown variable [47, 54]. Additionally, Belhumeur and colleagues showed that the image of a shaded Lambertian surface is consistent with a continuous set of related surfaces [4]. Due to this bas-relief ambiguity, the 3D shape of a Lambertian surface cannot be uniquely solved with a single viewpoint and an unknown light source.

One way the visual system can help resolve these ambiguities is therefore to assume the direction of light. We define the light direction according to a coordinate system in which the eye is positioned along the positive z-axis, the x-axis is horizontal (positive to the right), and the y-axis is vertical (positive up). We can therefore describe the direction of the light as a vector in 3-space (Figure 2.2). Assuming a light source at

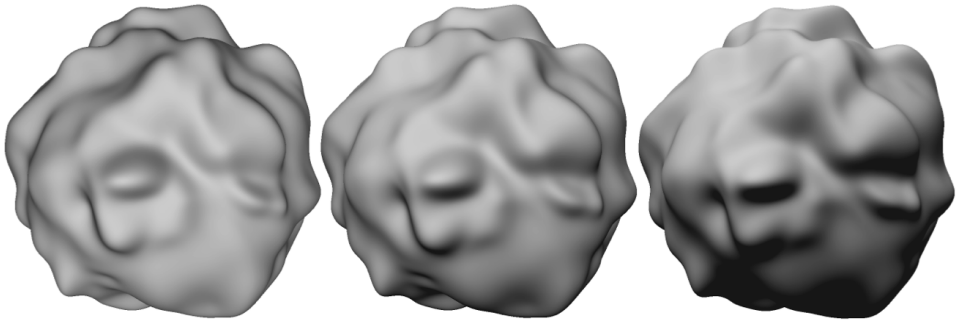

Figure 2.1. *The shading of an objects surface depends on the light source direction. The light is directed from the viewpoint in the left image, from 22° above the viewpoint in the middle image, and from 44° above the viewpoint in the right image. The object is positioned identically in each of the three views. In this chapter, we present an experiment designed to test how shape perception is affected by changing the angle of the light direction. We found the lighting used in the center image led to the most accurate estimations of 3D shape.*

infinity, we can ignore the magnitude of this vector and specify the light direction using two parameters in a manner analogous to surface slant and tilt [58]. Surface slant is the angle between the viewing direction and the surface normal. Surface tilt is the angular difference between the positive x-axis and the surface normal after it is projected onto the xy-plane. We can similarly define a light direction according to its slant and tilt relative to the viewing direction. Lighting slant, in this case, is the angle between the viewing direction (z-axis) and the light direction vector.

Previous research has shown that the visual system assumes light is coming from above and slightly to the left of a shaded object [59, 41]. Several studies have investigated how lighting and shading affect shape perception [41, 42, 32, 33], yet none have specifically examined variations in the slant of light.

We conducted a study to determine if changes in the slant of light affect 3D shape perception when shading is the only information available. We instructed observers to estimate the local surface orientation at specified sites on a virtual 3D object by orienting a gauge figure to appear normal to the surface [31]. By varying the light slant throughout the experiment, we were able to analyze subject settings as a function of illumination direction to determine the direction at which performance was optimal. If the visual system assumes a particular angle between the light direction and the view direction, we expect subject settings to be most accurate when the actual lighting is consistent with this assumption.

We rendered the surfaces using a local Lambertian shading model. We did not render specular reflections or cast shadows. We used this model because of its simplicity and its common usage in computer graphics applications. Due to its lack of photorealism however, the results of our study do not necessarily apply to viewing objects in the real world. Our findings do have a direct application in computer

graphics. Knowledge of the assumed light position is relevant to recent work on automatic lighting design [57, 17, 35], as well as non-photorealistic rendering techniques designed to affect the perception of 3D shape [10, 55]. If such a system positions the light away from the assumed direction, a viewer may be less likely to accurately perceive the shape of the object.

The main contributions of this chapter include the following:

- We measure shape perception as a function of the slant of light in order to directly evaluate whether the visual system assumes a particular slant angle when presented with local Lambertian shading without cast shadows. We find that observers assume the light is slanted 20-30° above the viewpoint.
- We isolate shading from other cues to shape, such as silhouettes and familiarity, in order to examine how shading alone affects the perception of surface orientation.

2.2 Related Work

To recover 3D shape from the image of a shaded surface, the visual system often makes use of information other than shading. Mamassian and Kersten demonstrated that the occluding contour of an object can be a significant cue to surface orientation [42]. In their experiment, subjects were instructed to orient a gauge figure to estimate local surface orientation for an object that is shaded under various lighting conditions. This work provides evidence that shading does affect surface perception, but it is important to note that their observer responses were not significantly different when the shading information was completely omitted and only the silhouette remained. The authors conclude that observers use the occluding contour of the object rather than shading to estimate surface orientation.

This finding may be due to Mamassian and Kersten's use of a simple object for probing local surface perception. Its surface was relatively uniform, and the silhouette provided adequate information for estimating surface orientation. By including the silhouette, this study failed to investigate the effect of shading alone. Given the limited number of lighting conditions, their experiment was also not specifically designed to examine how varying the angle between the view vector and the light direction would affect surface perception.

The familiarity of a shape may also provide additional information for estimating local surface orientation. Koenderink et al. [32] investigated how shading affects surface perception using a gauge-figure technique, yet their use of the human form as a stimulus image may have confounded the results. In the same way that the occluding contour of an object provides additional shape information beyond the shading, an observer's familiarity with an object may influence the results in this

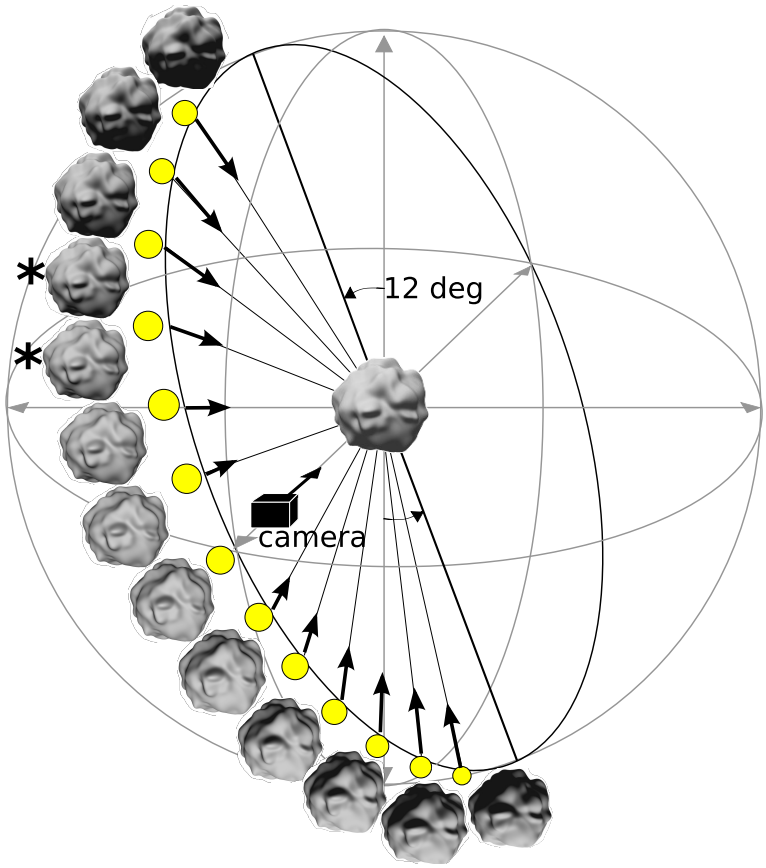

Figure 2.2. *Lighting directions tested in the experiment. We varied the angle between the light source and the view vector within a range +/- 66° above and below the viewpoint in 11° increments. Previous research has shown that people assume the light is above and to the left an average of 12° from vertical. Thus, we constrain our light directions to a plane that has been rotated about the z-axis 12° counter-clockwise. The images to the left of the light source directions illustrate how an object would be shaded according to these light directions. The asterisks denote the light directions which led to the most accurate estimations of 3D shape in our study.*

type of experiment. Because Koenderink's study was not specifically designed to address this question, it is unclear how it may have affected the outcome.

There is also research suggesting that the global convexity of a shape may influence the perception of local surface orientation for shaded objects. Langer and Bülthoff [33] showed that performance in a surface discrimination task was better for surfaces of globally convex objects compared to surfaces from globally concave objects. Their surfaces were rendered using only three different lighting directions.

The light-from-above prior is an important assumption a viewer can make to disambiguate the shape of a shaded surface. Sun and Perona experimentally investigated this prior in a study in which subjects were asked to identify if shaded ellipses appeared convex or concave [59]. By rotating the shading gradient of elliptical discs

about the visual axis, they effectively rotated the location of the light source as well. The results were consistent with an above-left prior. The study was not designed to specifically probe the angle between the light direction and the viewing direction.

This light-from-above prior was confirmed with line-drawing stimuli meant to represent the embossed surface of a fronto-parallel object [41]. In these stimuli, dark lines indicated a part of the surface facing down and white lines indicated the highlights of a surface facing up. Given the simplicity of these stimuli, the lighting could only be perceived to be in one of two positions: above or below the pattern of lines. The authors of this study did not specifically address whether task performance was affected by the slant of light.

More recently, Caniard and Fleming [6] addressed how shape estimation can be affected by changes in illumination conditions. They instructed subjects to perform both a surface-matching task, as well as local surface estimation task, while varying the location of the light source. Their results show an effect of light position, but they did not specifically vary the slant of the light in a systematic way.

These studies examined how lighting and shading affect shape perception. They did not specifically vary the angle between the light direction and the view direction, or they did not eliminate other pictorial cues which an observer could use to disambiguate the shape of a shaded surface. The experiment we describe in this chapter addresses both of these concerns.

2.3 Methods

We used a gauge-figure task [31] to test surface perception while varying the angle between the light direction and the viewing direction. We presented images of shaded irregular surfaces on a computer display and subjects were instructed to orient a gauge figure to appear normal to the surface of the displayed object.

2.3.1 Subjects

We collected data from three subjects familiar with psychophysical tasks. None had seen the stimuli before or were aware of the specific hypothesis under investigation. All subjects had normal visual acuity. If they required an optical correction, they wore it during testing.

2.3.2 Setup

The experiment was conducted using a desktop computer running Windows 2000 on an Intel processor. The 19-inch CRT display was gamma-corrected to linearize the luminance function for each color channel. We tested each subject monocularly after

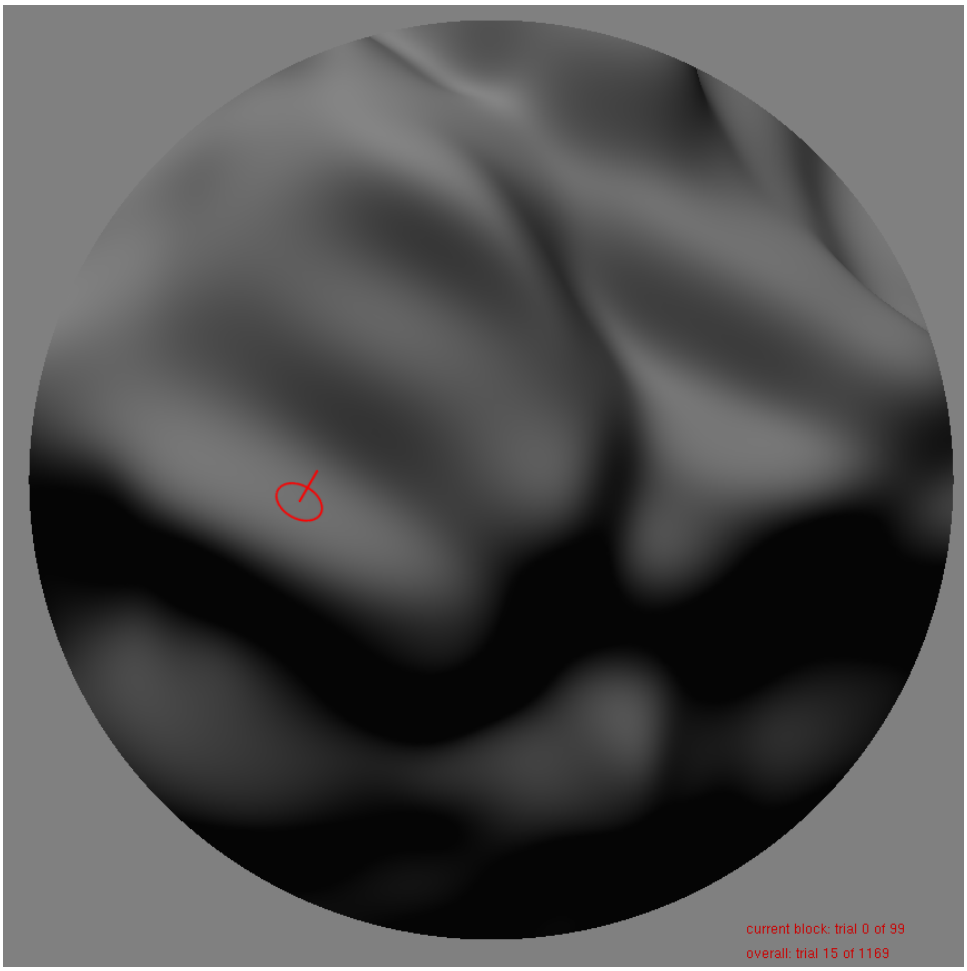

Figure 2.3. *Example screenshot from the experiment. In this figure, we show a synthetic aperture to approximately demonstrate how the stimuli appeared to the subject. In the experiment, we positioned a physical aperture between the viewer and the display screen to guarantee that the subjects would not mistake the edges of the aperture for the circular silhouette of the object.*

positioning them 45cm from the screen with their viewing eye directly in front of the center of the display. The subject's head position was maintained throughout the experiment using a custom-fitted bite bar. The experiment was self-paced. Subjects responded using a mouse. The CRT was viewed through a physical aperture to prevent the observer from seeing and thereby making use of the occluding contour to judge surface orientation. The experiment was conducted in a dark environment.

2.3.3 Stimuli

We generated smooth irregular 3D models for rendering our stimuli. We created each shape using a triangular mesh model of a sphere containing over 300,000 triangles. To create variations in the surface, we randomly selected vertices from the model

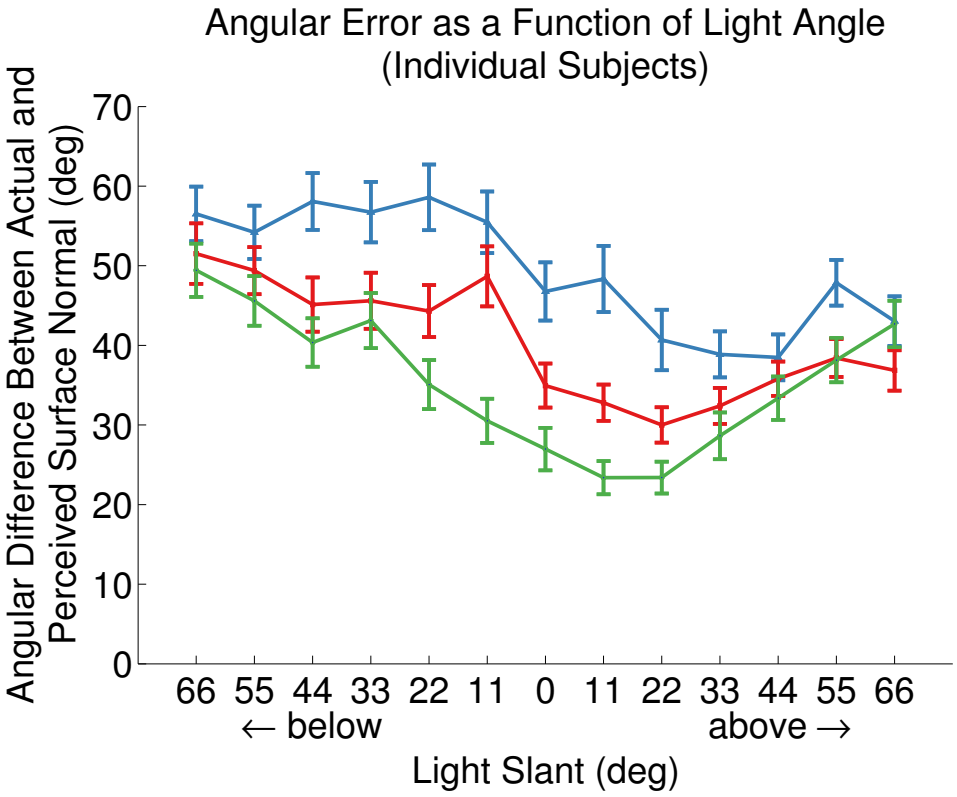

Figure 2.4. *Individual measurements of the angular difference (deg) between the actual and perceived surface normal as a function of light angle (deg). Each line represents data from one subject. The error bars represent the standard error of the mean. Although there are slight variations between subjects, each individual performed optimally when the light was slanted between 11° and 44° above the viewing direction.*

and increased or decreased their distances from the sphere's origin. We modified the locations of the neighboring vertices in a similar manner to keep the surface smooth. We shaded each object using a local Lambertian shading model without specularities or cast shadows. See Figure 2.1 for an example. See the Lighting section below for more information on the lighting conditions.

