

Specular reflections and the perception of shape

Roland W. Fleming

Department of Brain and Cognitive Sciences, MIT,
Cambridge, MA, USA, & Max Planck Institute for
Biological Cybernetics, Tübingen, Germany



Antonio Torralba

Computer Science and Artificial Intelligence Laboratories,
MIT, Cambridge, MA, USA



Edward H. Adelson

Department of Brain and Cognitive Sciences, MIT,
Cambridge, MA, USA



Many materials, including leaves, water, plastic, and chrome exhibit specular reflections. It seems reasonable that the visual system can somehow exploit specular reflections to recover three-dimensional (3D) shape. Previous studies (e.g., J. T. Todd & E. Mingolla, 1983; J. F. Norman, J. T. Todd, & G. A. Orban, 2004) have shown that specular reflections aid shape estimation, but the relevant image information has not yet been isolated. Here we explain how specular reflections can provide reliable and accurate constraints on 3D shape. We argue that the visual system can treat specularities somewhat like textures, by using the systematic patterns of distortion across the image of a specular surface to recover 3D shape. However, there is a crucial difference between textures and specularities: In the case of textures, the image compressions depend on the first derivative of the surface depth (i.e., surface orientation), whereas in the case of specularities, the image compressions depend on the second derivative (i.e., surfaces curvatures). We suggest that this difference provides a cue that can help the visual system distinguish between textures and specularities, even when present simultaneously. More importantly, we show that the dependency of specular distortions on the second derivative of the surface leads to distinctive fields of image orientation as the reflected world is warped across the surface. We find that these "orientation fields" are (i) diagnostic of 3D shape, (ii) remain surprisingly stable when the world reflected in the surface is changed, and (iii) can be extracted from the image by populations of simple oriented filters. Thus the use of specular reflections for 3D shape perception is both easier and more reliable than previous computational work would suggest.

Keywords: 3D shape perception, specularity, shape-from-texture, shape-from-shading, gloss, natural image statistics, material perception, modeling

Introduction

Figure 1 shows a computer-generated image of a perfectly polished mirror. Most observers agree that they have a vivid impression of the object's three-dimensional (3D) shape. This is surprising given that many of the cues that are traditionally thought to be important for shape perception are absent from the stimulus. Specifically,

1. The image is stationary and thus there are no cues to shape from motion.
2. There is only a single image, and thus there is no consistent information from binocular stereopsis (because the disparity field is uniform).
3. The object has been rendered as a perfectly smooth surface with uniform reflectance and thus there are no scratches, pigmentations, or other markings attached to the surface that could provide shape-from-texture information.
4. The image contains no shading in the traditional sense of the word, (i.e., smoothly graded variations

in intensity arising from a Lambertian surface) because the surface is a mirror that is riddled with specular highlights.



Figure 1. A computer-generated image of a perfectly mirrored (specular) surface. Most observers report having a vivid impression of the object's 3D shape, even though the image contains no motion, stereo, texture, or shading. Indeed, the image consists of nothing more than a distorted reflection of the world surrounding the object, and yet somehow we can interpret these patterns to recover the 3D shape.

Indeed, when we look at the image, all that we see is a distorted reflection of the scene surrounding the object, and yet somehow we are able to interpret these warped patterns to recover the 3D shape. How do we do this? What information is present in a single static image that allows us to perform this task? What assumptions does the visual system have to make?

The apparent difficulty of interpreting specular reflections

At first sight, our ability to estimate an object's 3D shape from the reflections in its surface is quite baffling. Reflections are extremely unstable. Unlike texture markings or shadows, specularities slide over the surface and change shape whenever the object, viewer, or environment moves. A feature in the surrounding scene, such as a building or tree, is generally warped into a complex irregular shape when reflected in a specular surface. This makes it extremely difficult to locate, track, and interpret reflections of even quite simple environmental features.

Furthermore, in the case of a perfect mirror, the image consists of nothing more than a distorted reflection of the world surrounding the object. Thus, a specular object, such as a polished kettle, produces a different image every time it is placed in a different scene. Put another way, specular surfaces inherit their appearance solely from their environment: Every visible feature belongs to the world surrounding the object rather than to the object itself. Thus, as the object is moved from scene to scene, the image changes dramatically. Despite this, the 3D shape appears quite stable, as shown in [Figure 2](#).

To make matters worse, because the image is just a reflection of the world, it is possible to produce almost any arbitrary image from a mirrored surface by carefully manipulating the environment surrounding the object. Thus a perfectly smooth object could be made to appear to have dents or bumps simply by distorting the scene, and the visual system would have no way of knowing that it is the environment rather than the shape that is responsible, because the image data would be identical. Consequently,

many possible combinations of shape and scene are consistent with a given image ([Figure 3](#)), and yet somehow the visual system must reject the infinite false interpretations to recover the one correct shape.

Thus, mathematically speaking, the task of recovering an object's shape from the image reflected in its surface is hopelessly ill posed, and surely a difficult perceptual inference. Indeed, it has even been suggested that it might not be possible to solve this problem at all for single static images (Oren & Nayer, [1996](#)) and that humans are poor at it (Savarese, Li, & Perona, [in press](#)), although we show here that they are not. Despite this, previous psychophysical research has shown that specular reflections generally improve human shape estimation (Blake & Bülthoff, [1990, 1991](#); Mingolla & Todd, [1986](#); Norman, Todd, & Orban, [2004](#); Todd & Mingolla, [1983](#); Todd, Norman, Koenderink, & Kappers, [1997](#)), although the relevant image information has yet to be identified. How does the visual system use specular reflections when they depend so much on the world surrounding the object? In what way do specular reflections constrain shape? How can the relevant information be extracted from the image?

An alternative way of posing the problem

In this work, we argue that the apparent difficulty of interpreting specular reflections is deceptive, and that it is possible to re-pose the problem in terms of simple image measurements that are diagnostic of shape but which remain relatively stable across changes in the environment. We argue that the interpretability of specular reflections depends on the particular way in which we conceive of the patterns reflected in the object's surface. By reformulating the role of the surrounding world, we show that it is possible to treat specularities somewhat like surface texture, and thus to recover shape from specular reflections by analogy to the recovery of shape from texture.

To make this clear, we will now contrast two ways of representing the scene. First, let us consider the surrounding environment as a complex physical world composed of discrete recognizable objects, such as buildings or trees. To



Figure 2. The image of a mirrored object is simply a reflection of the world surrounding the object. Thus the image changes dramatically when the object is placed in three different scenes.

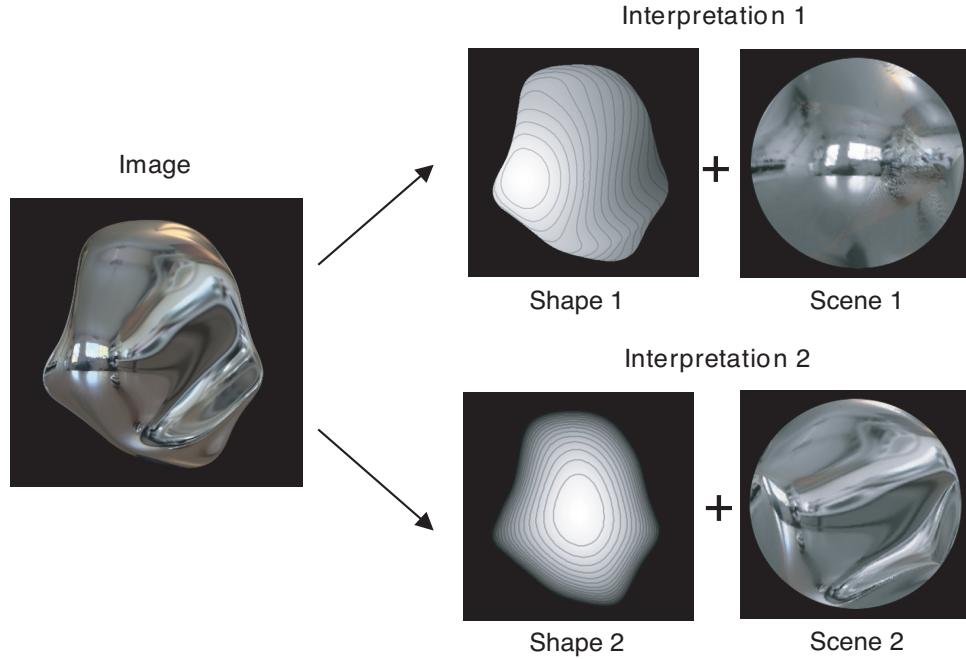


Figure 3. A given image of a mirrored object is consistent with many different shapes. For example, the same image could be created by placing Shape 1 in Scene 1, or by placing Shape 2 in Scene 2.

recover shape from specularities, the visual system would first have to locate and recognize the distorted reflection of a specific environmental object, such as a rectangle that has been warped into an irregular wedge shape. Then, the visual system would have to estimate the deforming transformation that has been applied to the shape of the reflection by the geometry of the surface. In theory, once this transformation is known, the visual system could recover the 3D shape of the surface that is responsible for the distortion. To take a simple example, if the reflection contains a curve while the corresponding environmental feature is actually a straight line, then the visual system can use the degree of 2D curvature in the image to estimate the 3D curvature of the reflecting surface.

Some variation on this reasoning has been the default approach in most previous computational work on the problem (for a review, see Oren & Nayar, 1996). For example, in elegant computational work, Savarese and Perona (2001, 2002) have shown that it is possible to reconstruct the 3D shape of a curved mirror from a single static image when a standard checkerboard pattern is reflected in the surface.

The primary disadvantage of this formulation is that the visual system can only interpret the distorted reflection of an object if it knows what the undistorted object looks like. Thus, this approach requires that the visual system has access to an accurate model of the surrounding scene, or at least makes strong assumptions about the world (e.g., lines are usually straight). However, the human visual system is not normally confronted with such carefully calibrated scenes. It seems quite unlikely that the visual system is capable of building a full model of the environment sur-

rounding an object in a realistic setting. Furthermore, to reconstruct a surface by “inverse optics” is computationally extremely complex. It is not yet clear how such complex computations could be implemented by simple neural mechanisms. We reason, therefore, that there must be a robust alternative strategy that (i) does not require a model of the surrounding environment, and (ii) can be expressed in terms of relatively simple image measurements that can be readily implemented by known biological substrates. The basis of the alternative strategy is to change our conception of the reflected world.

The intuition is as follows. We argue that the world can be treated somewhat like a “texture” whose image statistics (e.g., amplitude spectrum and distribution of orientations) are quite well conserved across scenes. Although the precise locations of physical structures, such as people or trees, change completely from scene to scene, the basic “texture” of the world remains quite stable (Field, 1987; Dror, Leung, Willsky, & Adelson, 2001; Dror, 2002). When this “texture” is reflected in a mirrored surface, it is distorted dramatically in a way that depends crucially on the surface shape. These distortions lead to continuously varying texturelike patterns across the image of the surface, which we call “orientation fields.” We argue that the visual system can recover strong constraints on the 3D shape of the reflecting surface directly from the distorted patterns, much as it can recover 3D shape of a textured surface from the patterns of distorted texture. This way the visual system does not have to interpret the distorted reflections of recognizable objects, and thus there is no need to construct an accurate representation of the scene surrounding the object. We have suggested previously that the visual system

can treat specular reflections somewhat like textures for the purposes of surface reflectance estimation (Fleming, Dror, & Adelson, 2003); here we extend the idea to the estimation of shape from specular reflections.