2.3.4 Gauge figure

We instructed subjects to indicate the surface normal at specified points on the surface using a gauge figure [31]. The gauge figure consisted of a single line with a circle at its base. The base appeared fixed to the surface (Figure 2.3). We presented the figure using perspective projection to provide additional information about its orientation. The circle was included as an additional cue to orientation. The line and circle were rendered without shading to avoid interfering with the shading of the surface. The orientation of the gauge figure was set randomly at the beginning of each trial. We rendered the gauge figure at a constant size so changes in its size would not reveal

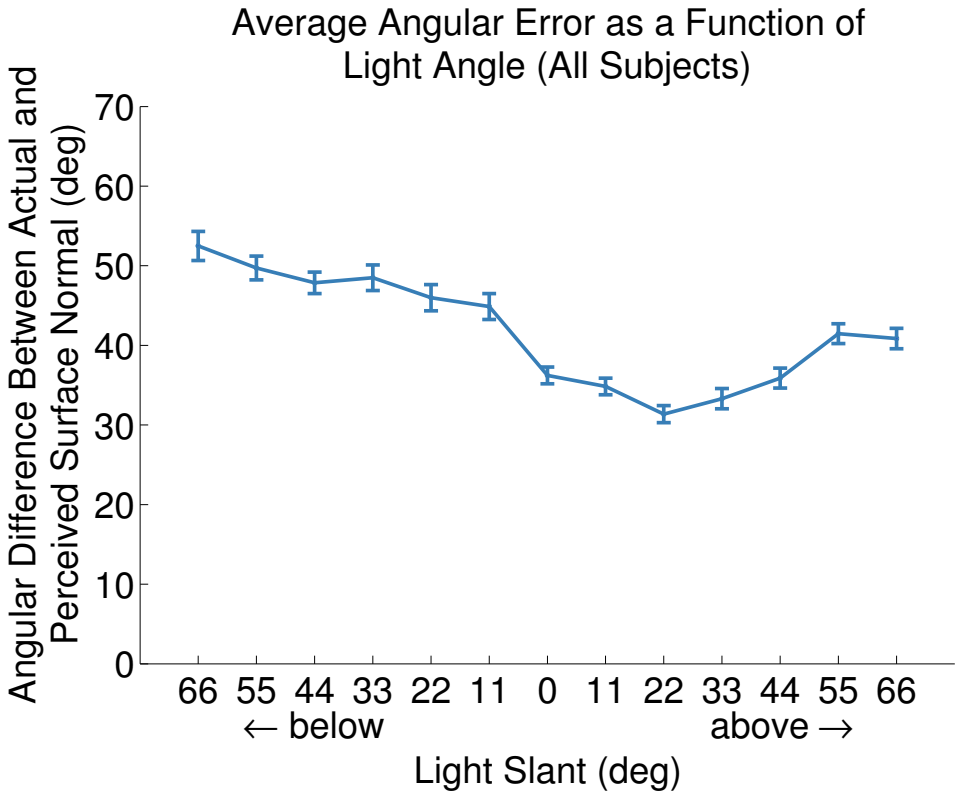

Figure 2.5. *Average angular difference (deg) between the indicated and actual surface normal as a function of light angle (deg). The data depicted in this chart are the combined average for all three subjects after normalizing the data. The error bars represent the standard error of the mean. A light angle of 0° corresponds to an illumination direction from the viewpoint. Errors are highest when the light is positioned below the object. Subjects performed best when the light was slanted 22° above the viewing direction.*

information about variations in depth along the surface. The task was conducted with monocular viewing to avoid a binocular cue to flatness that would have otherwise occurred.

2.3.5 Lighting

For each trial, we illuminated the object using a single directional light source. Because our goal was to study how the light slant (angle between light direction and the viewing direction) affects shape perception, we varied the direction of light such that its slant ranged between 0 and 66° in 11° increments over the course of the experiment. There is ample research suggesting a light-from-above prior, and a review of relevant literature revealed that the average assumed light source is above and to the left 12° from vertical [59, 41, 1, 8]. We therefore constrained our light source to always be tilted in this direction, rotating it to be slanted either above or below the

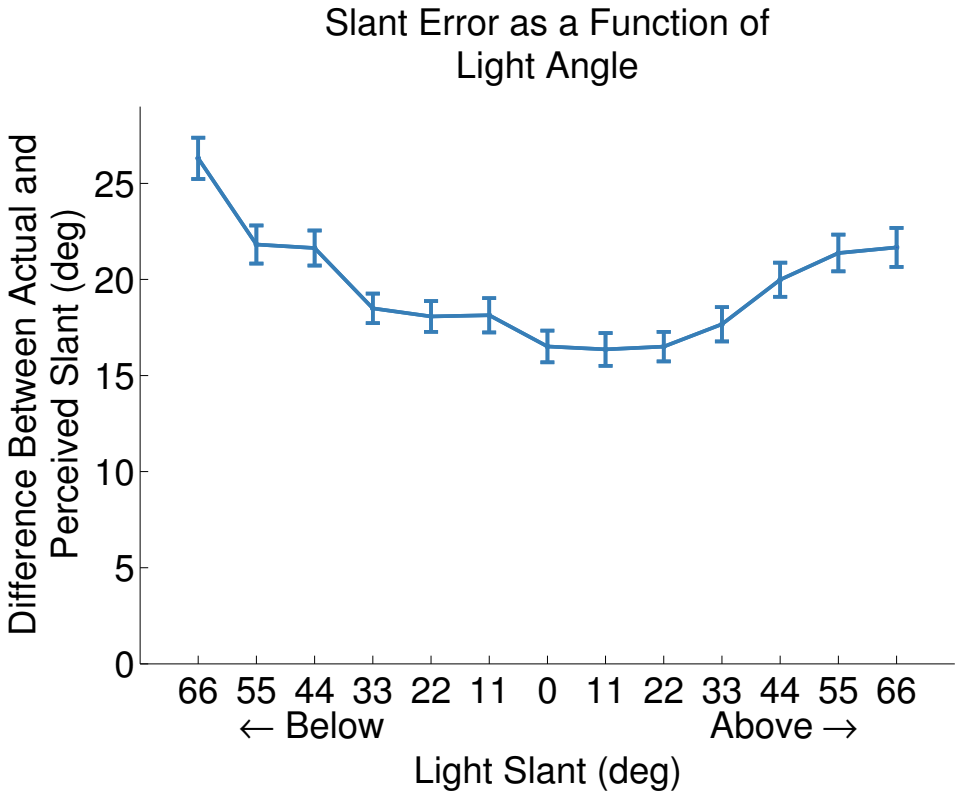

Figure 2.6. *Average slant error (deg) between the estimated and actual surface normal as a function of light angle (deg). Errors were minimized when the light was slanted approximately 0-22° above the viewpoint.*

viewing direction (Figure 2.2). In our coordinate system, this is equivalent to varying the light slant and keeping the light tilt fixed at 102°.

2.3.6 Experimental Design

We varied three parameters throughout the experiment: lighting direction, surface slant, and surface tilt. We predetermined the parameters of each trial to guarantee that a uniform sampling of surface slants and tilts would be tested under each lighting direction. We presented these trials to the subject in a randomized order. The angle between the light direction and the view direction varied from 0 to 66° in 11° increments, both above and below the viewpoint. We tested surface slants from 0 to 80° in 20° increments, and surface tilts from -120° to 180° in 60° increments. These parameters yield 390 unique combinations, each of which was tested three times for a total of 1170 trials.

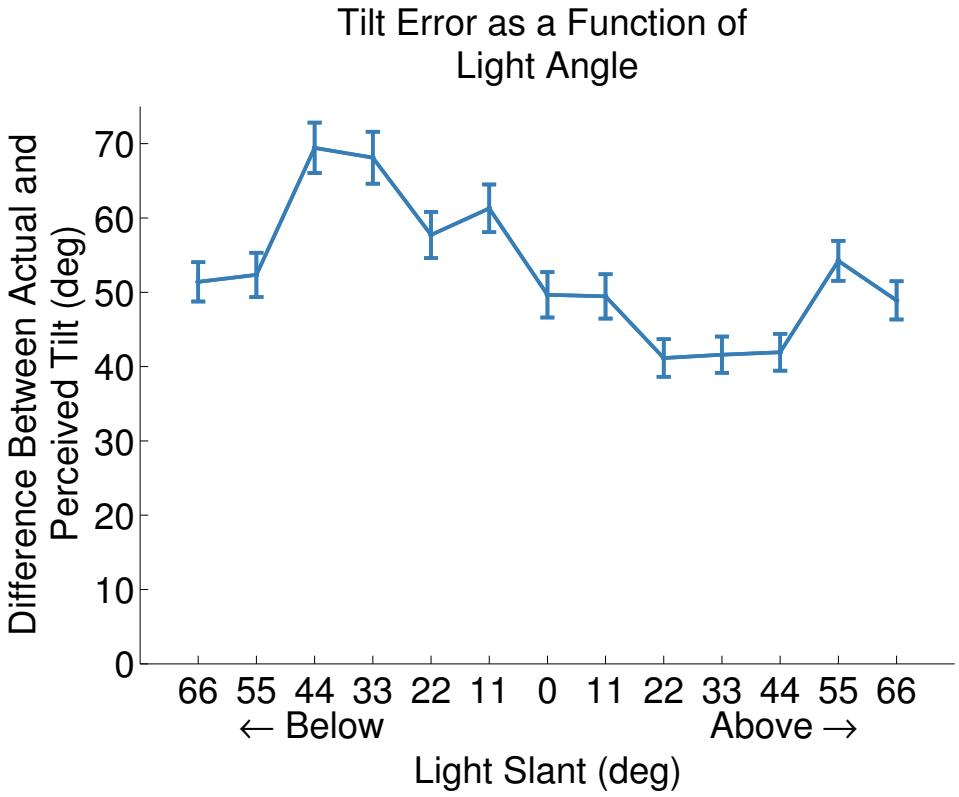

Figure 2.7. *Average tilt error (deg) between the estimated and actual surface normal as a function of light angle (deg). Similar to other measurements of error, subjects performed optimally when the illumination source was slanted 22° above the viewpoint.*

2.3.7 Procedure

For each trial, we presented the image of a surface to the subject with the gauge figure superimposed at the location to be tested. Using the mouse, subjects oriented the gauge figure until it appeared normal to the surface. The setting was made by clicking the right mouse button. Between trials, a gray screen was displayed for 500ms to minimize the after-image effect on the perceived shading of the next trial. The entire experiment was self-paced, and the data was collected over several one-hour sessions which were broken up into 20 minute blocks. See Figure 2.3 for a screenshot from the experiment.

2.4 Results

We recorded the surface location, the setting made by the subject, and the actual surface normal for each trial of the experiment. From these data, we calculated slant and tilt components for both the surface and the gauge figure. If the visual system assumes the direction of illumination is slanted away from the viewing direction at a

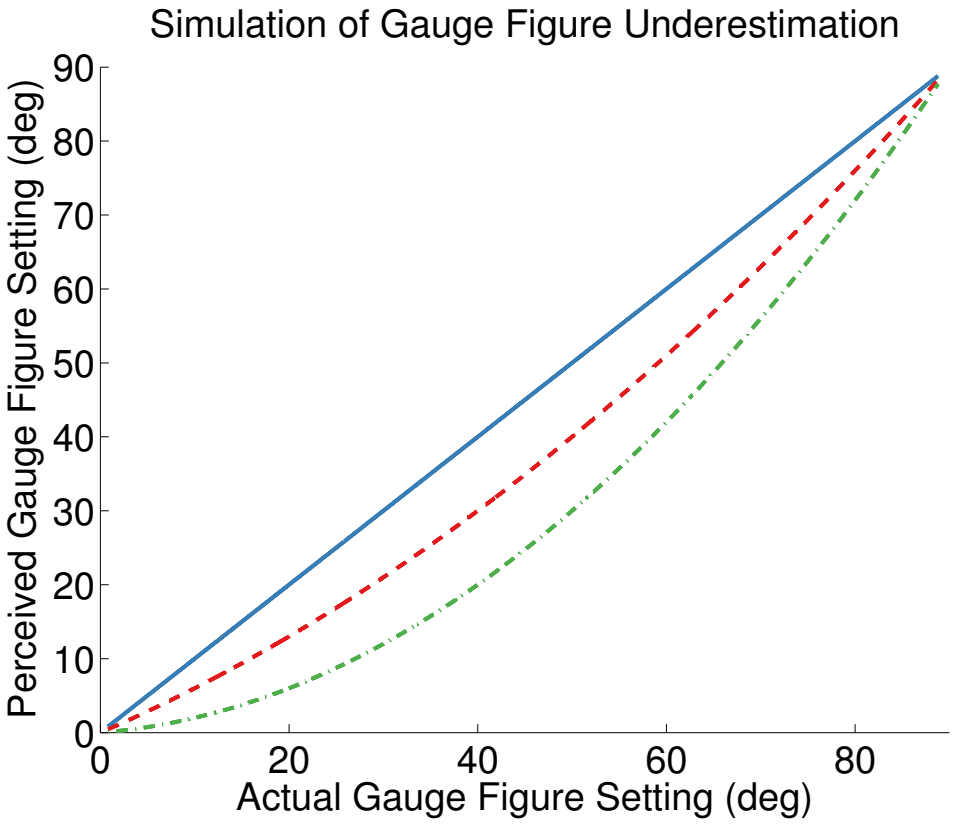

Figure 2.8. *Proposed model of gauge figure misperception. The solid blue line represents a veridical gauge figure model in which the setting perfectly represents the subject's perception. It is possible that subjects underestimated the slant of the gauge figure when it was greater than slant 0 and less than slant 90. We modeled both a 10° maximum underestimation (red line) and a 20° maximum underestimation (green line). We applied these functions to the data in order to assess how this misperception might have changed our results.*

particular angle, we expect the experimental settings to be most accurate when the lighting condition matches this assumption.

We first assessed subject performance by calculating the angular difference between the subject setting and the actual surface normal. Individual results for the three subjects are shown in Figure 2.4. The results averaged across the subjects after normalizing the data are shown in Figure 2.5. These data confirm the light-from-above prior. As the light direction moves from below the viewpoint to above it, average errors decrease. Errors are smallest when the light is slanted 20-30° above the viewing direction.

We also analyzed errors in surface slant and tilt estimation as a function of light direction. (Figures 2.6 and 2.7). Slant error was calculated as the angular difference (in°) between the actual surface slant and the slant indicated by the subject's setting. Tilt error was similarly calculated. These results are consistent with the previous

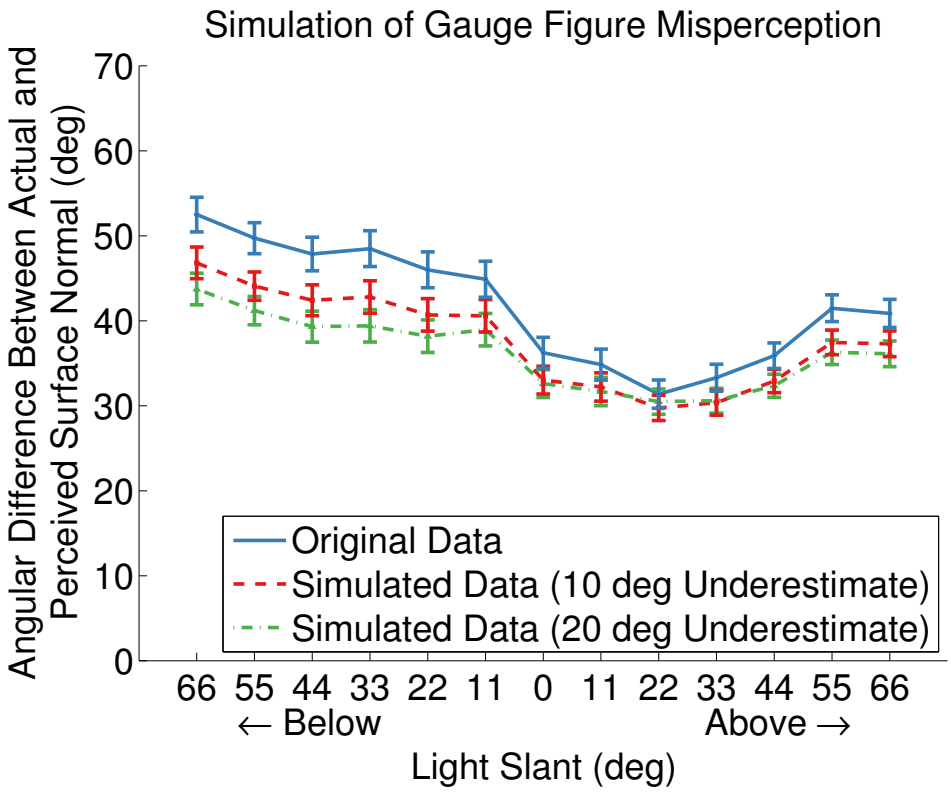

Figure 2.9. *Simulated effect that a systematic underestimation of the gauge figure slant would have on shape perception. These data represent the difference between the actual and perceived surface normal. Although the error introduced by the misperception of the gauge figure changes the data slightly, the overall results remain the same. We conclude that an underestimation of the gauge figure slant would not affect our findings.*

findings, confirming a preference for light from above as well as showing that errors were smallest when the light was slanted 20-30° above the viewing direction.

Although these three measures of error are consistent with each other, we consider the angular difference between the actual surface normal and the subject's setting to be a better estimate of performance. Slant and tilt errors can sometimes be misleading. When the surface slant is near zero, small errors in setting the gauge figure can disproportionately contribute to large errors in tilt. Likewise, large errors in setting the gauge figure may not be reflected by slant alone if the error is primarily in the tilt direction.

These findings provide further evidence for a light-from-above prior. Regardless of how we measured the error, subjects performed best when the light direction was above the viewpoint. Furthermore, these results show that the most accurate settings were made when the angle between the light direction and the view direction was 20-30°. The results of our study suggest that the visual system may assume light is slanted at this angle when presented with Lambertian shading without cast shadows.

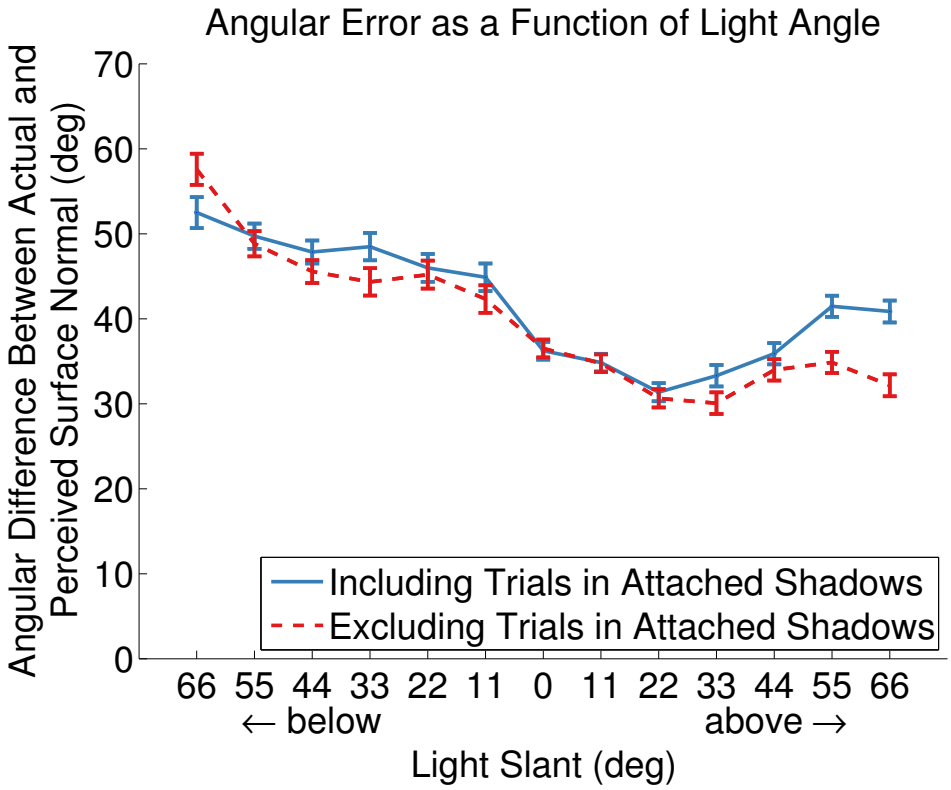

Figure 2.10. *Average errors after removing trials in which the gauge figure was in attached shadow. This graph shows the angular difference (deg) between the indicated and actual surface normal as a function of light angle (deg). We excluded 14% of the trials for each subject. We remove these trials from the analysis because they often occurred in areas of the surface that lacked local shading variation.*

2.5 Error Simulation

In most perception studies, it is impossible to verify that the results accurately reflect observer percepts. The experimental technique may introduce errors that affect the subject responses. In our case, it is possible that subjects misperceived the slant of the gauge figure. Our experiment was not designed to measure this type of error directly. In order to assess whether such errors might have affected our results, we simulated the error by applying a simple error model to our data. Our model of the error is designed to overestimate the magnitude of these errors, and is therefore a conservative model.

Consider the following example: we present a surface slanted at 55° but the subject perceives it to be 45°. If the experiment does not introduce any additional error, then we would expect the subject's setting to be 45°. This setting would be ideal because it would accurately reflect perception. Alternatively, if the subject systematically underestimates the gauge figure's slant, he or she would likely set it to be less than

45°. In this case, the experiment would be introducing an error, and the subject's response would no longer be an accurate measure of perception.

We modeled this error as a simple quadratic function (Figure 2.8). We do not expect subjects to misperceive the gauge figure's slant near 0° or near 90°, and our model of the error reflects this assumption. We simulated both a 10° and 20° maximum underestimation of the gauge figure slant. In both cases, the simulated data produce slightly different results (Figure 2.9), but the general findings remain the same. Errors in surface perception are minimized when the light is slanted 20-30° above the viewpoint. This simulation demonstrates that a systematic underestimation of the gauge figure slant would most likely not affect our findings.

2.6 Discussion

We demonstrated experimentally that viewers make more accurate estimations of surface orientation for shaded images when the light direction is slanted approximately 20-30° above the viewing direction. These results provide evidence that the visual system assumes this lighting direction in order to disambiguate the shape depicted in images rendered using local Lambertian shading without cast shadows.

One of the goals in computer graphics is to convey 3D information on a flat display screen. For a 2D image plane, there are non-pictorial cues to depth that specify a different shape than the effect simulated by shading. Previous research has demonstrated that the brain combines different cues to depth in an optimal fashion in order to achieve an overall depth estimate [19]. In addition to shading, the visual system utilizes cues such as disparity, defocus, and motion parallax to help discern the 3D shape of an object [63, 21]. When viewing objects in the real world, all of these cues provide a consistent 3D shape, yet this is not the case with simulated 3D scenes on a 2D display. In fact, the non-pictorial cues in these situations specify that the observer is really viewing a flat surface. Thus, shape cues such as shading need to be effective enough to overcome the flatness cues and thereby yield the desired percept for the simulated object.

Regardless of how we measured subject error, performance was always found to be best when the light was approximately 20-30° above the viewpoint. At higher angles, errors increased even though these conditions were still consistent with a light-from-above prior. One explanation is that these extreme light directions illuminated the object with a more raking angle, and a greater portion of the surface was oriented more than 90° away from the light source. These surface areas correspond to attached shadows, and thus are completely dark. This shading is due to the local diffuse shading model (Eq. 2.1). Without adequate variation in the shading of the surface around the gauge figure, it may be impossible to make any estimation of the surface orientation.

To asses how this situation may have affected the results, we ran a second analysis of the data excluding trials in which the gauge figure was positioned in an area of

attached shadow (Figure 2.10). We excluded 164 of the 1170 trials from each subject (14%). In general, we found a similar pattern of errors in this analysis. The main difference is that errors were slightly smaller when the light was directed from the highest angles above the viewpoint. Based on this analysis, it is unlikely that the overall findings were significantly affected by trials located in regions of attached shadows.

There is also the possibility that cast shadows would affect our results. We used a local Lambertian shading model which does not include computations for cast shadows. The visual system may make incorrect surface estimations when cast shadows would otherwise exist. As such, our results primarily apply to local diffuse shading models that do not compute cast shadows. Despite the simplicity of this model, it is commonly used in computer graphics applications.

As previously mentioned, there is evidence that the visual system has a prior for global convexity [33], which could affect shape perception. Although we do not control for this prior in our experimental design, it is unlikely to affect our findings. Because our stimuli are globally convex, any effect of the convexity prior would most likely improve shape perception and weaken our findings. The fact that we still find an effect of light direction despite this prior only strengthens our results.

It is also possible that our class of surfaces influenced the overall findings. Depending on the geometry of the surface, certain lighting directions may provide more useful shading information compared to other lighting directions. These differences may not be the same for all of the surfaces used in our study, which may affect subject performance. Figure 2.11 provides example images of different surface types after applying the results from our study. Previous research has attempted to quantify the amount of shading information in an image [57, 17], but it is still not known how the visual system ultimately makes use of this information. Further work is needed to distinguish the affect of surface geometry from the visual system's assumptions about lighting.

We acknowledge that there are many sources of error in this type of shape perception study. The flat display screen provides conflicting depth cues, the gauge figure may be misperceived, and the shading may produce errors. We do not know how these sources of error combine to affect the recorded observer percepts. In addition, we cannot infer the strength or shape of the error distribution attributed to changes in the light direction alone. This type of characterization of the assumed light angle would require further study.

2.7 Conclusion

In this study, we have confirmed the visual system assumes light is above when viewing the image of a shaded 3D surface. Additionally, we have demonstrated that the viewer's perception of shape is more accurate when the angle between the light direction and viewing direction is 20-30° above the viewpoint. This experiment provides

evidence that the visual system assumes this angle of lighting when presented with local Lambertian shading information without cast shadows.

2.7.1 Future Work

There are two remaining questions which deserve further study. First, we would like to investigate the extent to which the gauge figure task accurately reflects observer percepts. Some of the error in the experiment may be attributed to the task itself, so it would be useful to measure this effect. One approach would be to conduct a similar experiment using a known, cue-rich stimulus such as a real object. We can assume the subject would accurately perceive the object's shape. Any errors in the shape perception task could therefore be attributed to the task itself.

A second question to explore is whether more realistic shading, such as Lambertian shading with cast shadows, would affect the findings. As previously noted, subjects may incorrectly estimate surface orientation for regions where cast shadows would normally exist. An investigation of lighting priors may be more accurate if the stimuli provide more realistic shading cues.

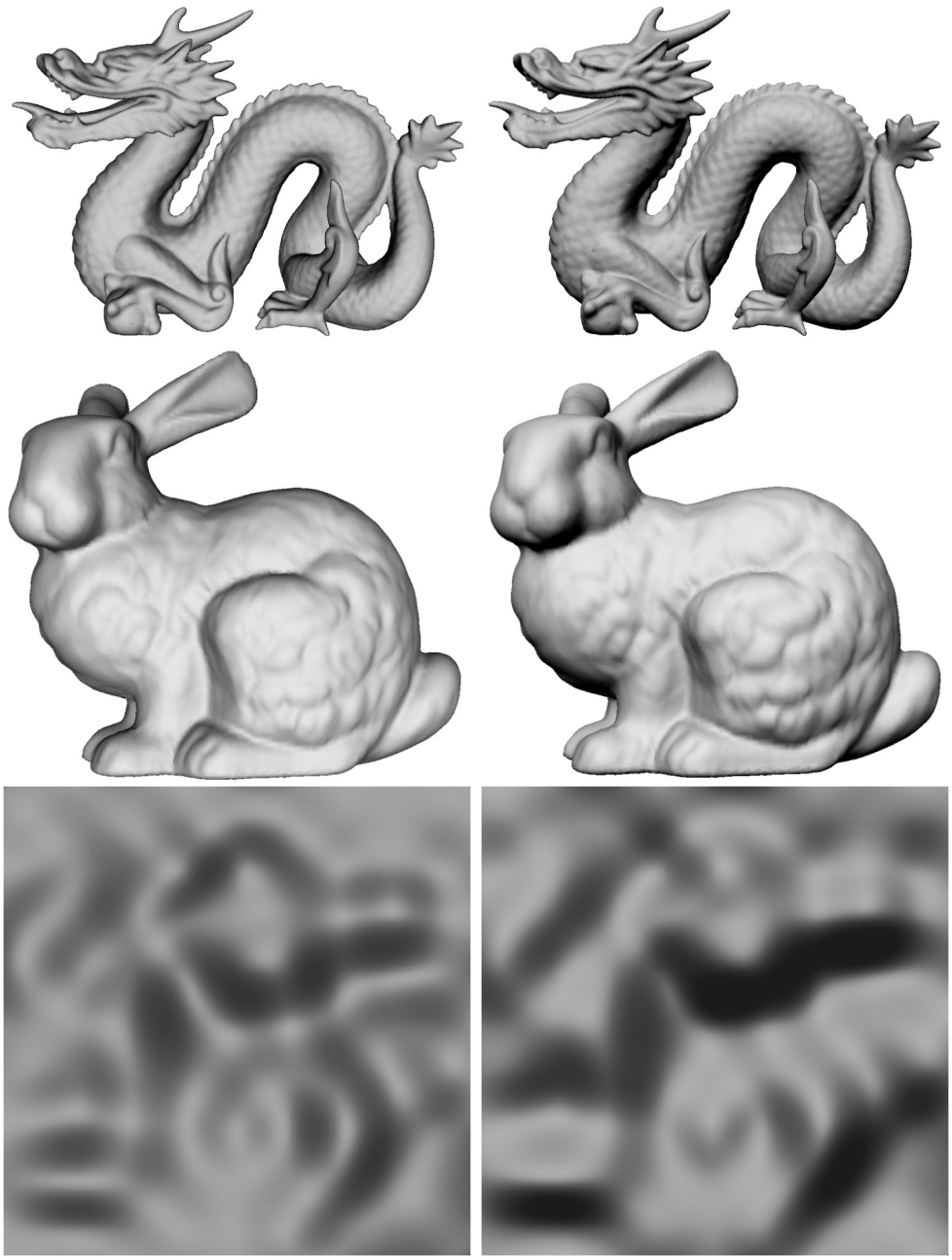

Figure 2.11. *Example surfaces displayed using two different light directions. The images on the left were shaded with the light at the viewpoint. The images on the right were shaded with the light angled 26° above the view vector. We found shape perception was best for the lighting direction used in the images on the right.*

Chapter 3

Lighting

3.1 Introduction

The variation in luminance across the image of a surface provides information about the 3D shape of the surface, the material of which the surface is composed, and the lighting or illumination falling on the surface. Although the physics of light transport is well understood [26], it remains unclear how human observers estimate reectance, lighting, and shape from a single image. Consider the simple case of Lambertian reectance, a single distant light source that is always visible and within 90- of any surface normal, and no interreections. Under these assumptions, the local diffuse shading equation is

$$I(x, y) = \rho\big(\vec{N}(x, y) \cdot \vec{L}\big) \tag{3.1}$$

where ρ is the constant albedo, $\vec{N}(x, y)$ is the surface normal at point (x, y) in the image, and $(\vec{L})$ is the vector pointing toward the light source. Thus, the observed luminance in the image is determined by the reflective properties of the material and the orientation of the surface normal relative to the lighting direction. Because shape, lighting direction, and material properties all determine the observed image, one cannot in general solve for any one of those properties without knowing the other two. The problem of solving for any of these terms (shape, lighting direction, or reectance) is consequently ill-posed.

Human observers have stable percepts of lighting direction [30] and stable percepts of shape [32] in shaded images, which makes sense if the objects shape or the lighting direction is known, respectively. However, observers also have stable percepts of lighting direction when the shape is not known [29] and stable percepts of shape when the lighting direction is not known [20]. The latter observations suggest that observers are making assumptions about the two unknown properties in order to solve for the third. Here we investigate how observers use the available sensory data

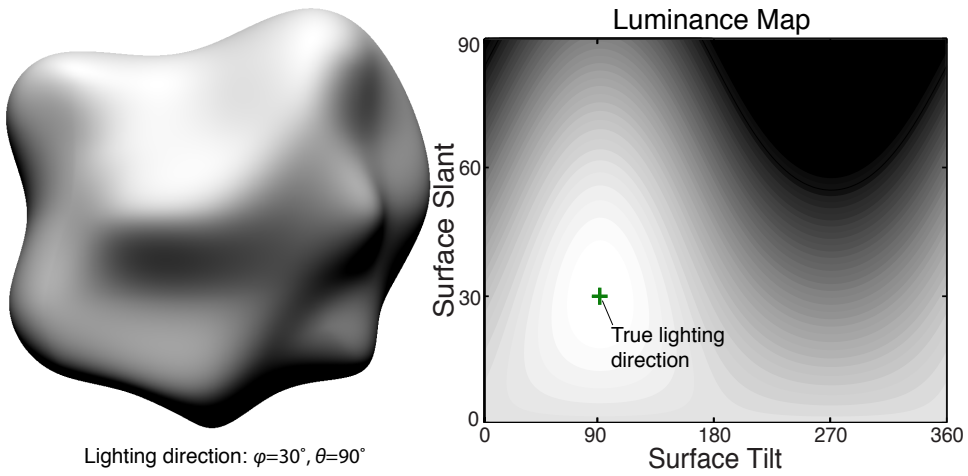

Figure 3.1. *Shaded object and its associated luminance map. The object (left) is illuminated from above (light slant $\phi = 30°$, light tilt $\theta = 90°$). The luminance map plots the observed luminance value for each surface orientation in the image. The green cross indicates the true lighting direction.*

and assumptions to turn an ill-posed problem of estimating lighting direction into a solvable one.

3.1.1 Models

3.1.1.1 Shape-based method

The lighting information contained in shaded objects is illustrated in Figure 3.1. On the left is an irregular object illuminated by a distant point light. Assume that the observer knows the surface shape. Then one can determine how the observed luminance varies as a function of the 3D orientation of different parts of the object. On the right, we plot luminance as a function of surface slant and tilt, where slant ϕ is the angle between the line of sight and the surface normal, and tilt θ is the direction of that angle relative to horizontal [58]. The plot is regular with a clear peak. Assume also that the surface material is Lambertian and the reectance is constant. Then from Equation 3.1, the luminance at each point in the luminance map on the right is informative about the direction of the light source. For instance, the point of maximum luminance has a surface normal pointing toward the light. Points with half the maximum luminance have surface normals that point $60°(cos(60°) = \frac{1}{2})$ away from the light. Assuming that the surface albedo is constant and that the surface slant and tilt are estimated with some degree of accuracy, we can rewrite Equation 3.1 as the product of two vectors:

$$I(x, y) \propto \vec{N}(x, y)^T \vec{L}, \tag{3.2}$$

where $\vec{N} = \begin{bmatrix} \vec{N}_x \\ \vec{N}_y \\ \vec{N}_z \end{bmatrix}, \vec{L} = \begin{bmatrix} \vec{L}_x \\ \vec{L}_y \\ \vec{L}_z \end{bmatrix}$, and $\|N\| = \|L\| = 1$.

Using the positive values ($I(x, y) > 0$), we can set up a linear system of equations:

$$A\vec{L} = B, \tag{3.3}$$

where $A = \begin{bmatrix} \vec{N}_x(x_1, y_1) & \vec{N}_y(x_1, y_1) & \vec{N}_z(x_1, y_1) \\ \vec{N}_x(x_2, y_2) & \vec{N}_y(x_2, y_2) & \vec{N}_z(x_2, y_2) \\ \vdots & \vdots & \vdots \\ \vec{N}_x(x_k, y_k) & \vec{N}_y(x_k, y_k) & \vec{N}_z(x_k, y_k) \end{bmatrix}$

and $B = \begin{bmatrix} \vec{I}(x_1, y_1) \\ \vec{I}(x_2, y_2) \\ \vdots \\ \vec{I}(x_k, y_k) \end{bmatrix}$ for k positive points in the image.

We can then estimate the lighting direction $\vec{L}$ using a linear least-squares approach:

$$L = (A^T A)^{-1} A^T B. \tag{3.4}$$

Thus, we can solve for the lighting direction when surface shape and reectance are known. To estimate the lighting direction in this way, the visual system must measure the luminances and the 3D orientations of points on the object (and the assumption of constant Lambertian reectance must be valid). Because this method depends on knowing or estimating the 3D surface geometry, we refer to it as the shape-based method for estimating lighting direction. If the luminance and orientation measurements are erroneous, the estimate of lighting direction will be correspondingly erroneous.

3.1.1.2 Image-based approach

Pentland [51] developed one of the first image-based methods for estimating lighting direction in an image. Precluding any direct estimation of the 3D geometry of the surface, but assuming the surface normals are isotropically distributed, he showed that the slant and tilt of the lighting direction can be estimated from the local shading derivatives. Lee and Rosenfeld [34] extended and improved the method, deriving the following estimate of lighting tilt θ:

$$\theta = arctan\left(\frac{E(I_y)}{E(I_x)}\right), \tag{3.5}$$

where $E(I_y)$) is the maximum-likelihood estimate of the image derivative along the y direction and $E(I_x)$ is the maximum-likelihood estimate of the image derivative along the x direction. The slant ϕ of the lighting direction is

$$\phi = arccos\left(\frac{3E(I^2)}{\lambda^2\rho^2} - 1\right), \tag{3.6}$$

where λ is the illumination brightness, ρ is the surface albedo, and $E(I^2)$ is the expectation that I^2 taken along the tilt direction θ. The slant and tilt estimates determine the lighting direction as follows:

$$L = \begin{bmatrix} sin(\theta)cos(\phi) \\ sin(\theta)sin(\phi) \\ cos(\phi) \end{bmatrix} \tag{3.7}$$

Thus, we can solve for the lighting direction using the 2D content of the image and an estimate of albedo and illumination brightness, provided that the global shape of the surface is convex. Because this approach is based on the 2D image information, we refer to it as the image-based method.

3.1.2 Outline

In this chapter, we first investigate whether human observers use a shape-based or image-based approach to estimate the lighting direction in a scene. We then examine humans ability to infer lighting direction when the material property was provided and shape information was indicated to greater or lesser extents. In so doing, we learn more about the computations and assumptions that viewers make while interpreting shaded images. When the sensory data (i.e., the image content) reliably specified the 3D shape, the observer could in principle determine the lighting direction (Equation 3.4). When the sensory data did not specify the 3D shape well, the observer could not determine the lighting direction from those data directly; in that case, the observer had to make assumptions about the shape and the lighting, and those assumptions were presumably based on previous experience.

In our analysis of the results, we use a Bayesian framework to fit a model to the data. We show that a simple model incorporating information from the sensory data and expectations based on previous experience fits the data well. The best-fitting model implies that observers use the sensory data and prior expectations, but that they rely on the prior expectations when the sensory data are unreliable. By making observers familiar with the material properties and manipulating the shape information, we were able to determine the usefulness of various shape cues and to measure the direction and variance of the prior for lighting direction.