Before discussing this formulation in detail, we present the results of a basic psychophysical experiment on the estimation of shape from specular reflections. It is now well established that specular reflections aid shape estimation in the presence of other cues, such as shading, texture, and stereo (Blake & Bühlhoff, 1990, 1991; Todd & Mingolla, 1983; Todd et al., 1997). However, to our knowledge, nobody has previously isolated this cue by testing our ability to estimate shape from purely specular surfaces that are reflecting realistic scenes.

We can derive two simple predictions from the idea that the visual system recovers shape directly from the texturelike orientation fields across a specular surface. First, subjects should be able to estimate 3D shape accurately even when they have no additional information about the scene surrounding the object (i.e., when the object is cropped out of its original context and shown against a neutral background). Second, as long as a scene has sufficient structure, the distorted reflection of the scene should produce the characteristic orientation fields across the image. Thus, shape estimation should remain quite good across different realistic scenes. These predictions are supported by the demonstrations in Figures 1 and 2 as the images yield a vivid impression of 3D shape across changes in the reflected scene and in the absence of context. To corroborate this phenomenological evidence, we have conducted a psychophysical shape-estimation task.

Findings I: Psychophysics

To measure human 3D shape estimation, we used the standard “gauge figure” task (Koenderink, van Doorn, & Kappers, 1992; Mamassian & Kersten, 1993, 1996). A screenshot of the task is shown in Figure 4(a). Subjects were presented with computer generated images of irregularly shaped objects with perfectly mirrored surfaces. Their task was to adjust the 3D orientation of a series of gauge figures to create a map of perceived surface normals.

Subjects

Subjects were two naïve observers who were paid for participation, and one of the authors (RF). All subjects had normal or corrected-to-normal vision.

Stimuli

Stimuli consisted of single static images of three irregular shapes. Each shape was rendered in three different real-world scenes, making a total of nine conditions. The rendering was performed using a set of “light probes,” which were captured photographically from locations in the real world (Debevec, 1998; Debevec et al., 2000). Light probes are spherical (360 deg x 180 deg panoramic) images that

capture the set of all rays converging on a point in the world. Rendering an object with a real-world light probe recreates the image that would be acquired if the synthetic object had actually been placed at that location in the world. This allows us to render perfectly specular surfaces that yield highly realistic images.

Stimuli were rendered and tone-mapped for display using RADIANCE (Ward, 1994). The surfaces were represented as triangle meshes of around 8×10^5 polygons. Surface reflectance was set to an ideal mirror (i.e., a specular reflectance gain of 1) with no diffuse reflection, no transmission, and no spread (blur) of the specular component. For the purposes of ray tracing, the light probes were treated as illumination arriving from infinite distance, as described elsewhere (Dror, 2002; Fleming et al., 2003). However, the focal point of the observer was set at finite distance from the object (i.e., perspective rather than orthographic projection). Images were initially rendered at a high resolution of 3072 x 3072 pixels, and down-sampled by a factor of 8 to 384 x 384 pixels to ensure high image quality. The objects were then cropped smoothly out of their original contexts and shown against a black background.

Procedure

Prior to the experimental conditions, subjects practiced the gauge-figure task with an additional stimulus that was a different shape from the experimental stimuli, and which was rendered with texture, diffuse shading, and specular highlights.

Experimental stimuli were presented in three blocks. Each block consisted of all nine conditions in pseudo-random order such that consecutive conditions contained neither the same shape nor the same light probe.

For each condition subjects were presented with two versions of the same image simultaneously [Figure 4(a)]. The left image consisted of an array of all the surface normals that the subject would adjust. Initially this array was set to random 3D orientations at each location. The right image showed a single gauge figure for the surface normal that the subject was currently adjusting. The first normal to be adjusted was picked at random with each new condition in every block. The subject adjusted the 3D orientation via the mouse. The 2D coordinates of the mouse were intuitively mapped into 3D orientation of the normal, so that the subject felt that he or she was controlling the 3D position of the end-point of the gauge figure’s gnomon. Once satisfied with the setting, the subject moved onto the next normal in the array by clicking the mouse. Subjects were allowed to return to and adjust previous normals in the array, although they reported that they generally did not choose to do so as they found they could set the normals satisfactorily at the first pass. Subjects were given unlimited time to perform the task, but took on average between 3 and 4 s per surface normal.

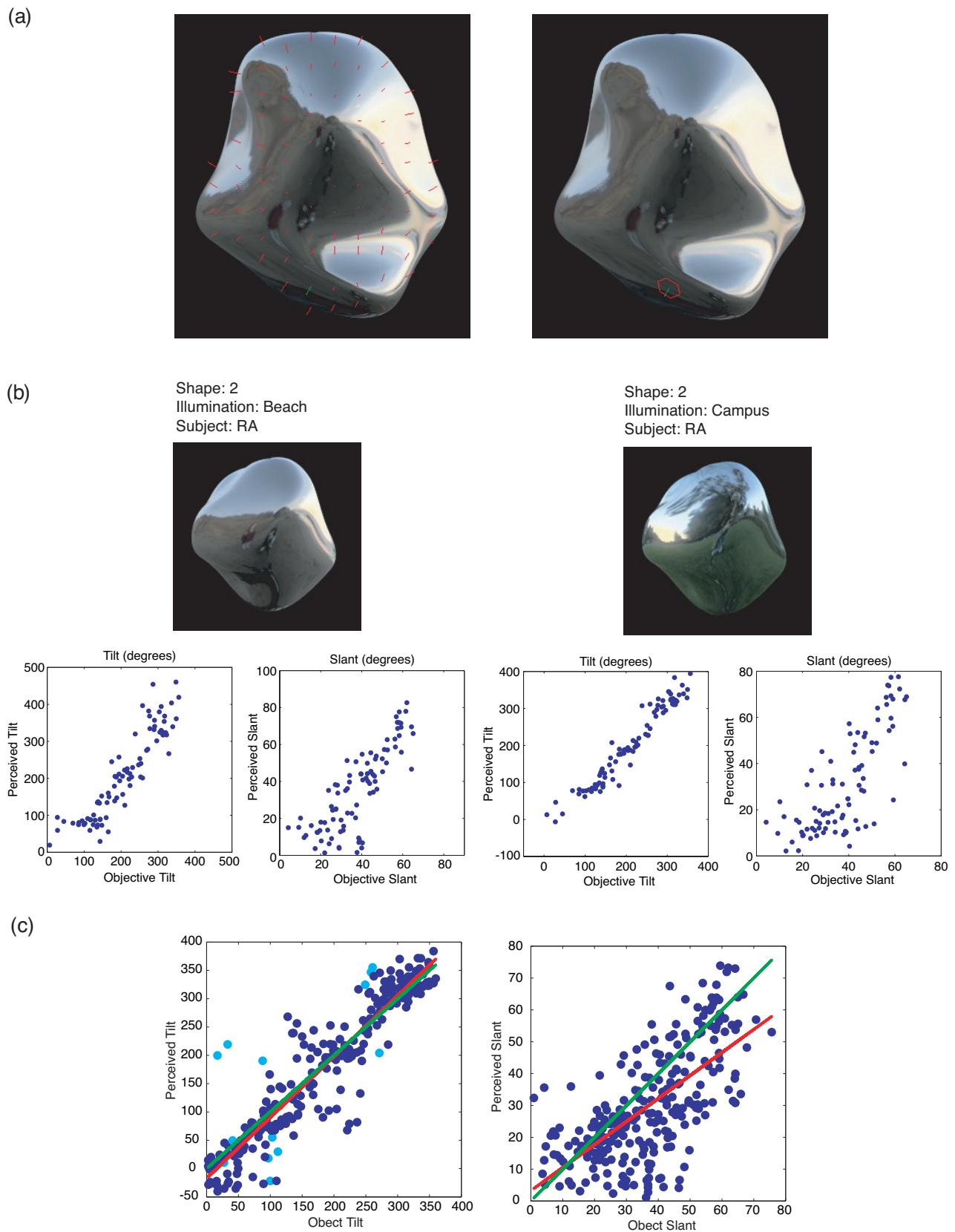


Figure 4. (a). Screenshot from gauge-figure task. Subjects adjusted gauge-figures to indicate surface normals. (b). Results of one subject. (c). Summary data pooled across subjects, illuminations, and shapes. Light blue dots show tilt estimates for which slant < 15 deg (i.e., objective tilt is ill-defined).

Results

In agreement with the demonstrations in Figures 1 and 2, we found that subjects were generally good at estimating the shapes of perfectly mirrored surfaces, even though the stimuli were presented without any context to specify the scene surrounding the object.

For the purposes of presentation, the 3D orientation of each surface normal can be represented as slant (orientation in depth) and tilt (orientation in the image plane). This is a standard azimuth and elevation representation of the hemisphere of possible responses (Stevens, 1983). Note that slant ranges from only 0 – 90 deg, while tilt varies from 0 – 360, hence the greater apparent spread of the data for the slant dimension. Note also that tilt is a circular dimension, which we have unwrapped for graphical purposes.

Example data from naïve subject RA are shown in Figure 4(b). The subject's estimates of both slant and tilt are quite accurate. Viewing the object under a different illumination also leads to accurate estimates of both slant and tilt.

Figure 4(c) shows data pooled across shapes, illuminations, and subjects. The green line represents ideal performance; the red line is the best-fit linear regression. Although we found that some shapes yielded slightly better performance than others, all measurements were well above chance performance. We conclude that subjects can reliably and quite accurately estimate the 3D shape of mirrored objects in realistic scenes, without any context to specify the scene surrounding the object. This suggests (i) that specular reflections are a sufficient cue for shape estimation, and (ii) that subjects do not need to construct a rich and accurate representation of the surrounding scene to recover shape from specular reflections.

Findings II: Theory and image analysis

So far we have argued that the visual system can recover 3D shape directly from the pattern of distorted reflections across a specular surface. We have shown that subjects can reliably and quite accurately recover the 3D shape of purely specular surfaces in the absence of context to specify the scene surrounding the object.

We will now explain in detail how powerful constraints on 3D shape can be extracted directly from the continuously varying texturelike patterns found on the surface of specular objects. We will first discuss some similarities and some key differences between textures and specularities. We will then demonstrate how local constraints can be extracted directly from the image by populations of simple oriented filters. Finally, we will measure the reliability and accuracy of these constraints for computer generated shapes rendered in realistic scenes.

Similarities and differences between specularities and texture

The idea that 3D space can be depicted using texture gradients dates back at least as far as the Renaissance. Since Gibson's (1950a) suggestion that texture gradients provide a visual cue to the inclination of a surface, the problem of shape-from-texture has received a considerable amount of attention both theoretically (e.g., Blake & Marinos, 1990; Cutting & Millard, 1984; Clerc & Mallot, 2002; Malik & Rosenholtz, 1997; Stevens, 1981; Super & Bovic, 1995; Witkin, 1981) and psychophysically (e.g., Buckley & Frisby, 1993; Cutting & Millard, 1984; Cumming, Johnston, & Parker, 1993; Gibson, 1950b; Li & Zaidi, 2000; Rosenholtz & Malik, 1997; Todd & Akerstrom, 1987; Todd, Oomes, Koenderink, & Kappers, 2004; Zaidi & Li, 2002).

The basic intuition behind shape-from-texture is depicted in Figure 5.