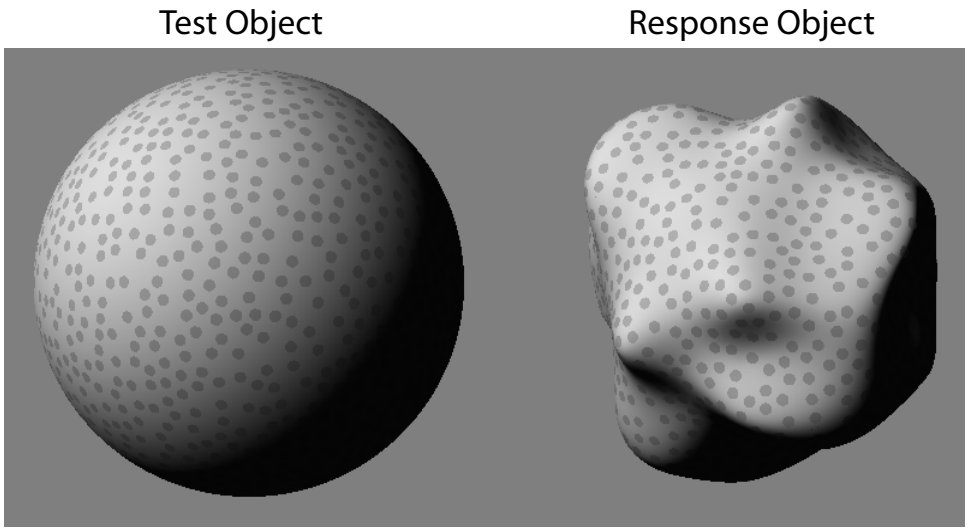

Figure 3.2. *Experimental stimuli. The test object was rendered with different sets of shape cues in different conditions; in the example above, all cues were present including having the familiar shape of a sphere. The response object was always rendered with all shape cues (disparity, texture gradient, global convexity, occluding contour, and shading). Observers adjusted the direction of the light on the response object until it appeared to be the same as the light illuminating the test object.*

3.2 General Methods

3.2.1 Apparatus

We displayed stimuli on a custom stereoscope with two arms that rotated about vertical axes co-linear with the rotation axes of the eyes [3]. Each arm used a mirror to position the image from a CRT in front of each eye. The CRTs were ViewSonic G225f displays with a resolution of 1280 $\times$ 1024. The physical distance from each eye to the appropriate CRT was 39 cm, so each pixel subtended 2.8 $\times$ 2.6 arcmin. We gamma-corrected each display to linearize the luminance function for the grayscale images. Except for the CRTs, the room was dark. The observers stabilized their head position using a bite bar fastened to an adjustable mount. We adjusted the separation of the rotation centers of the stereoscope arms to match each observers inter-ocular distance. We rotated the arms so that the vergence angle matched the viewing distance of 39 cm.

3.2.2 Lighting Parameters

We parameterized lighting direction in terms of slant and tilt in much the way Stevens [58] described surface slant and tilt. Lighting slant (ϕ) is the angle between a line from the eye (or cyclopean eye) to the center of the object and a vector from the center of the object to the light. Lighting tilt (θ) is the angle between the horizontal

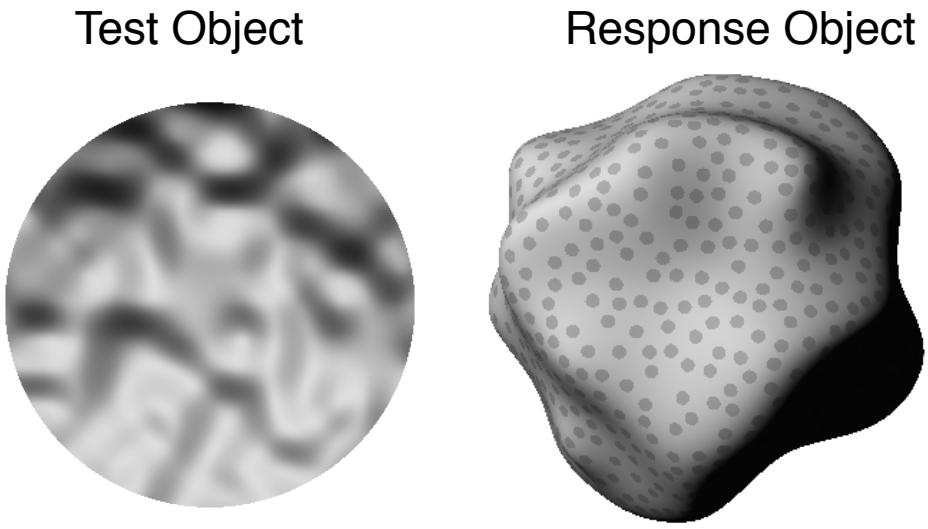

Figure 3.3. *Examples of stimuli used in Experiment 1. The test surface was always globally concave, and we varied the disparity information used to specify its 3D shape. The response object was convex. Observers adjusted the direction of the lighting on the response object until it appeared to be the same as the lighting direction on the test object.*

axis and the projection of the lighting direction onto the frontal plane. The slant and tilt of the light correspond, respectively, to the zenith and azimuth of the light [37].

3.2.3 Simuli and Procedure

On each trial, we presented two stimuli simultaneously: a test object and a response object (Figure 3.2). The two objects were rendered using the OpenGL graphics library and C++. The objects were composed of matte material (i.e., Lambertian reectance). We told the observers that the material was similar to matte paper. We also showed them that the luminance does not vary as a function of viewpoint. We made clear that the stimuli were composed of the same material as used in the demonstration. The response object contained all of the shape cues under investigation. We chose to examine a set of shape cues commonly found in the shape perception literature.

1. *Shading: Each pixel was given the appropriate luminance given the objects shape, Lambertian reectance, and the direction of the light source at infinite distance. The shading method correctly generated attached shadows. Cast shadows were not rendered, but the lighting slant was never greater than 45°, so there would have been few such shadows anyway.*

2. *Texture gradient: : We applied a texture by rendering small gray disks on the object. The disks were oriented in the tangent plane of the surface. We positioned the disks using a dart-throwing algorithm to generate a Poisson-disk*

sampling of the vertices [12]. The texture gradient fully specified the 3D shape of the object up to an unknown scale factor. Because the light source is at infinite distance, the scale factor does not need to be known to estimate the lighting direction, so the texture provided the information required to estimate lighting direction accurately (Equation 3.4).

3. *Binocular disparity: Each point on the object was given the appropriate horizontal and vertical disparities for the specified shape. This cue fully specified the 3D shape of the object and therefore provided the information needed to estimate light direction accurately (Equation 3.4).*

4. *Global convexity: When this cue was present, the object was a sphere with random radial perturbations. Because the object was approximately spherical, the orientation at a point on the surface was highly correlated with the points position in the image. For example, surface points above and to the right of the center of the image had tilts on average of 45°. Thus, observers could in principle have used such regularity to estimate light direction [37]. The fact that the surface was globally convex is consistent with the convexity assumption observers tend to make about surfaces [33, 40].*

5. *Occluding contour: When this cue was present, the silhouette of the object was visible. The silhouette provides information about 3D shape [22, 38] The slant of the surface at the occluding contour is 90° because that part of the surface is by definition orthogonal to the viewing direction. The tilt is equal to the orientation of the tangent to the contour at that point. Because surface orientation is known at the occluding contour, luminance values along the contour could provide useful information about the lighting direction. Of course, the surface at the occluding contour is invisible to the viewer, so one cannot measure luminance at precisely that point, but one can estimate the luminance by extrapolating from nearby points [46]. From these measurements, one can estimate the tilt of the light: Specifically, the tilt is perpendicular to the orientation of the contour at the brightest point on the occluding contour. We can see this relationship in the luminance map in Figure 3.1 where the brightest portion of the map at large surface slants indicates the light tilt.*

3.3 Experiment 1

We first investigated whether observers use a shapebased or image-based approach to estimate the lighting direction in a scene. To do so, we displayed irregular test objects that were globally concave and varied the disparity information specifying the 3D shape. The response object was always globally convex.

First, consider the predictions for image-based methods. If the lighting direction on the test and response objects were the same, the shading patterns on the two

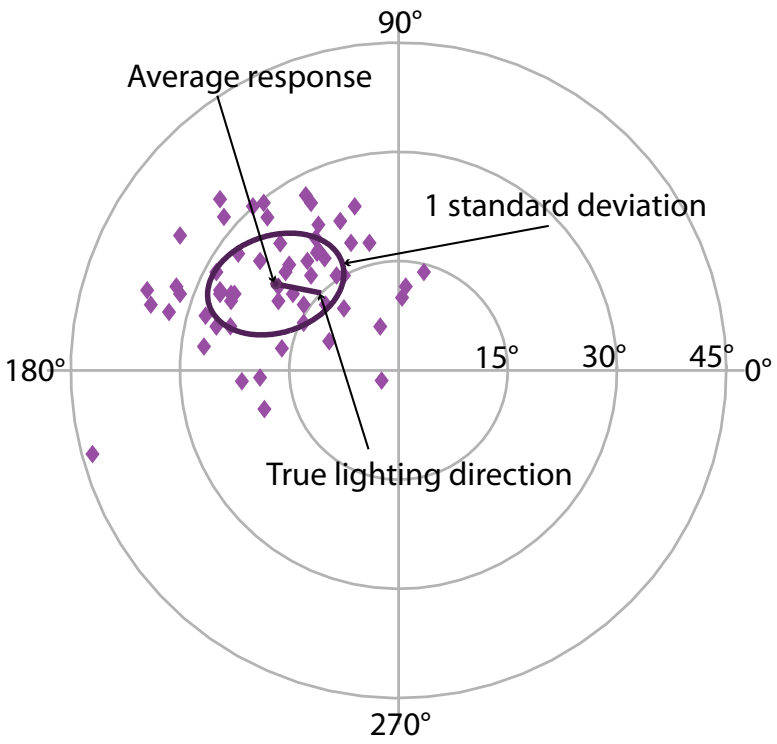

Figure 3.4. . *Responses from one subject for one lighting direction. The dots represent individual settings plotted in polar coordinates in which the angle is lighting tilt and the radius is lighting slant. The ellipse is centered on the mean of the settings and is adjusted in size and orientation to contain all of the settings out to one standard deviation from the mean. The ellipse axes are oriented along the principal components of the data, which are the orthogonal vectors that account for the maximum amount of variance (determined by a principal component analysis using the Karhunen-Loeve transform [25, 36]. We draw a line from the true lighting direction to the mean response to show the bias in the settings.*

objects would be in opposite directions (Figure 3.3). For example, if the lighting on both objects was from above, the response object would be brighter on the top than the bottom and the test object would be brighter on the bottom than the top. To match the shading patterns, the observer would have to set the tilt of the lighting direction on the response object 180° from the lighting tilt on the test object. Thus, image-based methods should yield tilt errors of 180°.

Now consider the predictions for the shape-based method. We manipulated the information specifying the test objects 3D shape by setting the disparities to zero (specifying a at surface) or to the correct values for the shape (specifying a concave surface). With zero disparities, the shape of the test object was ambiguous, and observers generally perceived the shape as globally convex [33, 40]. In this case, they would set the tilt of the lighting direction on the response object 180- from the tilt on the response object. With correct disparities, the shape of the test object was well specified and observers would therefore set the tilt of the lighting on the

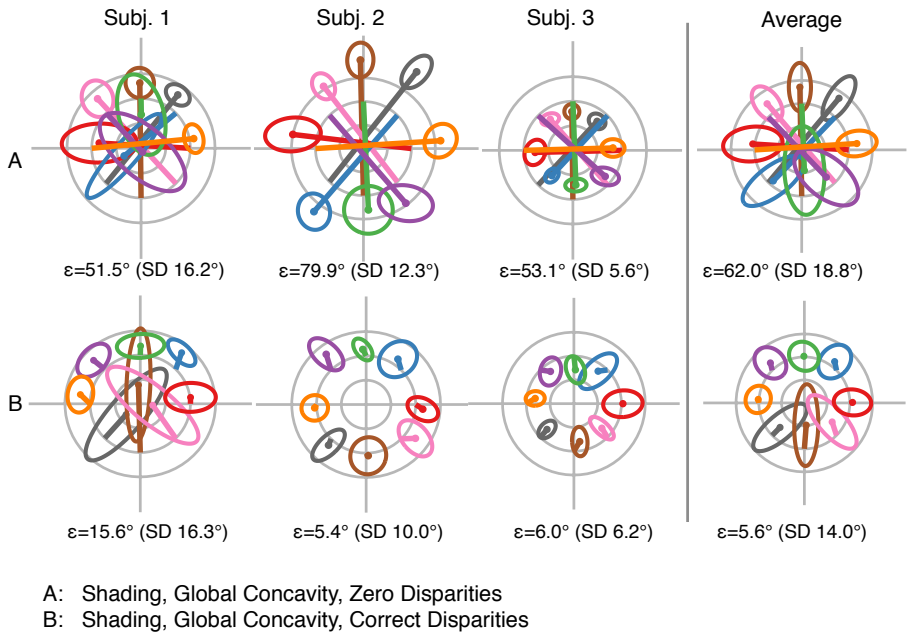

Figure 3.5. *Summary of data from Experiment 1. (Top) In Condition A, the globally concave surfaces were specified using only shading information. The disparities were set to zero. (Bottom) In Condition B, the same surfaces were specified with shading and correct disparities. Colors represent different lighting tilts on the test object. The ellipses summarize the mean and standard deviation of the settings for each condition with lighting tilt and slant represented as in Figure 4. The line segments connect the center of each ellipse (the average setting) to the true lighting direction. The numbers under each plot are the average angular error (ϵ) and the average standard deviation of the settings (SD).*

response object to a value close to the tilt of the lighting on the test object. Thus, the condition with zero disparities yields the same predictions for the image- and shape-based methods, and the condition with correct disparities yields entirely different predictions for the two methods.

3.3.1 Methods

3.3.1.1 Observers

Three female observers participated. They were 22-27 years of age and had normal visual acuity and stereopsis. They wore their optical corrections during testing. They were experienced psychophysical observers but were unaware of the experimental hypothesis.

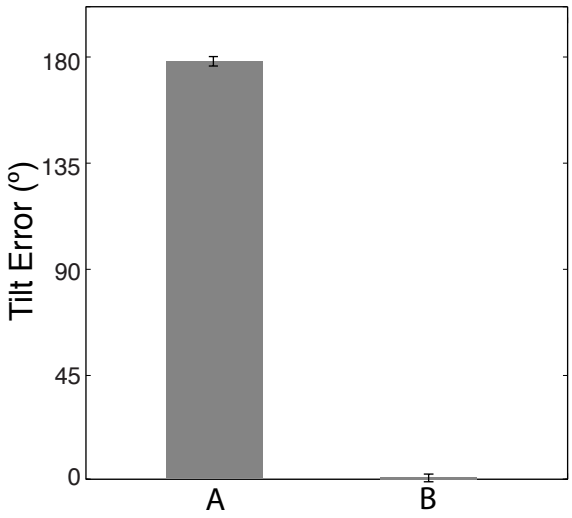

Figure 3.6. *Average tilt errors across observers for the two conditions. We calculated the tilt error by measuring the angular tilt difference between the average response and the true lighting direction. Error bars are standard errors of the mean. Observers averaged 177.6° tilt errors when the concave shape of the test object was poorly specified (Condition A). When the shape of the test object was well specified (Condition B), observers average 0.46° tilt errors.*

3.3.1.2 Lighting parameters

We presented eight lighting tilts (0, 45, 90, 135, 180, 225, 270, and 315°) while keeping the lighting slant at 30°.

3.3.1.3 Shape conditions

We presented the test objects with two different combinations of shape cues.

A: *Shading and global convexity:* We made disparity uninformative in this condition by presenting the test objects with zero disparity. They were shaded appropriately for a globally concave object. The image- and shape-based methods both predict 180° tilt errors.

B: *Shading, global concavity, and binocular disparity:* The test objects again were globally concave with appropriate shading. Disparities were correct and therefore specified their true shape. Image-based methods predict 180° tilt errors, and shape-based methods predict small errors.

We generated the 3D objects by subdividing the triangles of an initial control mesh. We created the spheres by subdividing an icosahedron (20-sided regular poly-

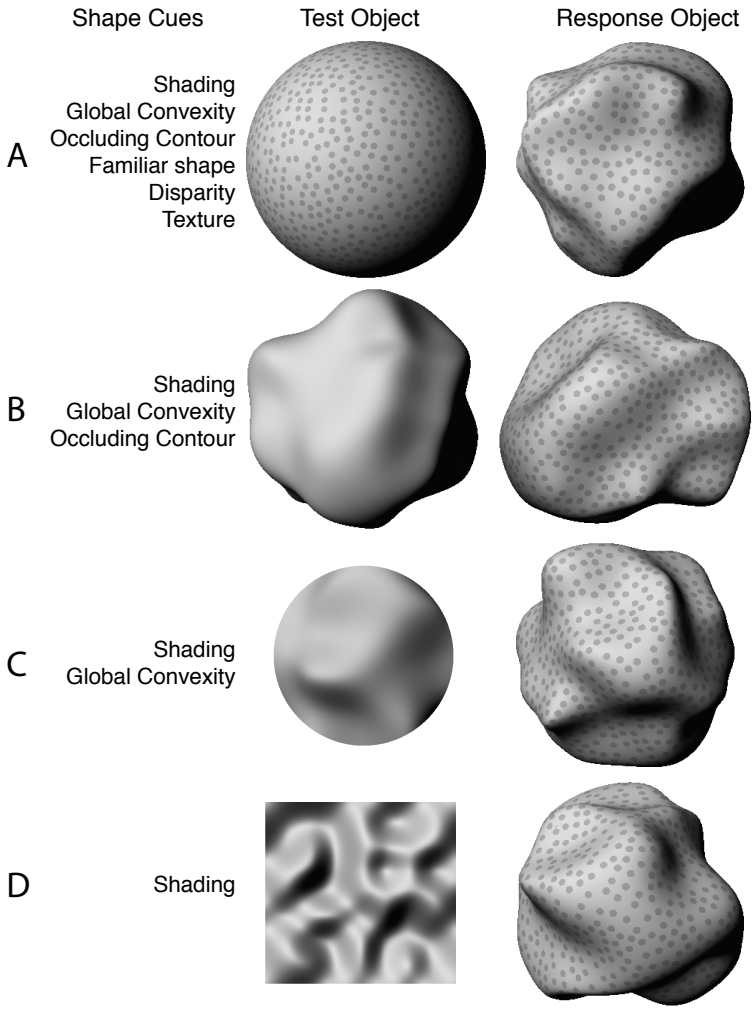

Figure 3.7. *Sample stimuli. The shape of the test object was specified by different sets of cues in each condition (listed on the left). The shape of the response object was specified by the same set of cues in all four conditions (disparity, texture gradient, global convexity, occluding contour, and shading).*

hedron) and normalizing the vertices to be equidistant from the origin of the object. The final spheres were composed of approximately 25,000 triangles. The irregular shapes were created using an implementation of the Catmull-Clark subdivision surfaces algorithm that generates smooth surfaces with C^1 continuity [7]. The resulting objects were spheres with random radial perturbations (Figures 3.2 and 3.3). We created the surface perturbations by randomly displacing the position of each vertex prior to the third iteration of the subdivision. We continued to run the subdivision algorithm until each object consisted of approximately 100,000 polygons. The globally at shapes were generated from a planar control mesh. The resulting objects were planes with random perturbations in depth.

We illuminated the test and response objects with point light sources at infinite distance, one source for the test object and another for the response. We told the

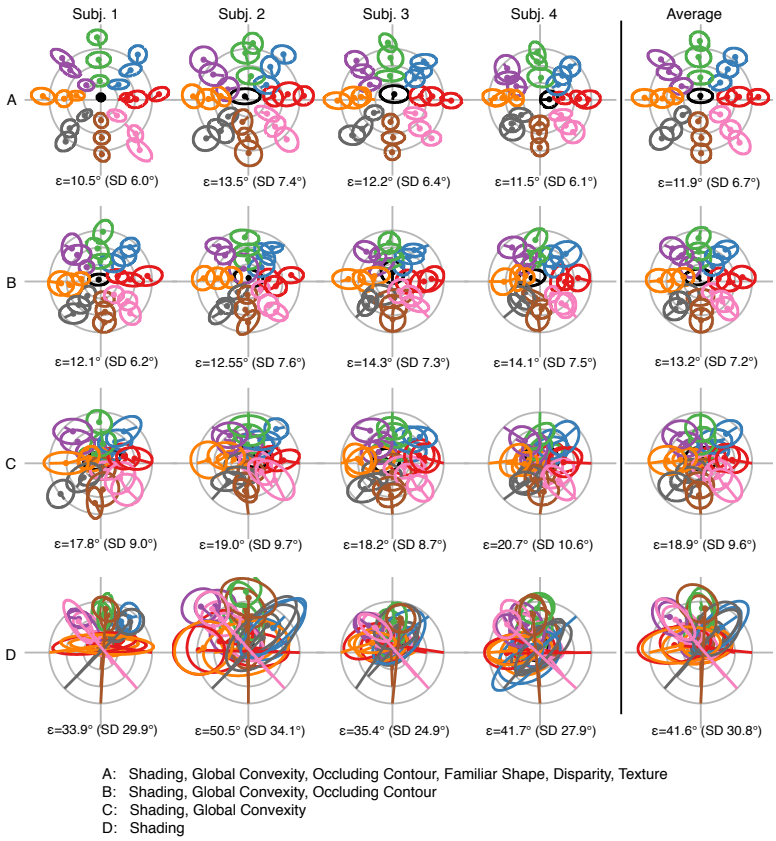

Figure 3.8. . *Summary of the data from all shape-cue conditions and observers. The columns show data from different observers and the average. The rows show the data from the four test-object conditions. Colors represent different lighting tilts on the test object. The ellipses summarize the mean and standard deviation of the settings for each condition, lighting tilt, and lighting slant as described in Figure 3.4. The line segments connect the average setting to the actual lighting direction. Numbers under each plot are average angular error (ϵ) and average standard deviation of the settings (SD)*

observers that the light was at infinite distance and thus similar to the sun. Observers moved a trackball to adjust the 2D orientation (ϕ, θ) of the light on the response object. Their task was to make the lighting direction on the response object match the perceived lighting direction on the test object. Lighting direction was not changed online with the trackball movement. Instead, after adjusting the trackball, observers clicked a button to update the lighting on the response object. Thus, they could not see changes in shading due to movement of the light source and, therefore, could not use light motion as an additional cue to shape. They kept making adjustments until the perceived lighting directions on the response and test objects were the same. They indicated that they were the same by clicking a mouse button. The test object appeared on the left for half the trials and on the right for the other half. A new pair of objects appeared on each trial. The shape of the response object was always well

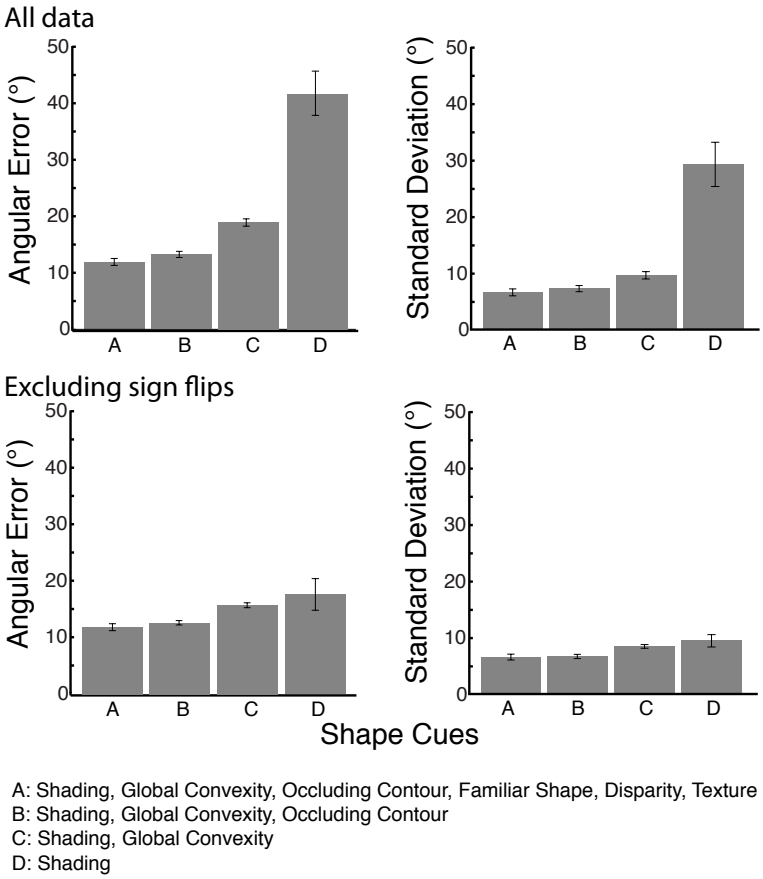

Figure 3.9. *(Upper row) Average angular errors (left) and average standard deviations (right) for each shape-cue condition using all trials. The errors and standard deviations are expressed in angular units irrespective of the light direction. Error bars represent standard errors of the mean. (Lower row) Average angular errors (left) and average standard deviations (right) for each shape-cue condition after excluding ips in the sign of the tilt. We define a sign ip as any trial with a tilt error between 135 and 225°.*

specified, so observers should have accurately perceived its shape regardless of how accurately they perceived the shape of the test object.

To measure the perceived direction of the illuminant, we could conceivably have used an estimation procedure such as asking observers to indicate light direction with a pointer. We chose not to use this approach because we had no way of knowing the mapping between perceived direction and pointer orientation, the so-called *response-mapping problem*. Said another way, one cannot know from the responses of such an estimation procedure which effects are due to the mapping between the percept and the response and which effects are directly indicative of the percept. By focusing on perceptual equivalence, we can be more confident that our results reect perceptual processes.

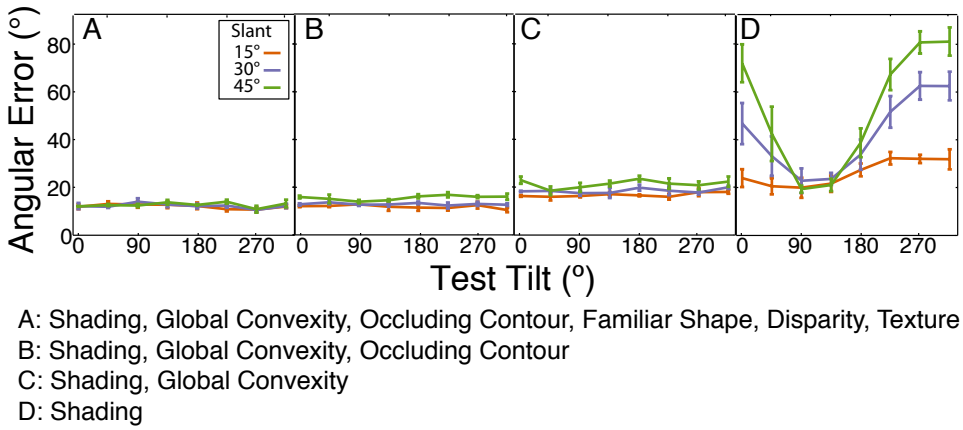

Figure 3.10. *Angular error for different lighting tilts. Each panel plots angular error averaged across observers as a function of lighting tilt. Different colors represent the data for different lighting slants, red for 15°, blue for 30°, and green for 45°. (A-D) The data from Conditions A-D. Error bars represent standard errors of the mean.*

Before running the experiment, we familiarized observers with the physics of lighting and shading by showing sample surfaces with various lighting directions. We did not use these surfaces as stimuli in the actual experiment. We collected 20 settings from each observer for each lighting direction in each disparity condition, yielding 320 settings per observer.

3.3.2 Results

Each setting is the lighting direction on the response object that the observer perceived as the same as the lighting direction on the test object. Figure 3.4 shows how we plot the settings for each observer in each condition. Each dot is one settingVa combination of lighting slant and tiltVin polar coordinates. The ellipses are best fits to capture one standard deviation in all directions. The line segments connect the actual lighting direction with the average setting.

Figure 3.5 summarizes the individual observer and average data for the two conditions. The columns and rows show the data from different observers and different conditions, respectively. All observers behaved similarly, so we can focus on the data averaged across observers, which are shown in the rightmost column. Without reliable shape information to specify that the test surfaces are concave, all observers made large errors in Condition A, primarily due to 180- tilt errors. The average angular error was 62.0° (Figure 3.5), and the average tilt error was 177.6° (Figure 3.6). When correct disparity information reliably specified the 3D shape, observers made 5.6° average angular errors (Figure 3.5), and the tilt error was only 0.46° (Figure 3.6).

Adding the correct disparity information had a significant effect on the results. Specifically, the errors in setting lighting direction were significantly smaller when the test objects shape was well specified.

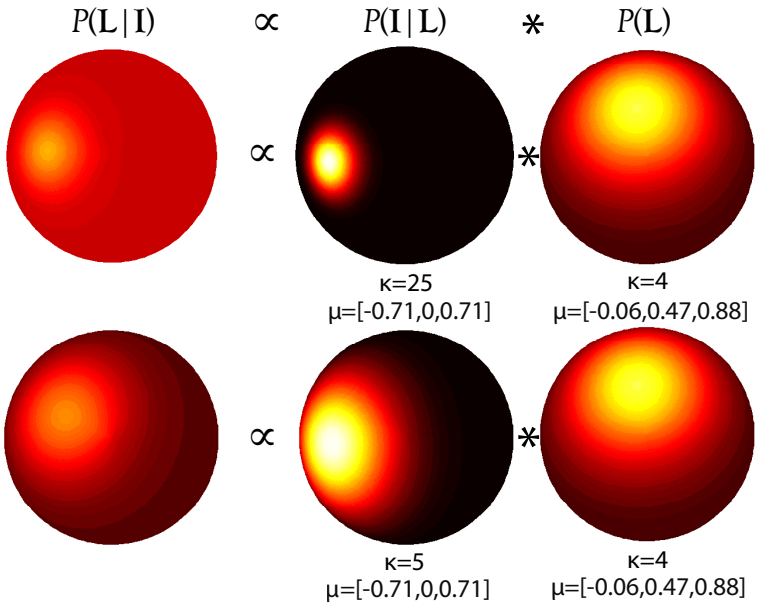

Figure 3.11. . *Von Mises-Fisher (VMF) distributions and implementation in Bayesian framework. (Upper row) From left to right, the VMF distributions represent the posterior, likelihood, and prior distributions in a Bayesian framework. The posterior is the product of the likelihood and prior. Here the likelihood has lower variance than the prior, so the posterior is similar to the likelihood. (Lower row) A likelihood with greater variance (lower κ) leads to a greater inuence of the prior on the posterior.*

3.3.3 Discussion

The same stimuli were used in the two conditions, so the 2D image content was identical. Thus, any image- based method would have yielded the same pattern of settings whether disparity was informative or not. This means that observers used the 3D shape information to match the true lighting directions even though it produced opposing 2D shading patterns on the test and response objects. The results, therefore, demonstrate that people use a shape-based approach to estimate lighting direction.

3.4 Experiment 2

We next examined how 3D shape information affects estimates of lighting direction. Specifically, we varied the shape cues used to specify the test objects in the same matching task. Our analysis of shading suggests that with reliable 3D shape information, observers should be able to accurately estimate the lighting direction. When the 3D shape is poorly specified, we expect observers to rely more on their prior expectations of lighting direction.

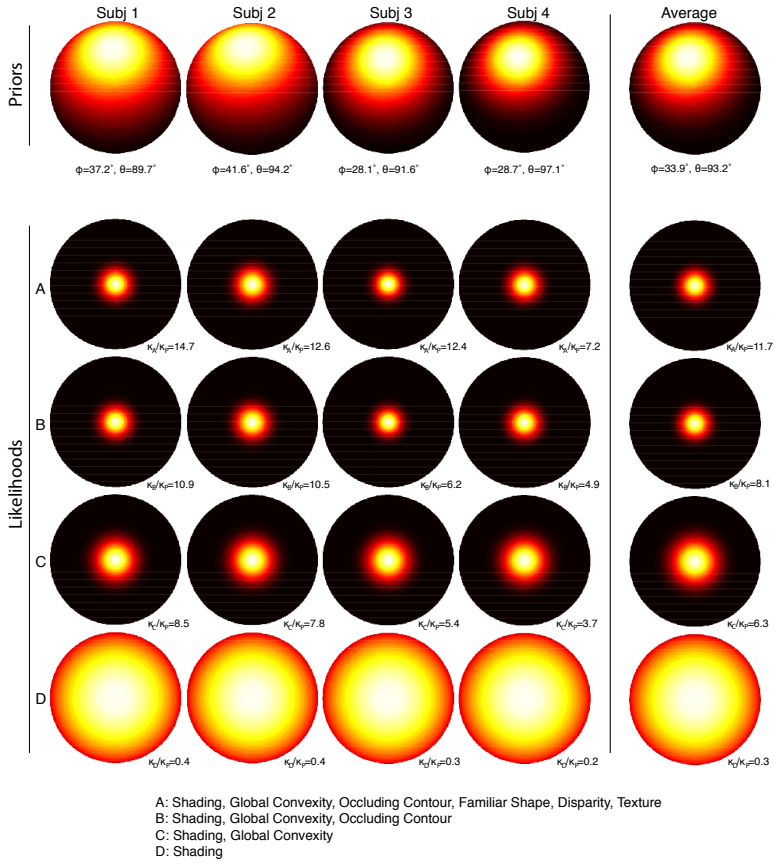

Figure 3.12. . *The likelihood and prior distributions that best fit each observers data. The first four columns represent the distributions for individual observers. The rightmost column represents the distributions derived from fitting the combined data from all four observers. The ratio of the best-fitting likelihood variance to the best-fitting prior variance is given below each likelihood panel. In this graphical representation, the means of the likelihoods have been set to [*$\phi = 0, \theta = 0$*], but the means were actually the true lighting slants and tilts for each lighting direction.*

3.4.1 Methods

3.4.1.1 Observers

Four female observers participated. They were 22-28 years of age and had normal visual acuity and stereopsis. They wore their optical corrections during testing. They were experienced psychophysical observers but were unaware of the experimental hypotheses.

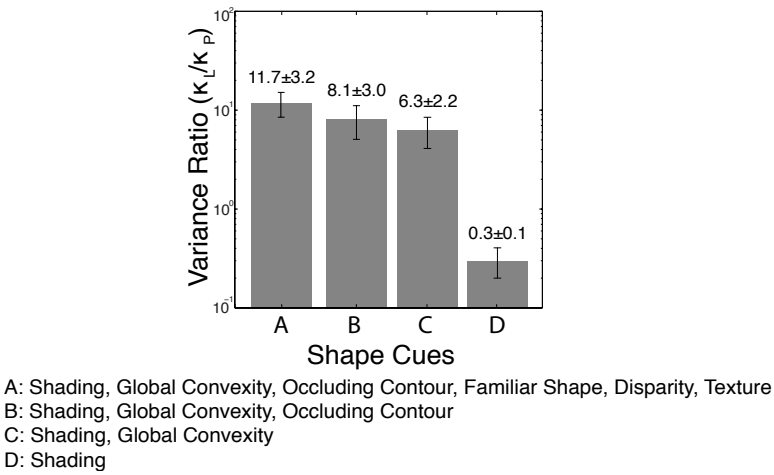

Figure 3.13. *The ratio of the likelihood κ_L and prior κ_P for the four shape conditions. The ordinate is the likelihood variance parameter (e.g., κ_A, κ_B) divided by the prior variance parameter (κ_P). The values that went into computing the ratio are the across-observer averages. Error bars represent the standard error of the mean.*

3.4.1.2 Lighting parameters

We presented each combination of four lighting slants (0, 15, 30, and 45°) and eight lighting tilts (0, 45, 90, 135, 180, 225, 270, and 315°). Tilt is undefined when slant is 0°, so we considered a total of 25 combinations of lighting directions.

3.4.1.3 Shape conditions

We presented the test objects with four different combinations of shape cues (Figure 3.7).

A: *All cues present:* The test objects were rendered with shading, global convexity, occluding contour, binocular disparity, and texture gradient. The cue of familiar shape was also present in that the test object was a sphere, a well-known shape. Because all cues were present in Condition A, the 3D shape was very well specified.

B: *Shading, global convexity, and occluding contour:* We eliminated disparity by presenting the stimuli monocularly. We eliminated the texture gradient by deleting the random-element texture. We eliminated familiar shape by using 3D shapes that were randomly perturbed as shown in Figure 3.7. By comparing responses in Condition B to those in Condition A, we could assess the combined contribution of the texture gradient, disparity, and familiar shape in specifying 3D shape and thereby aiding the estimation of lighting direction.