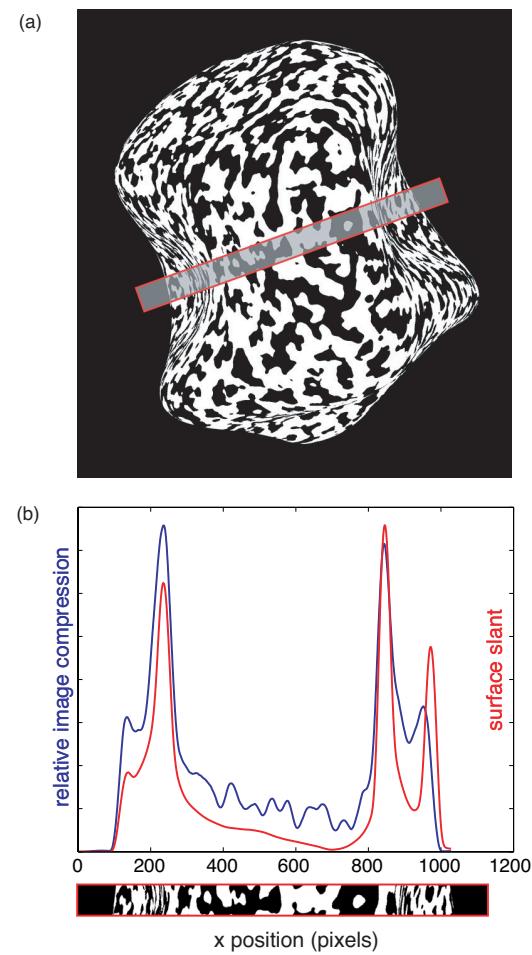


Figure 5. The intuition behind shape-from-texture. (a). A 3D shape coated in texture. In the image, the texture undergoes compressions due to foreshortening. (b). The pattern of image compression across the highlighted region of the image is plotted in blue. The objective slant of the surface is plotted in red. There is a good correspondence between the compression of the texture and the slant of the surface.

Consider an irregularly shaped object that is covered with a stationary isotropic texture, as shown in [Figure 5\(a\)](#). In the absence of shading or stereo the image leads to a vivid impression of 3D shape. The image information that carries this impression is thought to be the distinctive patterns of compression and rarefaction of the texture across the image. We can plot the degree of texture compression across the highlighted region, as shown in [Figure 5\(b\)](#). Superimposed on this plot is the objective slant of the surface at the corresponding points on the 3D model. We can see that there is a strong correspondence between the slant of the surface and the compression of the texture in the image. The important point is that there is a systematic relationship between the pattern of distortions in the image and some property of the 3D shape of the surface.

This basic intuition can be extended to specular surfaces. If we examine a specular surface that is reflecting a realistic environment, we see that the reflected world is dis-

torted into patterns of compression and rarefaction by the geometry of the surface. As with the case of texture, there appears to be some systematic relationship between the properties of the shape and the degree of compression of the reflected world. If this is the case, then in principle the visual system can recover properties of the 3D shape simply by measuring the patterns of distortion across the shape, much as it does with shape-from-texture. This is the intuition behind our formulation of the recovery of shape from specular reflections.

It is important to appreciate, however, that this is only an *analogy* between textures and specularities. The rules that relate 3D shape to the patterns of distortion are different for shape-from-texture and “shape-from-specularities.” We will now demonstrate these differences.

[Figure 6](#) contains an ideal planar surface that is rotated in depth. Note that when a plane is rotated in depth, the *first derivative* of surface depth changes, but higher derivatives remain constant at zero. In the left column, the surface is coated with a stationary isotropic texture; on the right, the surface is a perfect mirror. When we rotate the textured surface away from fronto-parallel, the corresponding texture elements in the image become compressed due to foreshortening. However, in the case of the mirror, as the surface rotates, all that happens is that the mirror selects different parts of the surrounding world and projects them into the image; the reflection is not compressed in any way [see [Figure 7\(a\)](#)]. Thus, in the case of textures, compression is a function of the first derivative of the surface, but in the case of mirrors it is not.¹

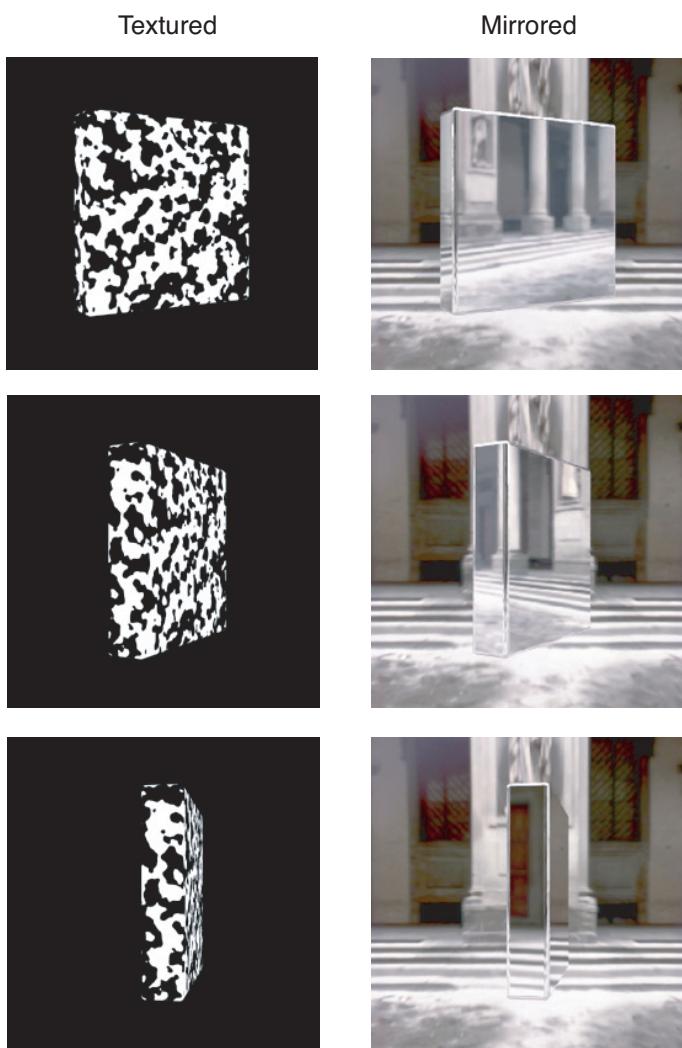


Figure 6. A planar surface at 30°, 60°, and 80° slant. In the first column the surface is coated in stationary isotropic texture. In the second column the surface is a perfect mirror. Note that the texture becomes increasingly compressed by foreshortening. However, the reflection in the mirror is not compressed at any surface orientation.

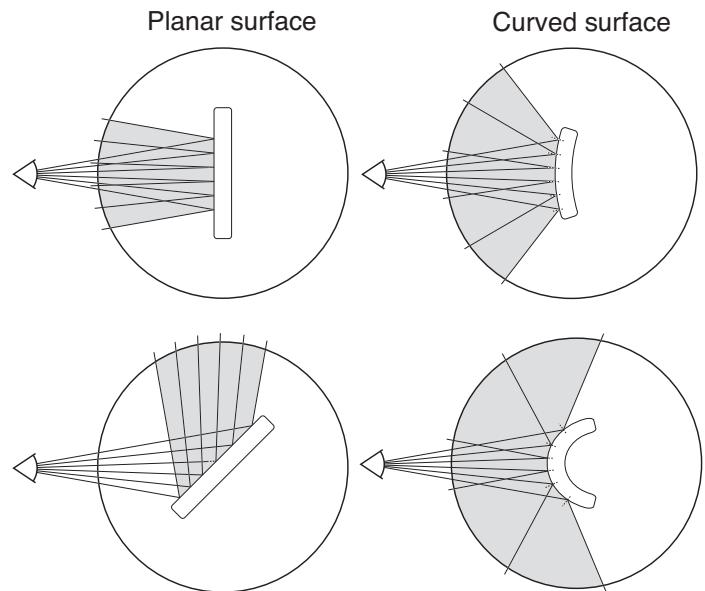


Figure 7. The geometry of mirror reflection for planar (a) and curved surfaces (b). The gray region represents the angular portion of the environment that is reflected into the image. The larger this angle, the greater the degree of compression of the image. (a). Note that rotating the flat plane has no effect on the proportion of the world compressed into the image. (b). By contrast, compression increases dramatically as a function of surface curvature.

In Figure 8, we consider what happens with a curved surface. Again, in the left column the surface is coated in texture, while on the right, the surface is a perfect mirror. Let us start with the sphere. In the case of the textured surface, there is a slight compression of the texture toward the edge of the sphere. This is because the first derivative of the surface increases as we move from the center to the edge of the sphere. In the case of the mirror, the image of the reflected world is also compressed. However, this compression has a different cause. A highly curved surface “sees” (i.e., points at) more of the world than a slightly curved surface, as shown in Figure 7(b). Thus, a surface with a large second derivative compresses a large angle of incident directions into a small portion of the surface. The more curved the surface, the greater the compression. If we conceive of the reflected world as a texture, then the degree of compression of the texture elements in the image is directly

related to the *second derivative* of the reflecting surface.

Note that in the middle of the sphere, the second derivative of the surface is equal in all directions and thus the image is equally compressed in all directions. However, toward the edge of the sphere, the second derivative is *large* in the direction perpendicular to the circumference, but *zero* in the direction parallel to the circumference. Hence, the reflection gets stretched into concentric streaks toward the edge of the sphere.²

To emphasize this relationship between the second derivative and image compression, let us consider what happens when the sphere is elongated into an egg-shape. In the case of the textured surface, when the egg is elongated all that happens is that more texture elements are recruited onto the surface. Because there is a small difference in the first derivatives, the texture is slightly less compressed along the principle axis of the egg. However, in the case of the mirror, there is a much more dramatic effect. In the direction of high curvature, the mirrored egg compresses many features from the world into a small portion of the image. By contrast, in the direction of low curvature, the surface compresses a relatively small angle of the surrounding world into a relatively large region of the image. Thus, the reflections are effectively stretched into parallel streaks along the direction of minimum curvature. Importantly, this means that surfaces that are anisotropic in curvature tend to produce patterns that are anisotropic in the image. The degree and direction of the anisotropy in the image carry information about the second derivatives at the corresponding location on the surface. This is the basis of the theory that we discuss in greater detail below. Previous researchers have noted that highlights are elongated along directions of minimum surface curvature (Beck & Prazdny, 1981; Blake & Brelstaff, 1988). Here, however, we elaborate in detail how the visual system can exploit this effect to recover constraints on 3D shape.

To summarize:

- (i) For textures, the compression in the image is a function of the *first derivative* of the surface.
- (ii) For specular reflections, the compression in the image is a function of the *second derivative* of the surface.

The dependency of specular reflections on the second derivative of the surface generally leads to characteristic anisotropies in the image. Specifically, whenever the minimum and maximum second derivatives are different, the reflected world is stretched in the direction of minimum surface curvature. In the extreme this leads to a characteristic pattern of striations along the direction of minimum second derivative, which we argue provides strong local constraints on 3D shape. Below, we also discuss how the different mappings can be used to distinguish between textures and specular reflections, but first we consider how the image compressions can be extracted from the image.