C: *Shading and global convexity:* We eliminated the occluding contour by presenting the stimulus in a circular software aperture. By comparing responses in this

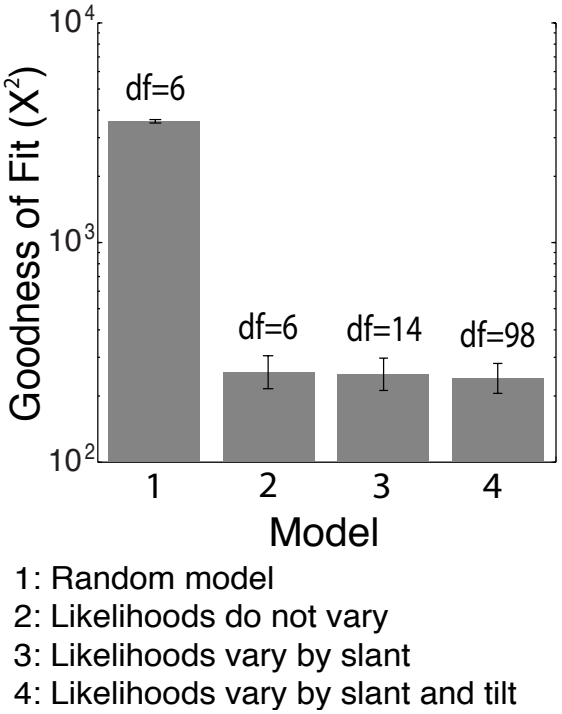

Figure 3.14. *Goodness of fit for the models tested. Chi square (χ^2) is plotted for each model. Model 1 is a random model in which the observed settings were randomly reassigned to condition and we then found the six parameters that provided the best fit to the reassigned data. Model 2 is the model used to generate Figure 3.12; we found the six parameters that provided the best fit to the observers settings across experimental conditions. Model 3 is similar to Model 2 except that the likelihood κ is allowed to differ for different lighting slants; the model has 14 free parameters. Model 4 is similar except the likelihood κ is allowed to differ for different lighting slants and tilts; it has 98 free parameters.*

condition to those in Condition B, we could determine the role of occluding contour in specifying 3D shape in the estimation of lighting direction.

D: *Shading only:* We eliminated global convexity by creating the stimulus from a frontoparallel plane (rather than a sphere) that was randomly perturbed in depth as shown in Figure 3.7. We clipped the stimulus with a square aperture to avoid an additional cue to convexity. By comparing responses in this condition to those in Condition C, we could determine the role of global convexity in specifying 3D shape and thereby aiding the estimation of lighting direction. Performance in this condition also tells us how well people can use shading alone to estimate light direction.

We collected 20 settings from each observer for each lighting direction in each shape condition, yielding 2000 settings per observer.

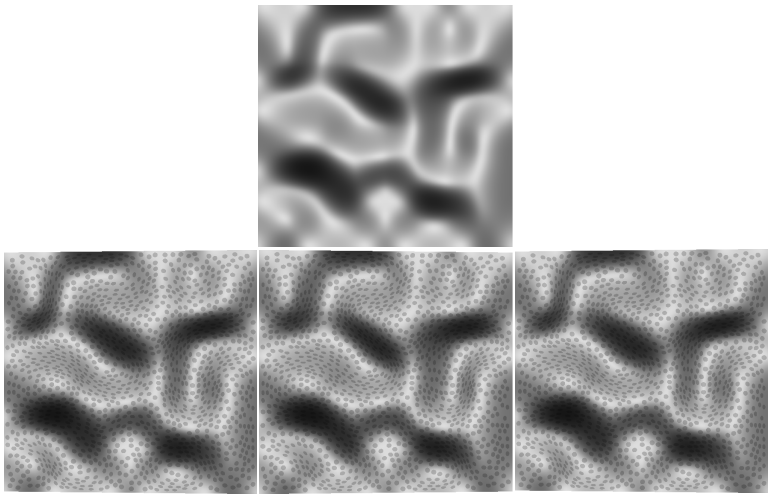

Figure 3.15. *Demonstration of how shape information affects the estimation of lighting direction. (Upper row) A shaded object in which the 3D shape is poorly specified. Most viewers of this image think the illumination is from above. (Lower row) The same object and illuminant, but now the shape is well specified. For cross-fusing, use the two panels on the left. For divergent fusing, use the two on the right. When the image is fused, the objects 3D shape is specified by disparity and texture as well as shading. If you are correctly fusing, you should see the brighter parts of the surface as slanted top near and bottom far. Most viewers of the lower panels now think the illumination is from below. The reliable shape cues in the lower panels allow one to estimate the true light direction (Equation 3.4). Those cues are not available in the upper panel, so the light-direction prior dictates the perceived light direction.*

3.4.2 Results

Figure 8 summarizes the individual observer and average data for the four test-object conditions. The columns and rows show the data from different observers and different conditions, respectively. The data were quite similar across observers, so we can focus on the data averaged across observers, which are shown in the rightmost column. Changing the set of available shape cues had a systematic effect on observers settings. The left panel of Figure 3.9 plots the average angular difference between the actual and responded lighting directions, while the right panel plots the standard deviation of the settings.

To determine which effects were statistically reliable, we conducted repeated-measures ANOVAs with angular error and with standard deviation as dependent measures. With angular error as the dependent measure, there were significant effects of shape-cue condition, lighting slant, and lighting tilt on angular error ($p < 0.001$ in all three cases); there were also significant interactions of condition and slant and of condition and tilt ($p < 0.001$ in both cases). With standard deviation as the dependent measure, there were significant main effects of condition, slant, and tilt

($p < 0.001$) and again significant interactions of condition and slant and of condition and tilt ($p < 0.001$).

Settings were most accurate in the full-cue condition (Condition A). Figure 3.8 shows that the angular errors in this condition were smallest and did not vary systematically with lighting tilt or slant. The average angular error was only 11.9° (Figure 3.9, upper row). The settings were also the most precise in this condition. The best-fitting ellipses in Figure 3.8 were small for all tilts and slants. The average standard deviation was only 6.7° (Figure 3.9, upper row).

The settings in Condition B were somewhat less accurate than those in Condition A. The average angular error and average standard deviation were slightly greater at 13.2° and 7.2°, respectively (Figure 3.9, upper row). These values were significantly greater than in Condition A: $t(6) = 1.6, p = 0.04$ (one-tailed) and $t(6) = 1.4, p = 0.05$ (one-tailed), respectively. The small decrease in performance shows that the cues of familiar shape, disparity, and texture provided useful information for specifying shape and thereby aided the estimation of lighting direction. It is somewhat surprising, however, that removing these shape cues had such a small effect; we will return to this observation in the Discussion section.

The settings in Condition C were less accurate than those in Condition B. The average angular error and standard deviation were now 18.9° and 9.6°, respectively (Figure 3.9, upper row). Both of these values were significantly greater than in Condition B: $t(6) = 6.7, p < 0.001$ and $t(6) = 4.5, p < 0.01$. The decrease in accuracy and precision means that the occluding contour (the cue not presented in Condition C) helped specify the shape of the test object and that observers used this greater specification to make better settings.

The settings in Condition D were much less accurate than those in Condition C. The average angular error and standard deviation were 41.6° and 30.8°, respectively (Figure 3.9, upper row). Both of these values were significantly greater than in Condition C: $t(6) = 5.7, p < 0.001$ and $t(6) = 10.2, p < 0.001$. These results show that observers are much better at determining the lighting direction when the object is globally convex than when it is not. The pattern of errors is particularly interesting. Figure 8 shows that when the lighting direction was below the line of sight (i.e., lighting tilt was between 180 and 360°), observers often made tilt errors of ˜180° in their settings. In other words, they perceived the light as above the line of sight even though it was below. This pattern of responses resulted in a bimodal distribution of settings and contributed to the large angular errors in Condition D (Figure 3.9, upper row).

This observation is clearer in Figure 3.10, which plots average angular error as a function of lighting slant and tilt. Notice that tilt had essentially no effect on error in Conditions A, B, and C but had a large and systematic effect in Condition D. In particular, large errors were observed when the tilt was between 180 and 360°, i.e., cases in which the actual light direction was below the line of sight. We also examined angular errors after excluding trials with tilt errors of ˜180°. Specifically, we excluded

trials for which the tilt error was between 135 and 225°. The results are shown in the lower row of Figure 3.9. The average angular error and standard deviation for Conditions A-C are nearly unchanged, but the average error in Condition D decreased from 41.6° to 17.7° and the standard deviation from 30.8° to 9.3°. Thus, the error pattern in Condition D shows that shading information alone is not sufficient for viewers to estimate lighting direction; when shape is not specified by other cues, they tend to see the light as coming from above the line of sight even when it is actually coming from below.

3.4.3 Bayesian model

To further analyze the data, we used a Bayesian framework to represent the information about lighting direction contained in the sensory data and the information provided by previous experience, and the means by which observers should combine such information. Bayes Rule provides the optimal method [27]:

$$P(L|I) \propto P(I|L)P(L). \tag{3.8}$$

The first term on the right side of the equation is the likelihood distribution, which represents the information in the sensory data (i.e., the image I). In this chapter, we do not present a generative model of how surface shape, material properties, and illumination combine to produce the likelihood distribution. We simply use the distribution to represent the light-direction information available in the sensory data. We assume that the likelihood distribution is unbiased. The second term on the right side of the equation is the prior, which represents the distribution of likely lighting directions independent of the sensory data. We know that observers tend to assume that light comes from above and slightly to the left [1, 41, 50, 59]. Observers should base their estimates of lighting direction on the product of the likelihood and prior, which is the posterior distribution on the left side of Equation 3.8.

We parameterized lighting directions in spherical coordinates, so we used Von Mises-Fisher (VMF) distributions to model the data. The VMF distribution is an isotropic continuous probability distribution that describes spherical data with a mean of μ and a concentration of κ. The distribution on a sphere for $x \in R^3$ is

$$\gamma(\mathbf{x}; \mu, \kappa) = \frac{\kappa}{4\pi sinh(\kappa)} e^{\kappa(\mathbf{x} \cdot \mu)} \tag{3.9}$$

where $\kappa \geq 1$ and $\|\mu\| = \|\mathbf{x}\| = 1$. The parameter 2 has the coordinates $[cos(\theta)sin(\phi), sin(\theta)sin(\phi), cos(\phi)]$, which correspond to the Cartesian coordinates of lighting slant ($0° < \phi < 90°$) and tilt ($0° < \theta < 360°$). The parameter κ is inversely proportional to the spread of the distribution, so as κ increases, the variance of the distribution decreases. Figure 3.11 shows some sample distributions.

We assumed that observers based their judgments on the peak of the posterior dis-

tribution, which is proportional to the product of the likelihood and prior (Equation 3.8). In particular, we assumed that the judgments were derived from the maximum of the posterior. We then found the likelihood and prior distributions that best predicted the observers responses for each experimental condition. In doing so, we assumed that the likelihood distributions were unbiased (that is, that the peaks of those distributions were centered on the true lighting direction). In finding the best-predicting distributions, the likelihoods had one free parameter κ for each of the four test-object conditions. Thus, we found κ_A for the data in Condition A, and likewise κ_B, κ_C, and κ_D for the appropriate data sets. The prior had two free parameters for the coordinates of the peak of the distribution (ϕ_P and θ_P) and one parameter κ_P for the spread. We found one set of prior parameters for all four conditions because we assumed an observers prior did not change over the course of the experiment.

Because we assumed unbiased likelihoods, we set the means of the likelihoods equal to the coordinates of the actual lighting direction in each condition. As we said, there were four parameters for the variances of the likelihoods (κ_A, κ_B, κ_C, and κ_D) and one for the variance of the prior (κ_P). However, the position of the maximum of the posterior is determined by ratios of likelihood κ and prior κ, so there were only four free parameters for κ. To deal with this constraint, we set κ_D to 1 and found the best values for the other four. Thus, we fixed the likelihood locations for all lighting directions within a shape condition. We found the best values for the six free parameters for the complete set of data from each observer using a non-linear, least-squares optimization routine (Matlabs lsqnonlin routine). The routine found the set of parameters that minimized Chi square (χ^2), the sum of the squares of the angular errors. We did not attempt to fit the variances of the observer settings.

Figure 3.12 displays the likelihood and prior distributions that best fit each observers data. The first four columns represent the results for the four observers and the rightmost column the results averaged across observers. The top row represents the best-fitting prior distributions and the next four rows the best-fitting likelihood distributions for Conditions A, B, C, and D, respectively. Note that the means of the likelihoods have been plotted at [0, 0] because many different directions were actually presented and could not be readily shown in one graph.

The results were strikingly similar across observers. For example, the prior distribution is centered above the visual axis for all four observers; specifically, the best-fitting tilt varies from 89.7 to 97.1°; tilts greater than 90° are counterclockwise from vertical. This result is consistent with the aforementioned light-from-above prior [1, 41, 50, 59]. The prior distribution is also roughly equally displaced from the origin in all four observers; the best-fitting slant varies from 28.1 to 41.6°. This result is nicely consistent with our earlier finding that the assumed slant for lighting direction is 20 - 30° above the line of sight [50]. The best-fitting likelihood distributions were also remarkably similar across observers. The spread of the distributions increased in quite similar fashion for all four observers as we took shape information away in going from Condition A to Condition D.

As we said earlier, the location of the maximum of the product of two VMF

distributions is determined by the ratio of the distributions variances. The ratio reects the degree to which the likelihood or prior determines the location of the posterior. In Figure 3.13, we plot the average ratio of the likelihood and prior variances –e.g., κ_A/κ_P – for all observers for each shape condition. The ratio is large in Condition A where all shape cues were present, which is consistent with the fact that observers made quite accurate settings in that condition. As shape cues were taken away, the ratio became smaller, which is consistent with the observation that observers made successively less accurate settings as their settings drifted toward above the line of sight. Indeed, the ratio is less than 1 for Condition D where only shading was available, consistent with the observers relying primarily on their prior expectation of lighting direction in that case.

We next investigated how well our model fit the data compared with other plausible models. To do this, we computed χ^2 for four models.

1. The first was a random model with six free parameters (the same six as in the model that generated Figure 3.12). In this model, we first randomly reassigned settings to conditions (with replacement) and then we fit the parameters to the data. This model provides an upper bound on our measure of goodness of fit for comparison with the fits of the other models.

2. The second was the model described earlier that was used to generate Figure 3.12. There are six free parameters in this model.

3. The third model was similar to the second except that the likelihood variance parameters (κ_A, κ_B, κ_C, and κ_D) were allowed to differ for each of the four lighting slants. The parameters of the prior were the same as in Models 1 and 2. Thus, this model has 14 free parameters (two for the prior and 12 for the likelihoods).

4. The fourth model was similar to the third except that the likelihood variance parameters were allowed to differ for each combination of lighting slant and tilt. The parameters of the prior were the same as in the above models. Thus, this model has 98 free parameters (two for the prior and 96 for the likelihoods).

Figure 3.14 shows goodness of fit (χ^2) for the four tested models. As one would expect, the random model provides a poor fit. The other three models provide roughly equivalent fits. Models 3 and 4 have many more free parameters than Model 2, yet Model 2 provides essentially the same fit to the data. We conclude that Model 2 – the one used to generate Figure 3.12 – provides the most parsimonious account for the data.

3.5 Discussion

3.5.1 Summary of results

In Experiment 1, we showed that observers use 3D shape information to match the lighting direction in a scene. In Experiment 2, we examined how specific shape cues affect observer estimates of lighting direction. Our results show, as expected, that accurate perception of lighting direction depends on reliable shape information. When the 3D shape of the object was specified by many robust shape cues, observers estimated direction accurately. When the shape was poorly specified, responses were very inaccurate: In that case, the perceived direction was above the view direction and slightly counter-clockwise from vertical no matter what the actual light direction was. We used a Bayesian framework to model the data. The framework combined light-direction information contained in the images with a light-direction prior. The prior was centered above and slightly to the left: tilt and slant of 93.2° and 33.9°, respectively.

We summarize the findings with a simple demonstration in Figure 3.15. The upper panel is a shaded image of a surface whose 3D shape is poorly specified. The surface is globally at, the occluding contour is not visible, and disparity and texture are not available; the shape is specified by shading only. Notice that the light source appears to be above the panel. The lower panel is the same shaded image, but now 3D shape is well specified by disparity and the texture gradient. It is now evident that the light source is actually below. The figure shows that the light direction is correctly perceived when the shaded objects 3D shape is well specified and is incorrectly perceived to be in the direction of the light-from-above

3.5.2 Light-direction prior

There is a great deal of evidence that viewers assume light comes from above and slightly to the left. Convexity- concavity judgments are consistent with an assumed lighting tilt of ˜110° [1, 23, 41, 44, 59]. Speed of visual search is greatest when the tilt is roughly the same value [8, 13, 28]. To our knowledge, only one previous study has made measurements relevant to specifying the slant of the lighting prior. O'Shea et al. [50] showed that 3D shape judgments of shaded objects were most accurate when the lighting slant was 20-30°. prior when we specify the 3D shape using only shading.

With the exception of Morgenstern and Murray [44], our task was very different from the ones used in the above-mentioned studies. We estimated the prior for light direction by having observers adjust the direction of the illumination on an object whose shape was well specified to match the perceived direction of the illumination on an object whose shape was poorly specified. The average tilt of the prior was 93.2° and the average slant was 33.9°. These estimates of the prior parameters are

remarkably consistent with the estimates from previous work despite the use of an entirely different task.

3.5.3 Perceiving lighting inconsistencies

Ostrovsky, Cavanagh, and Sinha [49] reported that people have considerable difficulty detecting inconsistencies in the direction of lighting in scenes composed of several objects. In a visual search task, they presented nine objects. In one condition, all nine were illuminated with the same light. In another, eight of the nine were illuminated with one light and one was illuminated with a light whose tilt differed by 90°. The task was to indicate whether the lighting was consistent or inconsistent. The shapes were reasonably well specified; the stimuli were most similar to the test objects of Condition B in our experiment. Ostrovsky et al. found that people could discriminate the inconsistent from the consistent displays, but performance was far from perfect. The relatively poor performance seems inconsistent with our results. In our Condition B, observers made accurate and precise settings. Average angular error was 13.2° and the standard deviation was 7.2°, values that are much lower than the 90- differences in the lights in the experiment of Ostrovsky et al. Their result is similar to findings that people have difficulty detecting inconsistencies in attached and cast shadows in complex scenes [14, 39].

Ostrovsky et al. hypothesize that the visual system can compute illumination direction for individual objects when shape is well specified and this is consistent with our data. They also speculate that multiple estimates of light direction from various objects may not support any accumulation into a group direction (p. 1311). Thus, the limit may have to do with accumulating estimates from individual objects into one global estimate of scene illumination.

3.5.4 Effectiveness of different shape cues

We observed a small decrease in performance between Conditions A and B. Thus, removing the robust shape cues of disparity, texture, and familiar shape had a small but noticeable effect. This result means that observers could use the cues of occluding contour, global convexity, and shading to estimate lighting direction reasonably accurately, which is consistent with previous work on illumination matching [53]. We also found a small decrease in performance between Conditions B and C, which means that removing the occluding contour had a discernible but small effect. This result suggests that observers could use the remaining cues of global convexity and shading to estimate light direction fairly accurately. There was a large decrease in performance between Conditions C and D, which suggests that global convexity had a significant effect on the ability to estimate light direction.

We first consider the information provided by the occluding contour. As we noted earlier, the variation in luminance near the occluding contour of an object can be

used to estimate the tilt of the lighting [46] (Figure 3.1). This estimation technique has been utilized effectively to detect illumination inconsistencies within photographs [24]. One cannot, however, estimate the slant of the lighting from this cue without inferring the shape of the rest of the object. Thus, the cues of global convexity and shading must have been the primary determinants of direction estimates.