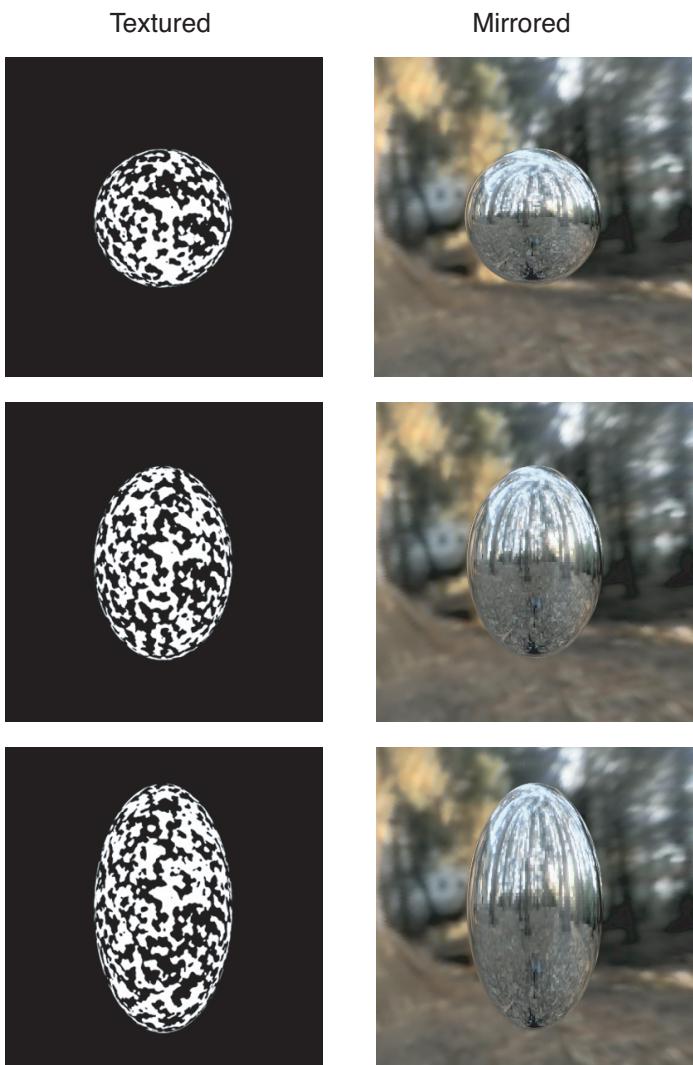


Figure 8. A sphere gradually elongated into an egg shape. In the left column the surface is textured; in the right column it is mirrored. Note that the texture is not preferentially stretched along the egg. By contrast, the reflected scene becomes stretched because of the lesser curvature along the vertical axis.

Extracting constraints on 3D shape using a population of oriented filters

We will now demonstrate how a population of simple oriented filters can measure local anisotropies, and, therefore, make image measurements that are directly related to 3D shape. For demonstration purposes, we will place mirrored surfaces in a synthetically generated scene with known image statistics, specifically, random noise with a $1/f$ amplitude spectrum.³ Note that this texture contains no recognizable objects, such as buildings or trees. We will consider the responses of a population of local image operators (filters) that are tuned to different image orientations (the details of these filters are described below).

Consider the spherical mirror in Figure 9(a). As we have already argued, a curved surface compresses many features from the world into a small portion of the image.

Thus the reflection of the noise is “miniaturized” in the surface of the sphere. However, at the center of a sphere, the compression is equal in all directions because the surface is equally curved in all directions. This means that there is no preferential stretching of the reflected texture in the image. Thus the close-up of this region contains a broad distribution of orientations, just as the surrounding world does. Let us consider the responses of the population of filters to the close-up of the surface. Because the close-up contains features at all orientations, all the filters in the population respond approximately equally strongly. The approximately flat population response indicates that the second derivative in the middle of the sphere is equal in all directions.

As before, we will now elongate the sphere into an egg-shape, which is highly curved in one direction and less curved in the orthogonal direction [Figure 9(b)]. As before,

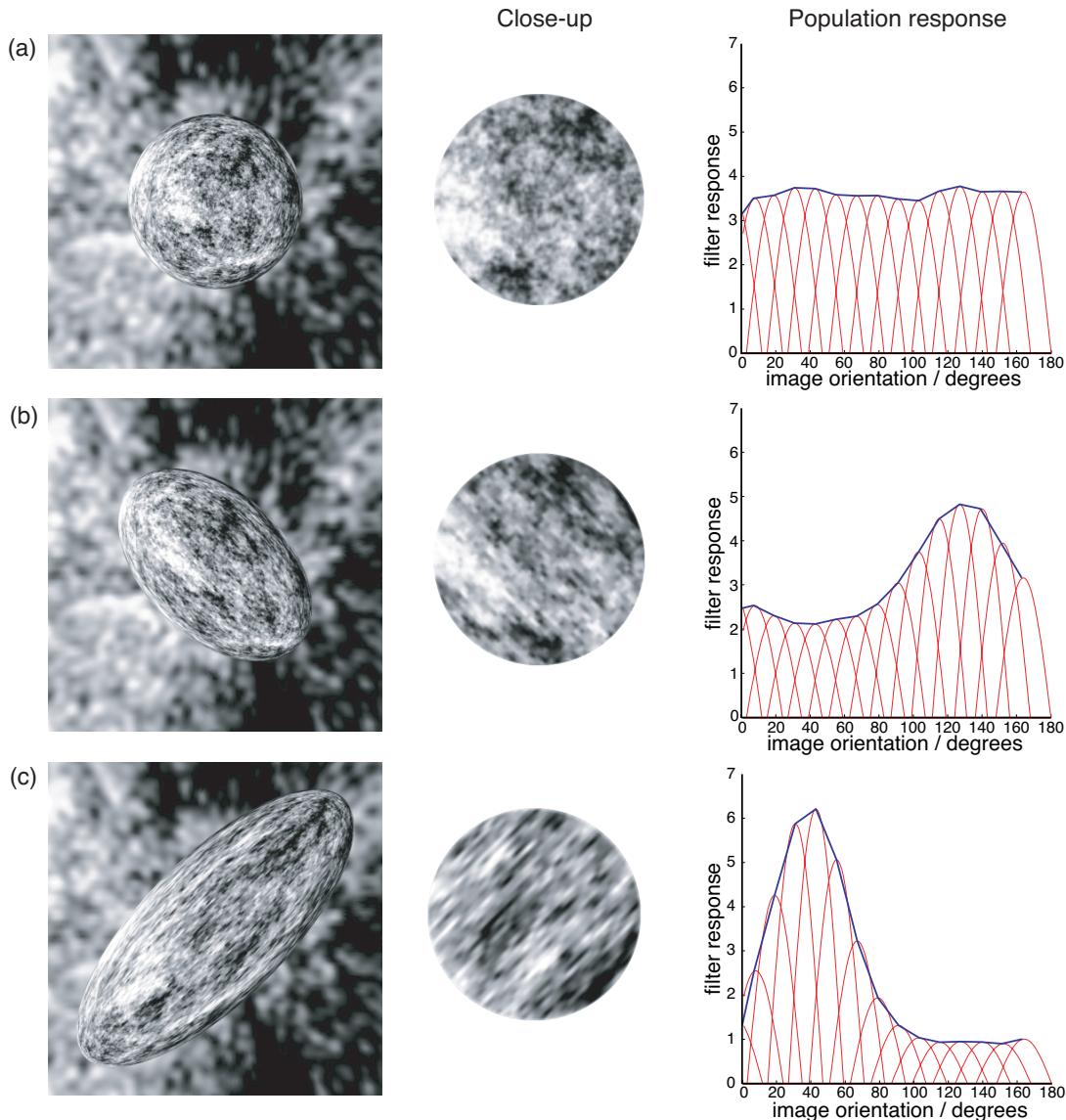


Figure 9. Mirrored surfaces in a world of $1/f$ noise, with responses of a population of oriented filters to the reflections. (a) A spherical mirror. (b) and (c) Egg-shaped mirrors. Note that the population response exhibits a peak that is aligned with direction of minimum surface curvature. Peak size increases with surface anisotropy.

the image is compressed in the direction of high curvature, and smeared out, by comparison, in the direction of low curvature. This smearing affects the orientations present in the close-up. Specifically, the reflected features become elongated into parallel diagonal streaks. Filters that are orthogonal to the streaks respond more weakly, while filters that are aligned with the streaks respond more strongly. Thus, the population response becomes peaked at the dominant image orientation.

Importantly, both the size and orientation of the population peak are directly related to the local 3D shape of the surface. To demonstrate this, we will rotate and elongate the egg to create the shape in Figure 9(c).

First consider what happens to the location of the peak response. By rotating the egg, we change the direction of minimum surface curvature. Recall that the reflection is most stretched in the direction in which the surface is least curved. Thus, when the direction of minimum surface curvature changes, the streaks rotate with the object. Accordingly, filters that were previously aligned with the streaks become suppressed, while different filters become enhanced, which causes the peak of the population response to shift. Thus, the orientation of the population peak provides a direct estimate of the direction in which the second derivative of the surface is smallest, which for brevity, we will call the *direction of minimum second derivative*.

Second, consider what happens to the size of the peak response. By elongating the egg, we have also changed the ratio between the minimum and maximum second derivatives. This exaggerates the stretching of the reflection, which makes the image more streaky, as shown in the close-up. Accordingly, the filters that are aligned with the streaks become enhanced, while the orthogonal filters become increasingly suppressed. Thus, the size of the population peak serves as a direct estimate of the relative magnitudes of the maximum and minimum second derivatives, which for brevity, we call *surface anisotropy*.⁴

Population codes are stable across realistic scenes

We have argued that a population of filters can estimate some local curvature properties of simple shapes, such as eggs, when placed in a standard scene with known statistics. However, can this theory be applied to arbitrary, complex shapes viewed in realistic scenes? For complex objects, the second derivative changes continuously across the surface. Accordingly, a simple feature in the real world, such as a straight line, can be warped into complex patterns in the image. How can the visual system decode these complex distortions without knowing the shape of objects that are reflected in the surface?

We have been arguing that the visual system does not attempt to interpret the warped reflection of recognizable environmental features. Rather, it simply treats the distorted reflections as a continuously varying “texture.” It is the continuous variation in the orientation content of this

texture that carries information about 3D shape. Specifically, we are suggesting that the visual system could apply the population coding strategy simultaneously at all locations in the image, to recover the direction and relative magnitude of the second derivative at all visible locations on the surface. For this to be a viable hypothesis, the way that the reflections “flow” across the image has to depend more on the shape of the object than on the reflected scene.

In this section we discuss the stability of reflections across changes in the scene that is being reflected in the surface. Before discussing empirical measurements, we will demonstrate the basic intuition. Consider the irregular 3D shape in Figure 10. The surface is shown reflecting three different scenes. At first sight the reflections of these three scenes in the surface look quite different. However, if we

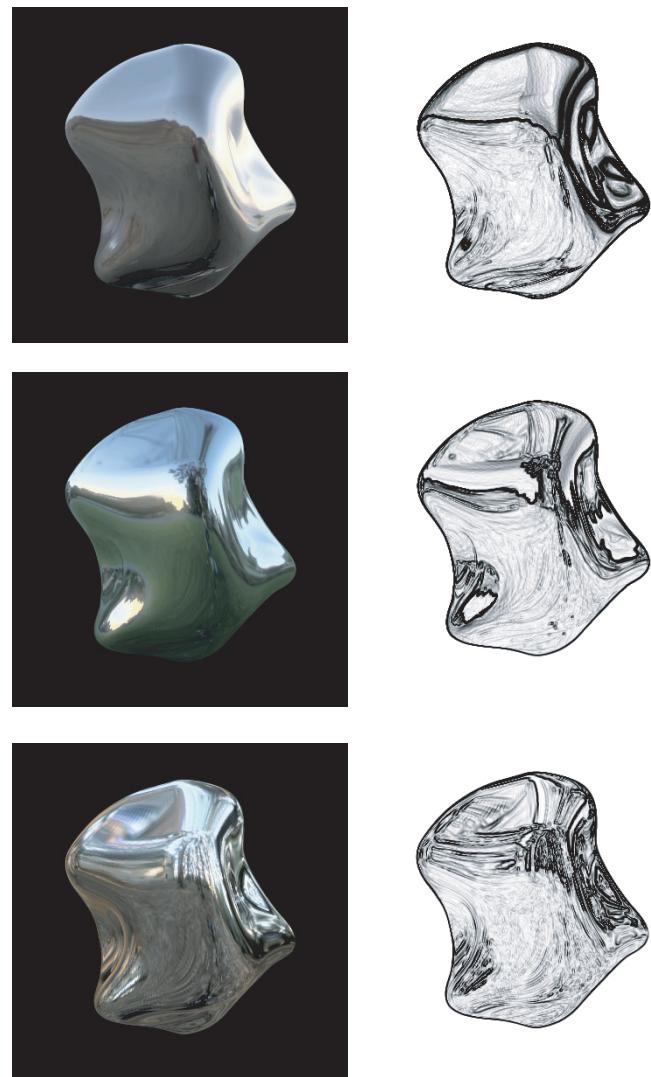


Figure 10. The orientation structure of mirrored surfaces. The top row shows a mirrored surface in three different scenes. Bottom row shows output of simple edge-detecting algorithm. Note that the dominant edge orientation remains quite stable across scenes.

pass the image through a simple edge-detecting algorithm, we see distinctive patterns of image orientation across the image, which are remarkably well conserved across the scenes. We suggest that the visual system uses these characteristic “orientation fields” as a cue to 3D shape. The fact that orientation fields can remain quite stable across scenes could account for the stability of 3D shape perception across changes in the reflected scene.

To test this idea empirically, we computer generated nine mirrored objects with different 3D shapes but identical silhouettes. We rendered each shape under nine differentDebevec light-probe illuminations, generating a 9x9 grid of images. Example images are shown in Figure 11. We

then calculated the responses of a population of oriented filters at each location in every image.

The model population of filters consisted of a simple, local first-derivative operator (i.e., a small odd-symmetric filter with only a single positive and a single negative lobe) that was “steered” through 24 equal orientation steps between 0 and 180 deg. The filters measure orientation energy, which is phase insensitive (i.e., they do not respond to the contrast polarity of the intensity variations, only to the orientation).

The implementation of the steerable pyramid algorithm that we used is described elsewhere (Simoncelli, Freeman, Adelson, & Heeger, 1992; Simoncelli & Free-

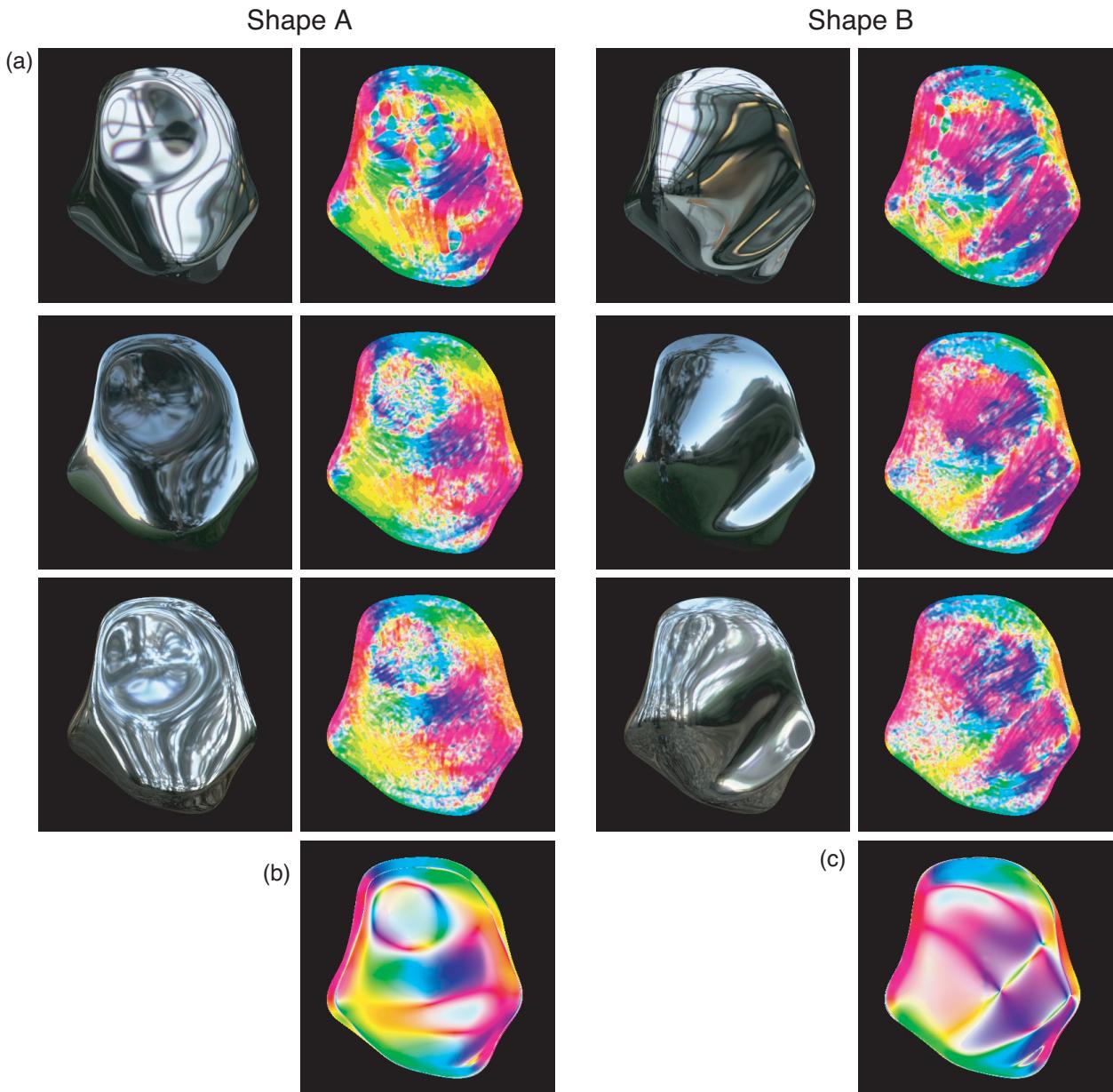


Figure 11. (a) Two shapes rendered in three different scenes, with corresponding orientation maps. Hue denotes peak orientation (estimated direction of minimum curvature), saturation denotes size of peak (estimated surface anisotropy). (b) and (c) Objective orientation maps, derived from shape model. Hue represents objective direction of minimum curvature, saturation represents objective surface isotropy.

man, 1995), and is available online at <http://www.cis.upenn.edu/~eero/steerpyr.html>. The steerable pyramids were built in the space domain (as opposed to the spatial frequency domain), using the command `buildSpyr`. We derived population measurements from the distribution of responses across the 24 different filter orientations at each image location. Because the filters simply measure the local derivative in image intensity, they operate at the finest possible spatial scale. We also tested filters at other scales and obtained comparable results.

The result for each image is an “orientation field,” which plots the population response at every image location. Example orientation fields are shown as color plots in Figure 11. We represent the orientation of the peak population response using hue; thus, for example, red means that the dominant local image orientation is vertical. We represent how defined the population peak is using color saturation.⁵ Thus, where the population peak is ill defined, the orientation map washes out to white, whereas, where the peak is clearly defined, the colors become vivid. Note that the orientation field can be thought of as an estimate of the *direction of minimum second derivative* and *surface anisotropy* at every visible location on the object’s surface.

We have found that orientation fields are diagnostic of shape, and remain quite stable as the object is moved from scene to scene. For example, in Figure 11, the orientation maps of Shape A are extremely similar across scenes, and quite different from those of Shape B. On average, pairs of orientation maps were well correlated if they originated from the same shape, even though the shapes were rendered in different scenes (population peak orientation: $r^2 = 0.92$; population peak size: $r^2 = 0.67$). By contrast, orientation maps were significantly less well correlated when the shape varied, even when the surrounding scene was held constant (population peak orientation: $r^2 = 0.79$; population peak size: $r^2 = 0.30$).⁶

This shows that although moving a specular object into a different scene can dramatically change the patterns of light and darks across the surface, the “texturelike” patterns remain surprisingly stable. Put another way, although the *luminance* content of the image varies considerably with the reflected scene, the *orientation* content of the image remains relatively stable across scenes. Thus the visual system can rely on orientation fields to provide reliable information about 3D shape, as an object is moved around in the world.

Population codes provide accurate information about shape

We have shown that the orientation field for a given shape is quite stable across changes in the scene. But do orientation fields provide *accurate* information about 3D shape? Recall that orientation fields constitute an estimate of the *direction of minimum second derivative* and the *surface anisotropy* at each visible location on the object’s surface. Are these estimates accurate? How do they compare to the objective curvatures of the 3D shape model? We will now

evaluate how well orientation fields estimate 3D curvatures by comparing the estimates with the objective values derived directly from the 3D shape model.

For comparison, objective second derivatives can also be displayed as color plots. This time, hue represents the objective direction of minimum second derivative (as opposed to the estimate derived from the image). Likewise, color saturation represents the objective anisotropy of the surface. Example objective orientation maps are shown in Figure 11(b) and 11(c). The correspondence between the objective and estimated orientation fields is quite striking for both Shape A and B.

We measured the error between objective directions of minimum second derivative and the population estimates at every pixel location for every image in the 9x9 grid. A histogram of errors is shown in Figure 12(a). Note that the distribution of errors is peaked around zero, and 74.78% of estimates fall within 30 deg of the correct value.

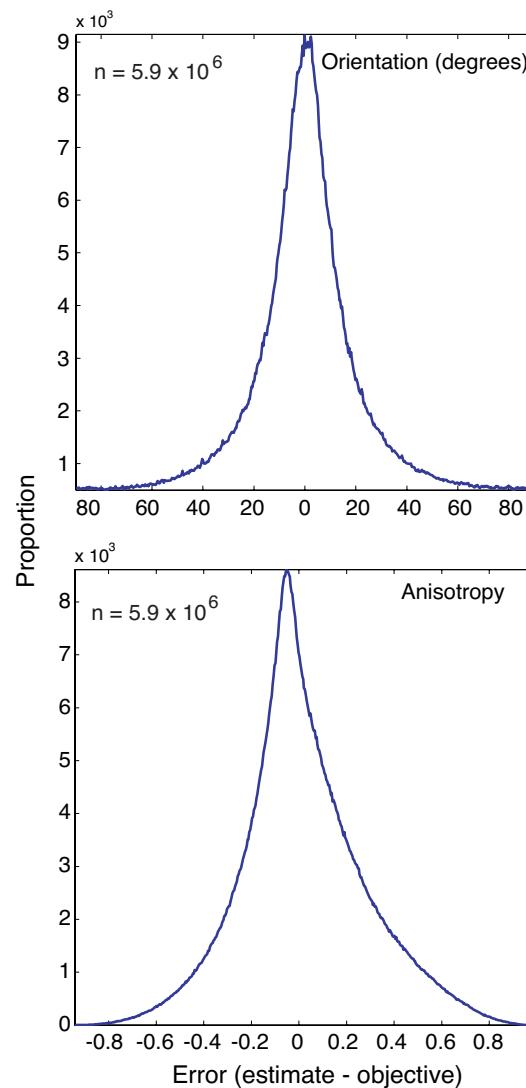


Figure 12. (a) Error between estimated and objective direction of minimum second derivative for all images in the 9x9 grid. (b) Error between estimated and objective surface anisotropy.