We next examine the lighting information available with globally convex stimuli and relate that information to observed performance. People tend to assume that surfaces are globally convex [33, 40]. This assumption is consistent with the test objects presented in Conditions A-C. As we said earlier, lighting direction can be recovered if the objects shape is known and the surface albedo is constant. The lighting information contained in shaded globally convex objects is illustrated in Figure 3.16. The stimulus is the irregular object in Figure 1 seen through an aperture so that the occluding contour is not visible. This stimulus corresponds to Condition C in our experiment. In constructing the luminance maps in the upper row, we assumed that the objects shape was estimated accurately. Thus, the plots of luminance as a function of surface tilt and slant are regular with clear peaks at the slant and tilt values corresponding to the slant and tilt of the light source. The luminance maps in the lower row were constructed with the same stimulus and lighting directions, but we assumed that the object is a sphere. By making this assumption, the observer can estimate the slant and tilt of each point in the image based on the coordinates of the point in the image. The luminance maps are of course less regular than in the upper row, but they still contain the same general pattern. Is there sufficient information to make a reasonably accurate estimate of light direction? We examined this by using the least-squares approach summarized by Equation 3.4.

We ran the analysis for each of the test stimuli from Condition C using two assumed sphere sizes. In the first analysis, the radius of the assumed sphere was equal to the average radius of the test shapes. The average angular error (angular difference between the estimated and true lighting directions) was 15.8° ($SD = 10.4°$). The blue stars represent the estimates. In the second analysis, the radius of the assumed sphere was equal to the radius of the aperture. This assumed shape is less consistent with the true 3D surface geometry, and the errors of the resulting estimates were slightly higher. The average angular error was 17.4° ($SD = 10.5°$). We also ran the analysis using stimuli rendered with a constant ambient light term. The ambient term changes the luminance values in the image, but the overall pattern remains the same and the resulting estimates were similar to the previous analyses.

We observed that the least-squares estimates were reasonably similar to observers settings for most of the stimuli. (The left panel represents one of the most accurate cases and the right panel one of the least accurate.) The similarity shows that human viewers could use a shape assumption in this case, an assumption of sphericity – to estimate light direction reasonably accurately when the actual stimulus is only globally consistent with the assumed shape.

It is also important to consider the illumination information when the stimulus is globally at, as it was in Condition D of our experiments. If the stimulus is a plane,

the luminance map is at: The angle between surface normals and a distant light source is constant, so the luminance is constant (Equation 3.4). Therefore, even a correct assumption about object shape (globally or locally correct) would not allow the observer to estimate light direction. Thus, an observer prior for lighting direction should dictate their responses as we observed in Condition D.

3.5.5 Applications

Our findings have implications for the construction of shaded images. In the absence of robust shape cues, such as texture, disparity, convexity, and familiar shape, our observers misestimated the direction of lighting with shaded images. When only shading was available, the Bayesian prior for light directionVslant of roughly 30° and tilt of slightly more than 90° (i.e., above and slightly to the left)Vdictated the estimate. Because of the interaction between light direction and 3D shape (Equation 3.1), misestimating the light direction can lead to misestimating the shape. Thus, to assure reasonably accurate perception of the 3D shape of shaded images, it is important to place directional lights near the prior. Such placement is relevant to recent work on automatic lighting design [17, 35, 57], and on non-photorealistic rendering techniques designed to affect the perception of 3D shape [55].

3.6 Conclusion

In Experiment 1, we found that observers use a shapebased rather than image-based approach to estimate the lighting direction of a scene. Analyses of the information contained in shading reveals that the lighting direction could in principle be correctly inferred if the reectance properties of the surface material are known and the 3D shape of the object generating the image is known. Our results from Experiment 2 confirm this expectation. We found that observers can match the lighting directions on two objects when the shapes of the objects are well specified. We found that they set the lighting direction quite inaccurately when the shape of one of the objects is specified by shading only; instead they set the lighting direction consistent with a light-from-above prior. Thus, shading alone does not provide sufficient shape information to estimate light direction accurately. We also found that global convexity is a very effective cue in determining light direction and that this finding is expected when one considers the information contained in a globally convex object. Given our results, algorithms for producing images with shaded objects should be consistent with viewer assumptions that the light comes from above and that most objects are globally convex.

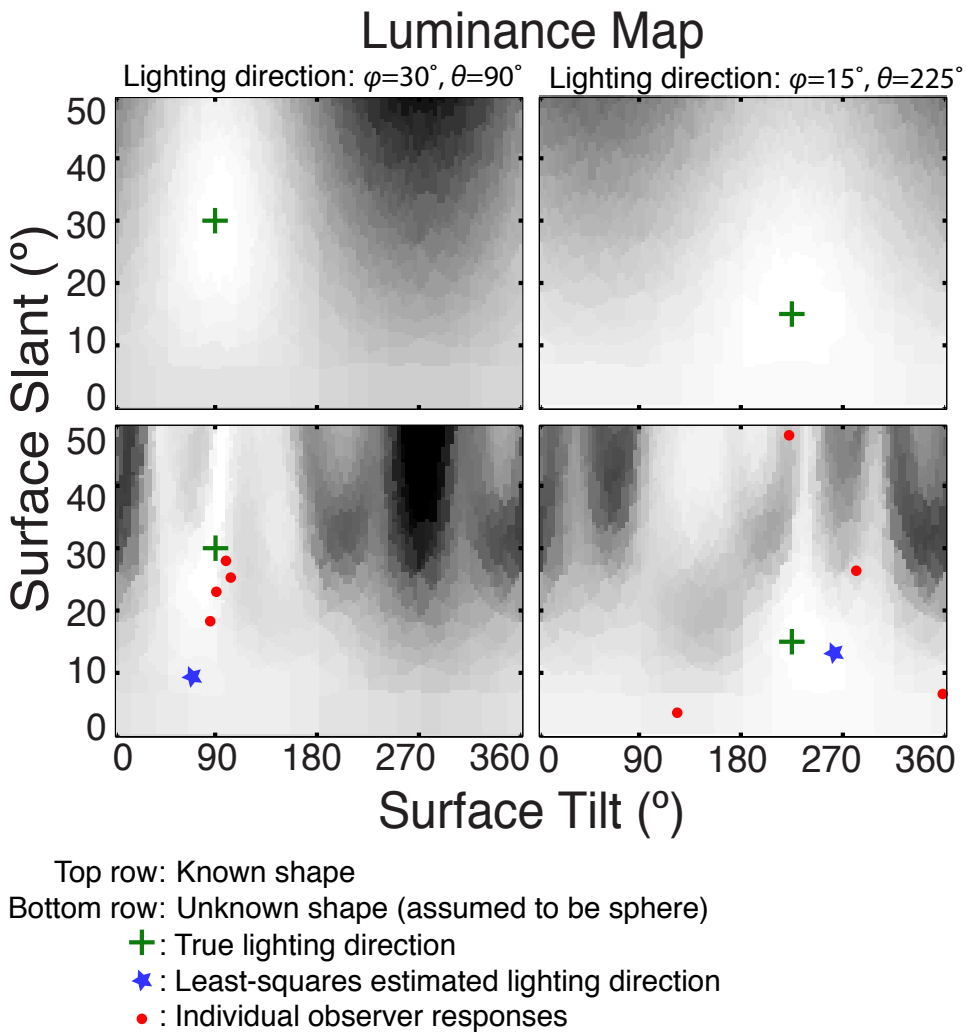

Figure 3.16. . *Luminance maps for estimated surface orientation of stimuli. The upper row shows the luminance maps for a stimulus in Condition C under the (unlikely) assumption that the 3D surface orientation was estimated accurately. The lighting direction was [$\phi = 30°, \theta = 90°$] on the left and [$\phi = 15°, \theta = 225°$] on the right. The green crosses indicate the true lighting direction. The lower row shows the luminance maps for the same stimuli under the assumption that the assumed shape is a sphere. Now the luminance map is distorted because the assumed surface orientation at each stimulus point is not necessarily correct. The green crosses again represent the true lighting directions. The blue stars represent the lighting directions estimated by the least-squares algorithm described in the Introduction section. Although the assumed shape is incorrect, the lighting-direction estimate is within 20° of the true direction. The red dots represent the settings from each of the four observers when presented these stimuli.*

Chapter 4

Material

4.1 Introduction

People encounter objects composed of different materials everyday. The image of such an object, whether on the retina or on a 2D image plane, is determined by the illumination, the 3D shape of the object, and the material properties. Although solving for the material properties is ill-posed (Chapter 3), most people find it trivial to distinquish different material types (e.g. platic, metal, wood), even under unknown illumination [16].

Gloss is a fundamental characteristic of material reflectance. Generally, the more light that is specularly reflected by the material (i.e. when the angle of reflectance is equal to the angle of incidence), the glossier the material appears. Thus, a material is more likely to give a glossy appearance when the surface is very smooth. Gloss can often be seen on different types of plastic, automative paints, and ceramic dinnerware. Taken to the extreme, a perfectly smooth surface would be a mirror.

At the opposite end of the spectrum are diffuse surfaces, which can often be found in the form of paper, unpolished wood, and many types of interior wall paints. Rather than reflecting most of the light in the mirror-reflection direction, the light is diffusely spread equally in all directions. Their appearance is not affected by viewing position. In this chapter, we discuss how disparity and defocus information can provide cues to distinguish glossy and matte surfaces. We describe the types of images which would be needed to study this question, and introduce a technique to produce them.

4.1.1 Material reflectance properties

The way different materials reflect light can be complex. Even seemingly simple materials such as paint can have multiple layers which scatter the light in complicated ways [43]. Material reflectance is typically described using the Bi-directional Reflectance Distribution Function (BRDF). Ignoring the effect of varying wavelength,

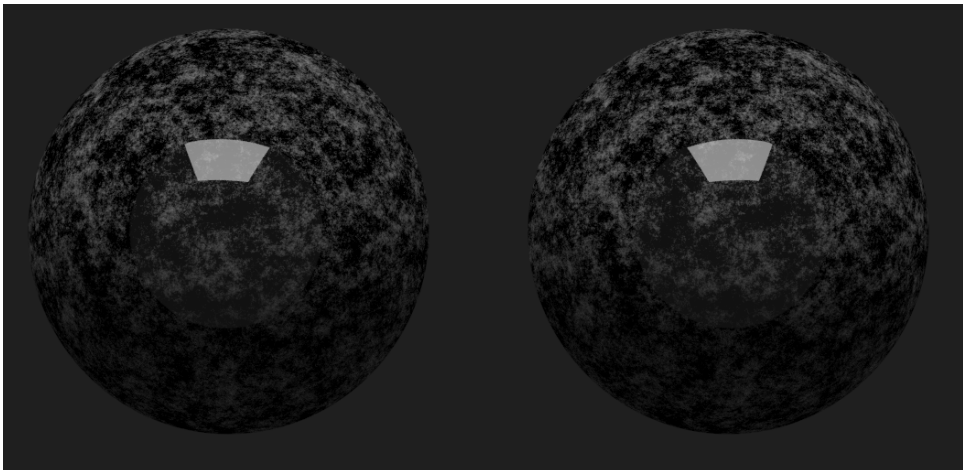

Figure 4.1. *Right eye's (left image) and left eye's (right image) view of a textured sphere with a highly specular surface (cross-fuse to view the image in stereo). The sphere is illuminated by an area light source above and behind the camera. The virtual image of the light source appears within the sphere, and is thus at a different depth from the camera. Because disparity is such a reliable depth cue, it is relatively easy to distinguish the reflected image of the light source from the textured surface of the sphere in stereo.*

the BRDF describes reflectance as a function of two light directions $f_r(\mathbf{i}, \mathbf{o})$, where $\mathbf{i}$ is the incoming light direction and $\mathbf{o}$ is the outgoing light direction. The BRDF is thus a description of how light reflects off of the surface of a particular material.

One of the most studied BRDFs was developed by Ward to provide a physically-based model that accurately describes anisotropic specular reflections. (Ward 1992):

$$f_r(\mathbf{i}, \mathbf{o}) = \frac{\rho_d}{4\pi} + \frac{\rho_s}{4\pi\alpha^2\sqrt{cos\theta_i cos\theta_o}} e^{-\frac{tan^2\theta_h}{\alpha^2}}. \tag{4.1}$$

In this model, the variables θ_i and θ_o are the angles of the incident and outgoing light directions with respect to the surface normal, and θ_h is the half angle between these directions [61]. The ρ_d and ρ_s terms control the proportion of total light that is reflected diffusely and specularly, respectively. The α term determines the roughness of the surface, and thus the spread of specular reflection. Small α values model very smooth and glossy surfaces, while large α values yield diffuse reflections like those from matte surfaces. Although the Ward model is capable of describing anisotropic specular reflections, we only consider the isotropic case.

4.2 Material perception

Recent research has identified different sources of image information which the visual system may use to distinguish glossy and matte materials. Fleming et al. examined

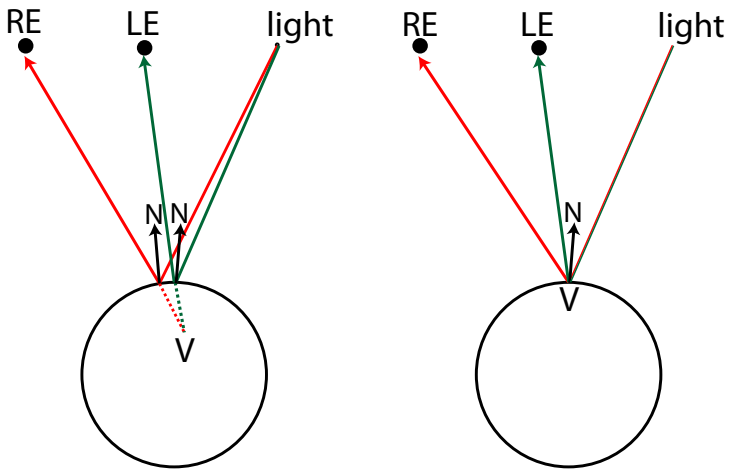

Figure 4.2. *(Left) The geometry of a specular reflection of a point light source on a sphere viewed in stereo. The virtual image (V) of the light source is located at the intersection of the reflection light rays which reflect off of the surface and intersect with left eye (LE) and right eye (RE). (Right) Diagram showing how the specular component of the BRDF can be manipulated to change the disparity of the virtual image. The specular reflections for the right eye uses the reflection angle based on the left eye's position. The same ray from the light source will now intersect each eye, which has the effect of moving the virtual image to the sphere's surface.*

how the statistics of glossy reflections can be used to distinguish specular reflections from matte textures [15]. Similarly, the statistics of the luminance histogram may provide useful cues to distinguish these material types. Glossy materials tend to have high-contrast specularities which skew the histogram of the luminance in a characteristic way [45, 2]. The dynamic range of the image may also reveal materal properties, primarily because specularities under natural illumination tend to have higher luminance magnitudes [11].

Under binocular viewing conditions, disparity has recently been shown to provide useful information to perceive glossy surfaces [56, 64]. This is because the virtual image of a glossy reflection is usually at a different depth than the object's surface [5]. Because disparity and defocus are geometrically similar [18], defocus may also provide enough information to distinguish glossy and matte materials. In this chapter, we explore the methods which would be required to examine this question, as well as the techniques used to generate the stimuli for the proposed experiment.

4.2.1 Disparity cue

Consider a polished sphere composed of a textured material such as marble, and polished such that it is smooth and reflects light with sharp specularities (Figure 4.1). In other words, the material of the sphere has both a diffuse component and a specular component. The sphere is illuminated by a single area light source above

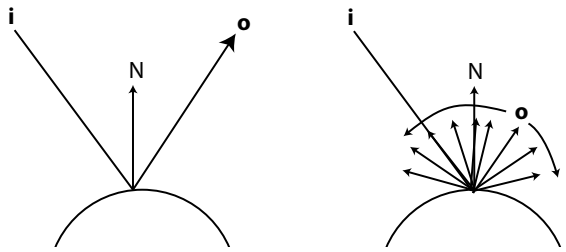

Figure 4.3. *The geometry of BRDF reflectance. (Left) For a specular surface, all or most of the incoming light i reflects off of the surface according to the mirror reflection angle. The angle between i and the surface normal N is equal to the angle between o and the N. (Right). For a diffuse material, the incoming lighti i is diffusely reflected equally in all directions.*

and behind the viewer. The light source is reflected by the surface of the sphere and appears as a slightly distorted (due to the curvature of the sphere) white square in the image. As previously mentioned, the image of this sphere is determined by 1) the 3D shape of the object, 2) the material of which it is composed, and 3) the illumination. When viewed monocularly and without additional information, solving for the material properties of the sphere is ill-posed and it is impossible to determine if the surface is glossy or not.

Consider the same scene and assume the 3D shape of the object and the illumination are both known. Although two of the three scene properties are known, solving for material is still ill-posed because it is still impossible to disambiguate the diffuse component from the specular component. The white specularity could just as easily be a mark of white paint sitting on the surface of the sphere or it could be the reflection of a light source. If viewed in stereo however, the virtual image of the reflection will be at a different distance than the surface of the sphere (Figure 4.2).

Under these conditions, identifying whether the sphere has a glossy surface now becomes a matter of detecting differences in the relative disparities of the image content. The depth of the specularity in Figure 4.1 is slightly behind the textured surface of the sphere. Because disparity is such a strong cue to depth, the difference in depth between the specular reflection and the textured surface of the sphere is easy to perceive. This cue is strong enough to affect an observer's perception of material gloss [5, 56, 64]. Figure 4.4 provides an example of how disparity can provide useful information to distinguish glossy and matte surfaces.

4.2.2 Defocus cue

Disparity can be used to precisely solve for the absolute distance of objects when the interocular and focal distances are known. Held et al. showed that defocus is geometrically the same as disparity [18]. Whereas disparity is created by differences in the positions of the two eyes, defocus blur is created by differences in the positions

Figure 4.4. *Physically-correct (top) and manipulated (bottom) specular reflections for a sphere illuminated by an environment map. Natural illumination can affect material perception [16], and the top image should provide a strong percept of gloss. Even when viewed monocularly, the dynamic range and statistics of the image suggest a glossy appearance. Viewed in stereo, the bottom image should appear matte because there is no dispartiy difference between the textured surface of the sphere and the specular reflection content. It should appear as if the scene has been painted on to the surface. Our method to generate these types of images accurately manipulates the disparity cue without affect the other image cues to gloss perception.*

of the light rays as they pass through the aperture of the eye. Using the small-angle approximation, the equations describing disparity δ and blur β in radians can be rewritten as follows:

$$\delta \approx I \left(\frac{1}{z_o} - \frac{1}{z_1} \right) \tag{4.2}$$

$$\beta \approx A \left(\frac{1}{z_o} - \frac{1}{z_1} \right), \tag{4.3}$$

where I is the interocular distance, A is the pupil diameter, z_0 is the focal distance, and z_1 is the distance to another point in the scene. In other words, the two sources of

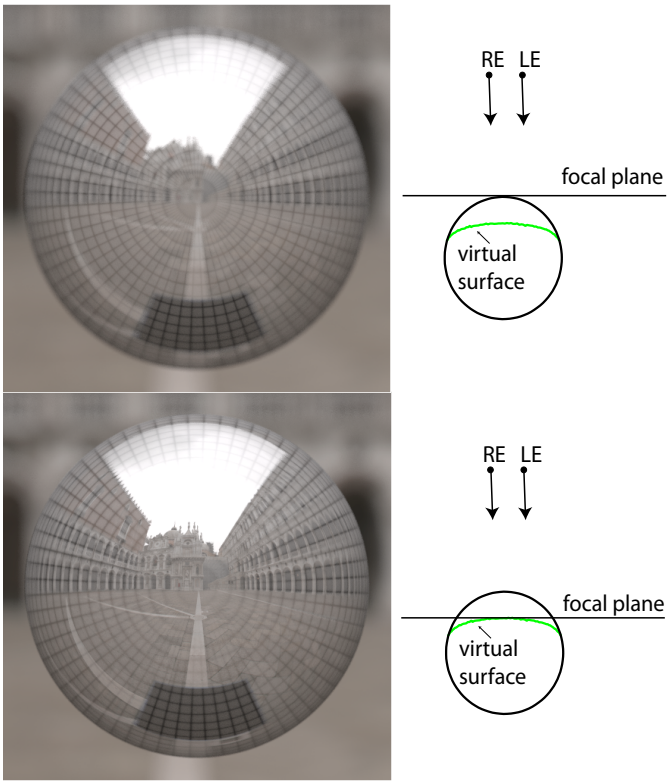

Figure 4.5. *Defocus as a cue to material perception. The left side of the figure shows the computer-generated images of a chrome sphere with a grid texture on the surface. The images are rendered with physically correct blur. The diagrams on the right side show the position of the focal plane in a top-down view of the rendered scene for each image. In the top image, the camera is focused at the surface of the sphere. Because the virtual image of the reflection is positioned within the sphere, the reflection appears slightly blurry. In the bottom image, the focal plane is now approximately at the distance of the virtual image. In this image, the reflection appears sharp while the grid pattern on the surface of the sphere is blurry. This difference in blur may provide useful information to distinguish glossy and matte materials.*

information share the same geometric constraints. Thus, defocus may provide useful information to distinguish glossy surfaces from matte surfaces as long as the difference in blur between the surface and the virtual image is detectable by the observer. Figure 4.5 illustrates how blur can provide useful information to distinguish glossy surfaces from matte surfaces.

4.3 Experimental stimuli

As previously mentioned, there are several different sources of information which may be cues to identify whether a material is glossy or matte. Our goal is to generate im-

ages which allow us to independently manipulate the disparity or defocus cue without affecting the other properties such as the dynamic range or image statistics.

4.3.1 Physically-based rendering

Physically-based rendering techniques synthesize the image of a light-field for a specified scene by simulating the physics of light transport. One method to create these images is to trace light rays in a scene and calculate the rays which intersect with the camera's center of projections. Due to Helmholz's law of reciprocity, it is typically more efficient to trace a ray from the camera's center of projection through the image plane into the scene. If the light ray intersects a surface, the corresponding point in the image is simply the sum of the light incident to the surface point for all possible incoming directions (Figure 4.3).

Because we model material reflectance as the sum of the diffuse and specular component (Equation 4.1), these calculations can be made independently. For physically-correct reflections, both the diffuse component and the specular component intersect with the centers of projection. This is true in stereo images as well, although the virtual image of the reflections may be at different distances than the surface of the object (Figure 4.2, Left). The top image in Figure 4.6 shows a sphere rendered with phyisically-correct reflections.

4.3.2 Manipulating disparity

To control the cue to material properties from disparity or defocus, it is necessary to manipulate the specular component of the BRDF for the material. As described in Equation 4.1, the specular component ρ_s is determined by the outgoing light direction $\mathbf{o}$ which intersects with the center of projection. In stereo, this component yields a virtual image at a different depth than the object's surface (Figure 4.2). As previously noted, we expect this difference in depth to effect perceived gloss. We expect images in which there is zero change in disparity or defocus across the surface of the object to be perceived as matte. We expect observers to perceive images in which there are non-zero changes in the disparity or defocus across the surface as glossier.

We manipulate the distances to this virtual image by adjusting the outgoing light direction $\mathbf{o}$ for the specular component and moving it away from the center of projection for the respective eye. In stereo, this has the effect of changing the disparity of the specular reflection (Figure 4.2). By moving the incoming light direction towards the other eye's center of projection, we move the virtual image of the reflection closer to the surface of the sphere. Figure 4.6 shows an example stereo image pair in which the reflection component of the BRDF was manipulated such that the left eye's reflection component was rendered from the right eye's image center of projection.

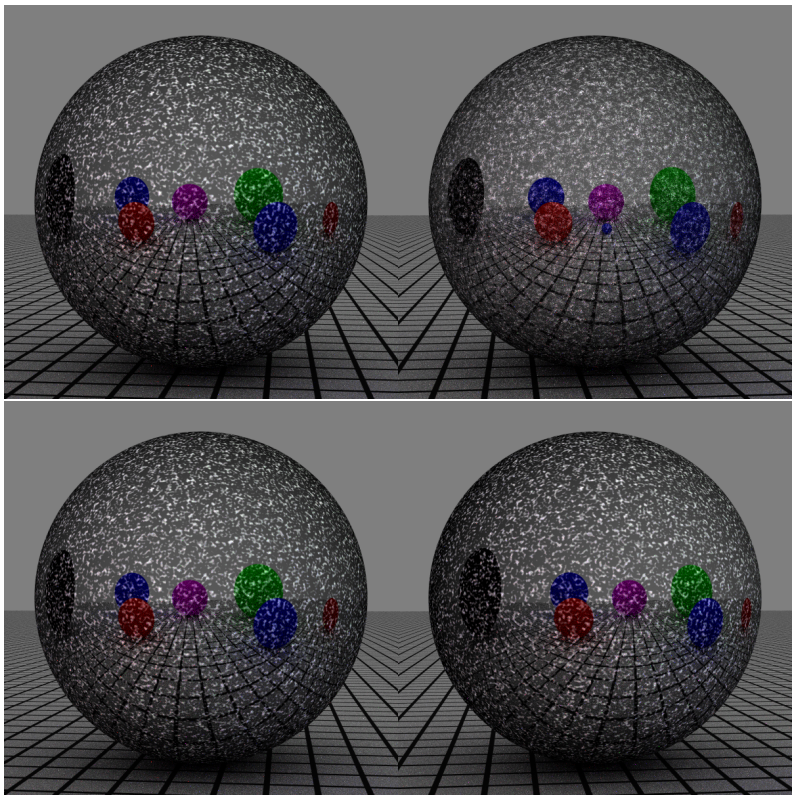

Figure 4.6. *Stereo view of sphere rendered with a physically-correct (top) and incorrect (bottom) reflections. The sphere has both a diffuse component (marble-like texture) and a specular component which reflects the incoming light. In the top image, the virtual image of the reflected colored appears behind the surface of the reflecting sphere. In the bottom image, the reflections for each eye's image were calculated using the mirror-reflected rays traced from the opposite eye's position. This manipulation does not affect the image statistics or dynamic range of the image, but the virtual image of the reflections now appear conincident with the surface of the sphere. We expect the sphere in the top image to appear glossy while the bottom sphere should appear matte.*

4.3.3 Manipulating defocus

To this point, we have mainly focused on how to manipulate the disparity information of the specular reflection in order to affect perceived gloss. Because the geometry of defocus is the same, the rendering process is similar as well. Instead of calculating rays for two different eyes, we simply calculate rays which sample the area of an aperture. By modifying the reflection angle of the rays depending on their position in the aperature, we affect the defocus of the specular component similar to how we manipulated the disparity component.

Our method for manipulating the specular reflections effectively isolates the effect of disparity or defocus from other cues which may contribute to material perception.

In particular, the statistics and dynamic range of the image remain nearly identical between the physically-correct renderings and the manipulated renderings designed to look matte. This effect is more compelling in Figure 4.4 in which a natural illumination map is used for the lighting. Natural illumination has been shown to provide a stronger gloss percept, and some research suggests it may be due to the statistical properties of reflections compared to matte textures [16].

4.4 Conclusion

Solving for any of the basic scene properties which determine an image – 3D shape, lighting, and material properties – is an ill-posed problem which can only be solved with additional information. The visual system is able to make use of different sources of information to try to recover these properties. Disparity provides useful information to distinguish when materials have a glossy We have explained how disparity may provide information to distinguish glossy and diffuse materials. Recent research suggests that defocus blur should provide similar information to observers, and we have described the types of images which would allow us to probe this research topic experimentally. We have developed a technique to create this type of content and have provided example images.

Bibliography

[1] Wendy Adams, Erich Graf, and Marc Ernst. Experience can change the light-from-above prior. *Nature Neuroscience*, 7(10):1057–1058, 2004.

[2] B Anderson and J Kim. Image statistics do not explain the perception of gloss and lightness. *Journal of Vision*, Jan 2009.

[3] B. T. Backus, M. S. Banks, R. van Ee, and J. A. Crowell. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39: 1143–1170, 1999.

[4] Peter Belhumeur, David Kriegman, and Alan Yuille. The bas-relief ambiguity. *IEEE Computer Science Society Conference on Computer Vision and Pattern Recognition*, pages 1060–1066, 1997.

[5] A Blake and Heinrich H Bulthoff. Does the brain know the physics of specular reflection? *Nature*, Jan 1990.

[6] Franck Caniard and Roland Fleming. Distortion in 3d shape estimation with changes in illumination. *Proceedings of the 4th symposium on Applied perception in graphics and visualization*, pages 99–105, 2007.

[7] E. Catmull and J. Clark. Recursively generated b-spline surfaces on arbitrary topological meshes. *Computer-Aided Design*, 10:350–355, 1978.

[8] Rebecca A. Champion and Wendy J. Adams. Modification of the convexity prior but not the light-from-above prior in visual search with shaded objects. *Journal of Vision*, 7(13):1–10, 10 2007.

[9] Robert L. Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1):7–24, 1982.

[10] D. DeCarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella. Suggestive contours for conveying shape. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, pages 848–855, 2003.

[11] Katja Doerschner. Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *Journal of Vision*, 10(4):1–9, Jan 2010.

[12] D. Dunbar and G. Humphreys. A spatial data structure for fast poisson-disk sample generation. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2006)*, 25:503–508, 2006.

[13] J. T. Enns and R. A. Rensink. Inuence of scenebased properties on visual search. *Science*, 247:721723, 1990.

[14] H. Farid and M. J. Bravo. Image forensic analyses that elude the human visual system. *Proceedings of the SPIE*, 7541:1–10, 2010.

[15] Roland W Fleming. Specular reflections and the perception of shape. *Journal of Vision*, 4(9):798–820, Jan 2004. doi: 10.1167/4.9.

[16] Roland W Fleming, Ron O Dror, and Edward H Adelson. Real-world illumination and the perception of surface reflectance properties. *J. Vis.*, 3(5):347–368, 2003.

[17] Stefan Gumhold. Maximum entropy light source placement. In *VIS '02: Proceedings of the conference on Visualization '02*, Washington, DC, USA, 2002. IEEE Computer Society. ISBN 0-7803-7498-3.

[18] Robert T. Held, Emily A. Cooper, and Martin S. Banks. Blur and disparity are complementar cues to depth. *Current Biology*, 22(5):426–431, 2012.

[19] James Hillis, Simon Watt, Michael Landy, and Martin Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12):967–992, December 2004.

[20] Y. X. Ho, M. S. Landy, and L. T. Maloney. How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, 6(5)(8):634–648, 2006.

[21] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):1–30, 3 2008. ISSN 1534-7362.

[22] K. Ikeuchi and B. K. P. Horn. Numerical shape from shading and occluding boundaries. *Articial Intelligence*, 17:141–184, 1981.

[23] H. L. Jenkin, M. R. Jenkin, R. T. Dyde, and L. R. Harris. Shape-from-shading depends on visual, gravitational, and body-orientation cues. *Perception*, 33: 1453–1461, 2004.

[24] M. K. Johnson and H. Farid. Exposing digital forgeries by detecting inconsistencies in lighting. *Proceedings of the 7th Workshop on Multimedia and Security*, pages 1–10, August 2005.

[25] I. T. Jolliffe. *Principal component analysis*, 2002.

[26] J. Kajiya. The rendering equation. *Computer Graphics*, 20:143–149, 1986.

[27] D. Kersten, P. Mamassian, and A. Yuille. Object perception as bayesian inference. *Annual Reviews of Psychology*, 55:271–304, 2004.

[28] D. A. Kleffner and V. S V. S. Ramachandran. On the perception of shape from shading. *Perception and Psychophysics*, 52:18–36, 1992.

[29] J. J. Koenderink, A. J. van Doorn, and S. C. Pont. Light direction from shad(ow)ed random gaussian surfaces. *Perception*, 33:1405–1420, 2004.

[30] J. J. Koenderink, S. C. Pont, A. J. van Doorn, A. M. L. Kappers, and J. T. Tood. The visual light field. *Perception*, 36:1595–1610, 2007.

[31] Jan Koenderink, Andrea van Doorn, and Astrid Kappers. Surface perception in pictures. *Perception and Psychophysics*, 5(52):487–496, 1992.

[32] Jan Koenderink, Andrea van Doorn, and Chris Christou. Perturbation study of shading in pictures. *Perception*, 25:1009–1026, 1996.

[33] Michael Langer and Heinrich Bulthoff. A prior for global convexity in local shape-from-shading. *Perception*, 30:403–410, 2001.

[34] C. H. Lee and A. Rosenfeld. Improved methods for estimating shape from shading using the light source coordinate system. *Articial Intelligence*, 26:125–143, 1985.

[35] Chang Ha Lee and Xuejun Hao. Geometry-dependent lighting. *IEEE Transactions on Visualization and Computer Graphics*, 12(2):197–207, 2006. ISSN 1077-2626. doi: http://dx.doi.org/10.1109/TVCG.2006.30. Member-Amitabh Varshney.

[36] P. Leong and S. Carlile. Methods for spherical data analysis and visualization. *Journal of Neuroscience Methods*, 80:191–200, 1998.

[37] J. Lopez-Moreno, S. Hadap, E. Reinhard, and D. Gutierrez. Light source detection in photographs. *Congreso Espanol de Informatica Graca*, pages 161–168, Sept 2009.

[38] J. Malik and D. Maydan. Recovering threedimensional shape from a single image of curved objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:555–566, 1989.

[39] P. Mamassian. Impossible shadows and the shadow correspondence problem. *Perception*, 33:1279–1290, 2004.

[40] P. Mamassian and M. S. Landy. Observer biases in the 3d interpretation of line drawings. *Vision Research*, 38:2817–2832, 1998.

[41] Pascal Mamassian and Ross Goutcher. Prior knowledge on the illumination position. *Cognition*, 81:B1–B9, 2001.

[42] Pascal Mamassian and Daniel Kersten. Illumination, shading and the perception of local orientation. *Vision Research*, 36(15):2351–2367, 1995.

[43] Stephen R. Marschner, Stephen H. Westin, Eric P. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. Image-based brdf measurement including human skin. In Dani Lischinski and Gregory Ward Larson, editors, *Rendering Techniques*, pages 131–144. Springer, 1999.

[44] Y. Morgenstern and R. Murray. Contextual lighting cues can override the light-from-above prior [abstract]. *Journal of Vision*, 9(8):65, 2009.

[45] Isamu Motoyoshi, Shin'ya Nishida, Lavanya Sharan, and Edward H Adelson. Image statistics and the perception of surface qualities. *Nature*, 447(7141):206–209, May 2007.

[46] P. Nillius and J. O. Eklundh. Automatic estimation of the projected light source direction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:1076–1083, 2001.

[47] John Oliensis. Shape from shading as a partially well-constrained problem. *Computer Vision, Graphics, and Image Processing*, 54(2):163–183, 1991.

[48] Michael Oren and Shree K. Nayar. Generalization of lambert's reflectance model. In *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 239–246, New York, NY, USA, 1994. ACM. ISBN 0-89791-667-0. doi: http://doi.acm.org/10.1145/192161.192213.

[49] Y. Ostrovsky, P. Cavanagh, and P. Sinha. Perceiving illumination inconsistencies in scenes. *Perception*, 34:1301–1314, 2005.

[50] J. P. OShea, M. S. Banks, and M. Agrawala. The assumed light direction for perceiving shape from shading. *In Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, pages 135–142, August 2008.

[51] A. P. Pentland. Finding the illuminant direction. *Journal of the Optical Society of America*, 72:448–455, 1982.

[52] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[53] S. C. Pont and J. J. Koenderink. Matching illumination of solid objects. *Perception and Psychophysics*, 69:459468, 2007.

[54] E. Prados and O. Faugeras. Shape From Shading: a well-posed problem? In *Proceedings of CVPR'05*, volume 2, pages 870–877, june 2005.

[55] Szymon Rusinkiewicz, Michael Burns, and Doug DeCarlo. Exaggerated shading for depicting shape and detail. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 25(3), July 2006.

[56] Yuichi Sakano and Hiroshi Ando. Effects of head motion and stereo viewing on perceived glossiness. *J Vis*, 10(9):15, Jan 2010.

[57] Ram Shacked and Dani Lischinski. Automatic lighting design using a perceptual quality metric. In *EG 2001 Proceedings*, volume 20(3), pages 215–226. 2001.

[58] Kent A. Stevens. Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, 46:183–195, 1983.

[59] Jennifer Sun and Pietro Perona. Where is the sun? *Nature Neuroscience*, 1(3): 183–184, July 1998.

[60] Kenneth E. Torrance and E. M. Sparrow. Theory for the off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1144, 1967. Article.

[61] Bruce Walter. Notes on the ward brdf. *Cornell Technical Report*, 2005.

[62] Gregory J. Ward. Measuring and modeling anisotropic reflection. *SIGGRAPH Comput. Graph.*, 26(2):265–272, 1992. ISSN 0097-8930. doi: http://doi.acm.org/10.1145/142920.134078.

[63] Simon J. Watt, Kurt Akeley, Marc O. Ernst, and Martin S. Banks. Focus cues affect perceived depth. *Journal of Vision*, 5(10):834–862, December 2005.

[64] Gunnar Wendt, Franz Faul, Vebjørn Ekroll, and Rainer Mausfeld. Disparity, motion, and color information improve gloss constancy performance. *J Vis*, 10 (9):7, Jan 2010.