Likewise we measured the error between estimated and objective surface anisotropy for every image [Figure 12(b)]. Again the distribution peak is close to zero, and 80.64% of estimates fall within 33.3% of the correct value.

We conclude that simple image measurements are capable of providing the visual system with reliable and accurate estimates of the direction of minimum second derivative and surface anisotropy at every visible location on a specular surface. Because these measurements remain quite stable across scenes, the visual system does not need to estimate the environment surrounding an object to recover 3D shape. Thus, specular reflections are easier to use for shape estimation than previous computational work would suggest.

Discussion

It is commonly believed that visual perception is achieved by a process of “inverse optics” (Helmholtz, 1867/1962; Poggio, Torre, & Koch, 1985), in which the visual system reverses the physics of image generation to infer the outside world from an image. When posed this way, recovering the shape of a mirror is extremely difficult because all visible features belong to the environment surrounding the object, rather than to the object itself. It would seem that the visual system would have to form an extremely sophisticated model of the environment to recover the object’s underlying shape. However, we have shown that the problem can be reformulated in terms of image measurements that are diagnostic of shape but which remain quite stable as the object is moved from scene to scene. This way, early visual processes could estimate curvature properties directly from the image, without having to build an explicit representation of the environment.⁷

The results of our psychophysical experiment show that subjects are good at recovering the 3D shape of perfectly mirrored objects. This can be contrasted with previous claims (Oren & Nayer, 1996; Savarese et al., *in press*). There are two notable aspects of the result. First, the fact that performance was good in the absence of any context implies that the image local to the surface of the object provides sufficient information to perform the task. Second, the fact that performance was good across changes in the reflected scene suggests that the information used by the visual system is relatively stable across image variations that are due to the scene.

To account for these results, we proposed that the visual system recovers shape from the patterns of distortion that occur when the world is reflected in a curved surface. Rather than evaluating the distortion of specific environmental features, the visual system can treat the image as a continuously varying texture whose statistics are determined by the 3D shape. The advantage of this is that 3D curvature properties can be estimated directly from the distribution of orientations passing through each location in the image,

without having to represent the environment surrounding the object.

We have shown that these image measurements can be performed by populations of simple local filters. Specifically, a population of filters tuned to different image orientations produces a peak response that is closely aligned with the direction of minimum second derivative. The relative magnitude of minimum and maximum second derivatives is specified by how well defined the population peak is. When applied in parallel to all image locations, we have shown that this population coding strategy provides accurate estimates of 3D curvature properties across a range of real-world scenes. It is worth noting that these measurements are at least biologically plausible, as it is well known that primary visual cortex contains cells that are tuned to different image orientations (DeValois, Yund, & Hepler, 1982; Hubel & Wiesel, 1959, 1962, 1968; Schiller, Finlay, & Volman, 1976).

The ambiguity of orientation fields

It is important to clarify that orientation fields provide a field of local *constraints* on 3D shape; they do not in themselves constitute a complete estimate of the shape model. Indeed, multiple 3D shapes are consistent with a given orientation field. We will now discuss some of the ambiguities that remain to be resolved.

Local image anisotropy does not specify the sign of local surface curvature (i.e., there is concavity vs. convexity ambiguity). This ambiguity is not unique to the interpretation of specular reflections: It is well known that shape-from-shading suffers from a similar limitation (Kardos, 1934; Ramachandran, 1988, 1990). There are a number of ways that this ambiguity might be resolved. First, it is generally believed that the visual system has a built-in preference (or “prior”) for convex interpretations (Hill & Bruce, 1993, 1994; Langer & Bülthoff, 2001; Mamassian & Landy, 1998; Symons, Cuddy, & Humphrey, 2000; Woodworth & Schlosberg, 1954). This prior may help to disambiguate the global sign of curvature of the object.

Second, enforcing mutual consistency between local interpretations is likely to reduce the number of possible interpretations quite dramatically, especially if the bounding contour of the shape is used to provide additional constraints (Howard, 1983; Koenderink, 1984). Li and Zaidi (2000) have shown that for textured surfaces, convexities and concavities lead to distinct orientation field patterns. It seems likely that a similar argument also applies to orientation fields generated by specular reflections. Although each local measurement is ambiguous in isolation, the patterns made by entire fields of local measurements seem to carry the necessary information.

Indeed, more generally there appears to be something about the global structure of orientation fields that carries information about the global form of the underlying surface. It is important to note that orientation fields are highly organized. Orientation varies smoothly across the

image as the distorted reflections twist and turn across the surface. It seems to be the organization of these patterns that specifies 3D shape. However, at present we do not know how to characterize this information.

Of course, any transformation that preserves the direction of minimum second derivative and the ratio of minimum to maximum second derivatives will, by definition, leave the orientation field unchanged. Examples of such transformations include scaling along the line of sight and affine shearing. If the orientation field remains constant, then the visual system would clearly require additional information to distinguish between shapes that are related to one another by these transformations.

However, as we have already stated, we are not claiming that orientation fields are the *sole* source of information about shape that can be derived from specular reflections, nor that orientation fields are the underlying “representation of shape” in the human visual system. Rather, our claim is that there exists a source of information that can be extracted from the image by relatively simple measurements, without reference to the objects surrounding the surface of interest. This information provides strong *constraints* on 3D shape.

Interactions with the occluding boundary

When we look at the image of an entire object, we see not only the internal structure of the surface, but also the “occluding contour” – the boundary of the object where the surface curves out of view. This contour also carries information about 3D shape (Koenderink, 1984). Is it possible that the impression of 3D shape that we get from mirrored objects results primarily from the occluding contour?

[Figure 13](#) suggests that this is unlikely. All four images have identical silhouettes, but the impression of 3D shape is very different. The three images that contain specular reflections look vividly more volumetric than the silhouette alone, and also look strikingly different from one another. This suggests that orientation fields carry more information about 3D shape than the bounding contour alone.

However, although the occluding contour is not a sufficient cue on its own, we believe it can provide extremely useful boundary conditions on the interpretation of orientation fields. Furthermore, for closed, globally convex objects, the orientation field becomes more reliable closer to the occluding boundary. The reason for this is that the second derivative of the surface increases as the object curves out of view. This suggests that removing the occluding boundary should have a detrimental effect on perceived 3D shape.

In [Figure 14](#), we take a couple of objects and remove the occluding boundary by cropping regions from the middle of the image using an irregularly shaped outline. Most observers agree that the vividness of the sense of 3D shape is reduced by this manipulation in images (c) and (d).

Nevertheless, it is difficult to know how much of this effect is due to the occluding contour per se, and how

much is due to the fact that cropping the image invariably removes some of the orientation field as well. In images (e) and (f), the same objects are shown cropped with a larger contour. These images yield a somewhat more compelling sense of 3D shape, even though the occluding boundary is still absent from the image. Many of the recesses and bulges become visible, and we regain the impression that some parts of the surface are closer to us than others. Thus, the occluding contour is not *necessary* for the recovery of shape from distorted reflections, although it certainly plays an important role.

Beyond mirrors

How general is the strategy that we have outlined? We have shown that simple image measurements can recover certain shape properties from *perfect mirrors*, but most objects in the world are not perfectly mirrored. Most materials scatter light in many directions and do not form perfect images of the world on their surfaces. How can our proposal be generalized to deal with a wider range of materials?

We will now consider two possibilities. The first possibility is that the visual system might be able to separate specular reflections from other surface properties (such as shading and texture), and apply the proposed measurements only to the specular component. If the visual system could somehow “skim off” the specular component of the image, then its orientation measurements would be uncontaminated by other surface properties. This way the visual system could apply our proposed strategy to any material

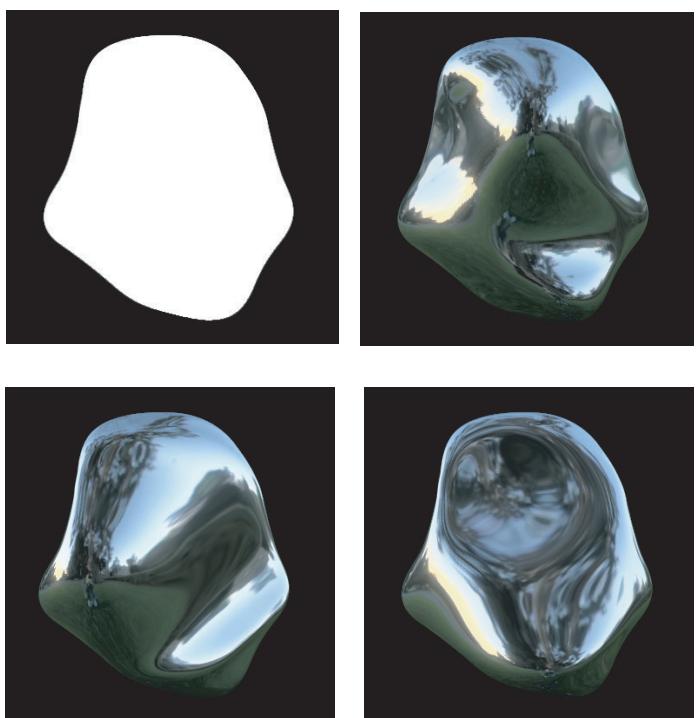


Figure 13. Four images with identical silhouettes, but dramatically different apparent 3D shapes. The silhouette alone leads to only a weak sense of 3D shape when compared to the other three images.

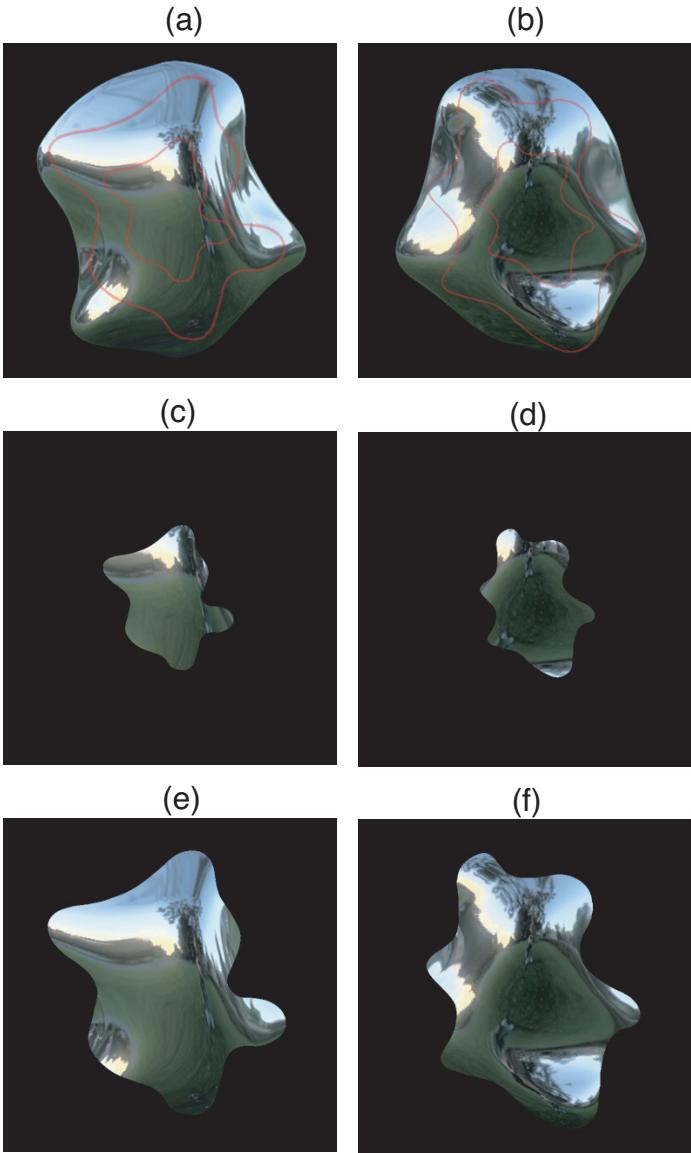


Figure 14. Effects of removing the occluding boundary on apparent 3D shape. (a) and (b), show original images. Red outlines indicate the regions that are cropped out in the following two panels. Note that when a small region is cropped out as in (c) and (d), the 3D shape percept is considerably impaired. However, when a larger region is cropped out, as in (e) and (f), the image largely regains its 3D appearance, even though the true occluding boundary is still missing from the image.

that has a specular component of reflection (e.g., a granny smith apple), and not only to perfect mirrors.

How plausible is this? It is important to note that the image of a glossy surface (such as plastic, or glazed ceramic) can be expressed as a simple linear sum of two component images: the matte component and the specular component. Put another way, specular reflections are *additive*: they are like a transparent layer superimposed on the underlying surface. Indeed, specular reflections can be thought of as a special case of Metelli's (1974) transparency.⁸ It is well known that the visual system can separate images of trans-

parent surfaces into the contributions of the background layer and the transparent filter through which it is visible (Adelson, 1999; Anderson, 1997; Heider, 1933; Koffka, 1935; Metelli, 1974; Singh & Anderson, 2002). We suggest that it is not unlikely that the visual system could separate specular reflections from the "background" surface that is visible through them. We discuss the separation of specularities from texture in greater detail below.

A second possible generalization could be that the visual system does not need to separate specular reflections from other types of surface reflectance. If other surface reflectance properties (e.g., diffuse shading) also lead to similar distinctive patterns of orientation across the image, then the orientation measurements that we have proposed could be robust across changes in surface reflectance, as well as across changes in the reflected scene.

The images in Figure 15 suggest that under some circumstances, orientation fields can be quite stable across changes in surface reflectance properties. Figure 15(a) shows a mirrored surface and its orientation field. The surface in 15(b) is a glossy plastic. Note that the detailed structure of the specular reflections is lost: the specularities are mere "highlights." Despite this, the orientation field continues to resemble the orientation field derived from the mirrored surface. In (c) we have roughened the surface so that the highlights become blurred. However, this blurring has little effect on the distribution of orientations at each image location, and thus the orientation field remains quite stable. This suggests that to use specular reflections for shape estimation, the visual system might not have to separate them from the underlying surface.

Previously, a number of authors have argued that the visual system could use the orientation structure of shaded images to estimate shape from shading. For example, Koenderink and colleagues have long argued that it is the "pattern of isophotes" across a diffuse surface that the visual system uses to recover shape from shading (e.g., Koenderink & van Doorn, 1980; Koenderink & van Doorn, 2003). More recently, Zucker and colleagues (e.g., Ben-Shahar & Zucker, 2001; Breton & Zucker, 1996; Higgins, Chen, Belhumeur, & Zucker, 2001) have repeatedly argued that shape-from-shading ought to be based on "shading flow." They note that diffuse shading leads to orientation fields that are stable across changes in albedo and cast shadows, and that these orientation fields can be used for the shape estimation and edge classification.

The orientation structure of shaded images is difficult to see because shading is so smooth. However, in Figure 16, we show the isophotes across a shaded Lambertian surface. This reveals the latent orientation structure of the image. These orientation patterns exhibit some clear similarities to the distorted reflections across the mirrored surface.

It is important to note that the orientation structure of shaded images is much less stable than for mirrored surfaces. Changing the direction of illumination can distinctly alter the pattern of isophotes across a shaded surface (Koenderink & van Doorn, 1980). However, the important

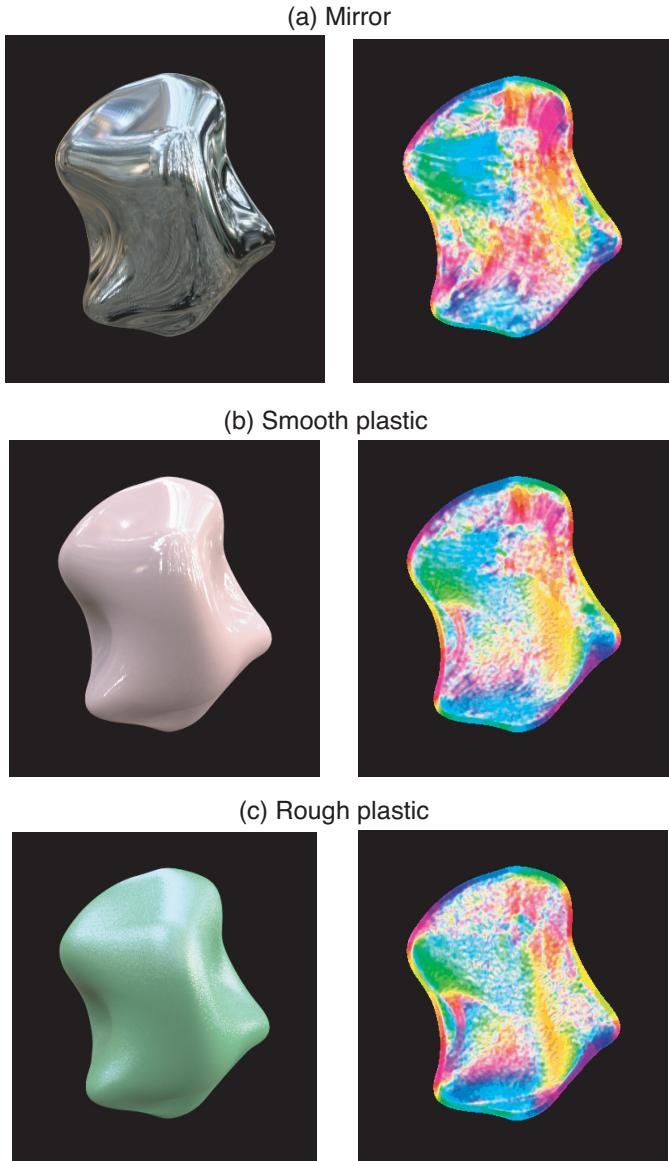


Figure 15. The population coding strategy generalizes to non-mirrored surfaces. (a) A mirrored surface. (b) A smooth plastic surface. (c) A rough plastic surface. Orientation maps remain quite stable across changes in material.

point is that the orientation structure of the images appears to carry information about 3D shape. We suggest, then, that specular highlights and diffuse shading may not provide fundamentally different cues to shape. Rather, they appear to operate with the same basic currency—orientation fields that can be extracted from the image by relatively simple image measurements.

Using orientation fields to distinguish between textures and specularities

As we have already mentioned, textures and specular reflections have some things in common. Both lead to stochastic patterns in images that undergo compressions and

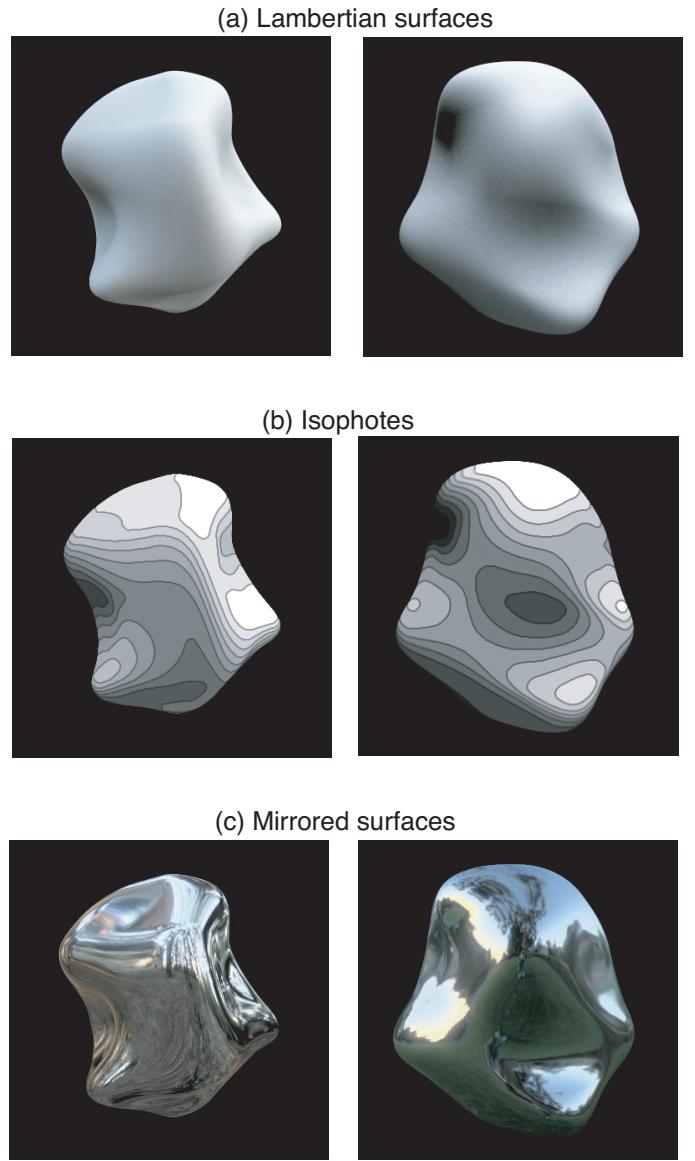


Figure 16. Revealing the latent orientation structure in diffuse shading. (a) two objects with diffuse reflectances. (b) Isoluminance contours of the images in (a). (c) Specular surfaces are presented for comparison. Note the similarities between the orientations in (b) and (c).

rarefactions that depend on 3D shape. And yet the visual appearance of a matte, textured surface is quite distinct from a glossy, specular surface. How can we tell them apart?

Under normal viewing there are many ways of distinguishing texture markings from specular reflections, including luminance or color information (Ullman, 1976; Klinker, Shafer, & Kanade, 1988; see also Yang & Malone, 2001); binocular disparities (Blake & Brelstaff, 1988; Blake & Bühlhoff, 1990, 1991), and characteristic motion fields (Koenderink & van Doorn, 1980; Oren & Nayer, 1996). A particularly vivid demonstration of the role of motion has been developed by Hartung and Kersten (2002, 2003). They have shown that distorted mirror reflections can be made to look like a pattern painted on a surface

simply by changing the way that they move when the object rotates. When the features slide across the surface, like well-behaved specularities, the object appears to be mirrored. However, when the same features are “attached” to the surface during motion, the appearance of the material changes dramatically, becoming matte and patterned rather than glossy. This is particularly impressive given that any single frame from the motion sequence leads to a vivid impression of a mirrored surface when viewed statically.

We have previously suggested that specular reflections of real-world scenes have characteristic image statistics (e.g., heavily skewed pixel histogram) that could help the visual system to distinguish reflections from textures (Fleming et al., 2003). Here we suggest that there is an additional cue that results from the different ways that textures and specular reflections are distorted by 3D shape.

Recall that the compression of textures depends (primarily) on the first derivative of the surface, while the compression of specularities depends on the second derivative of the surface. This means that a given shape will generally lead to different orientation fields in the image depending on whether it is glossy or coated with texture. In Figure 17, we demonstrate that this distinction can influence our sense of material quality.⁹

When a pattern is mapped onto the surface according to the rules for texture, the surface appears matte and painted (Figure 17, left column). By contrast, when the patterns are warped onto a surface according to the rules for reflection, the surface becomes somewhat more glossy-looking, even though the statistics of the patterns are unlike the real world (Figure 17, right column). Note that multiple factors can influence the apparent glossiness of the surface, especially the statistics of the patterns themselves. Here we have used patterns with ambiguous statistics in an attempt to isolate the source of information that comes from the distortion of those patterns across the surface.

One final example

We will now consider one final case to emphasize the circumstances under which textures and reflections lead to distinct orientation fields. Recall that reflections are compressed along *directions of high curvature*, while textures are compressed along *directions of high slant*. This means that the two orientation fields will be most different in shapes for which these two directions are most different. An example of such a shape is shown in Figure 18.

Along the longitudinal axis of the tube, surface curvature is zero, while around the circular cross-sections of the tube, the surface is quite highly curved. This means that glossy reflections tend to stretch along the tube, so that the orientation field is aligned with the long axis. In (a) we show the glossy surface and in (b) we show the dominant image orientation at each location across the surface.

Note that in the central bend of the object, the long axis of the tube slants away from the observer. This is interesting as it means that the direction of maximum curvature

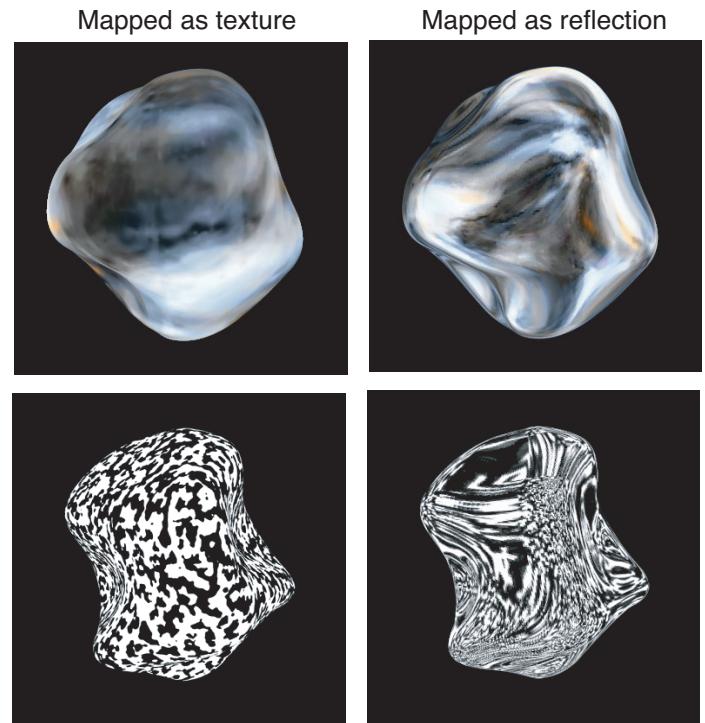


Figure 17. Apparent surface qualities can be influenced by the way that features are mapped onto the surface. In the left column the patterns are mapped according to the rules for texture. In the right column, similar patterns are warped onto the surfaces according to the rules for specular reflection. Observers generally agree that the images on the right look somewhat more glossy than the images on the left.

is almost perpendicular to the direction of maximum slant. When the surface is textured, as in (c), the orientation field will tend to be compressed into parallel rings that *cut across* the tube instead of running along it. This is shown in (d). To emphasize the difference, we can superimpose the two orientation fields, for this region of interest, as shown in Figure 19.

It is striking that both orientation fields lead to a vivid impression of 3D shape, although they are markedly different. If the visual system could somehow separate specular reflections from the underlying texture, then it could use the complimentary orientation fields as two convergent cues to the object’s 3D shape. Furthermore, the fact that orientation fields for textures and reflections can be so different may open the possibility of using image orientations themselves to distinguish between textures and reflections, even when they are directly superimposed in the image. This represents an interesting avenue for future research.

However, what is becoming clear is that the continuously varying *orientation structure* of images contains a wealth of information about the world, which remains to be fully explored. Orientation fields can carry reliable information about 3D shape and surface properties. Thus, populations of oriented filters can achieve much more than simple edge-detection.

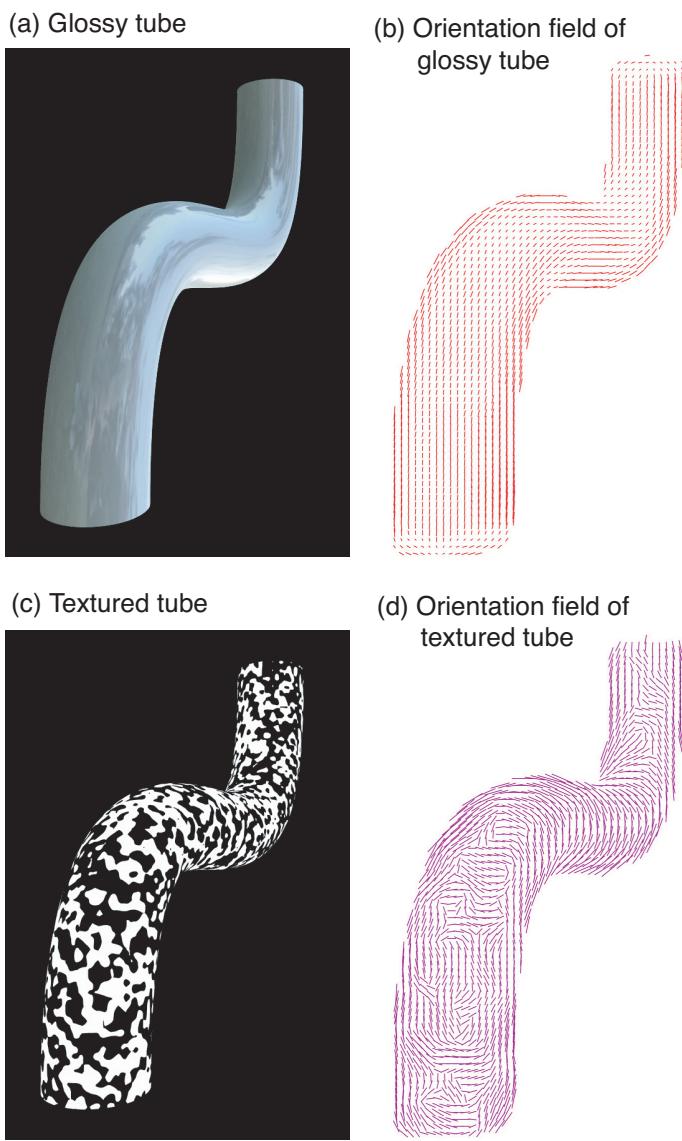


Figure 18. Textured and glossy versions of a tube-shaped object with corresponding orientation fields. Note that the orientation fields are distinctly different, especially in the region of the horizontal bend in the tube.

Conclusions

Many materials, including water, leaves, plastics, glazed ceramic, and metals exhibit specular reflections. It is well known that specular reflections aid shape perception, but the relevant image information has not previously been identified. Here we have presented a theory of how specular reflections could provide constraints on 3D shape.

At first sight, it is quite surprising that we can recover an object's shape from the distorted reflection of the world in its surface. As we noted in the "Introduction," the image of a perfectly specular object changes completely when the object is moved from scene to scene. Furthermore, to interpret the *distorted* reflection of an environmental feature (e.g., the warped image of a tree), it seems that the visual

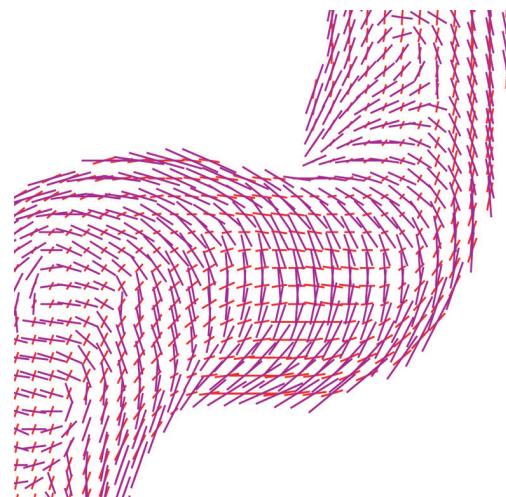


Figure 19. Orientation fields for texture (purple) and reflections (red) are shown superimposed to emphasise the differences. Note that at almost all locations, the two orientation fields have different orientations.

system would have to know the *undistorted* shape of that feature. In other words, it seems that the visual system would need access to a complete model of the world surrounding the object.

However, we have argued here that strong constraints on shape can be extracted directly from the image of a surface, without reference to the surrounding world. Specifically, we argued that the visual system treats specular reflections somewhat like a "texture" that is warped onto the surface. Thus the visual system can recover shape from specularities by analogy to the way that it recovers shape from texture.

We have shown, however, that there is an important difference between textures and specular reflections. A simple analysis of the geometry of projection reveals that the compression of texture is due largely to the slant of the surface (i.e., first derivative), while the compression of specular reflections depends on the rate at which the surface normal changes across the image (i.e., second derivative). The intuition behind this is that a highly curved surface "sees" (i.e., points at) more of the reflected world than a slightly curved surface and thus compresses more features into the same portion of the image. Importantly, when the surface has different curvatures in different directions, the reflections become dramatically distorted. In the extreme, the image is stretched into parallel streaks along the direction of minimum second derivative.

We then showed how these distortions can readily be extracted from the image by a population of filters tuned to different orientations. We showed that

- (i) the peak of the population response tends to align with the *direction of minimum second derivative*, while
- (ii) the size of the population peak indicates the *ratio of maximum to minimum second derivatives*.

The continuously changing curvatures across a complex shape lead to complex “texturelike” patterns across the surface of a specular object, which we call “orientation fields.” We argued that these orientation fields provide strong constraints on 3D shape.

We studied the orientation fields of specular surfaces that were rendered under a range of real-world scenes. We found that orientation fields provide accurate estimates of 3D curvature properties that remained surprisingly stable across changes in the reflected scene.

We have also performed a simple psychophysical experiment using the gauge-figure task. We found that subjects can reliably and quite accurately estimate the 3D shape of perfectly specular objects. There are three notable aspects of the results:

- (i) Subjects can perform the task even when the surface is a perfect mirror, and thus the image consists of nothing but a distorted reflection of the surrounding world.
- (ii) Subjects could perform the task even though the objects were cropped out of their original contexts and viewed against a neutral background, and thus there was no additional information about the world surrounding the object.
- (iii) Performance was quite reliable across changes in the reflected scene.

Together these findings support the idea that distorted reflections across a specular surface provide a stable, powerful source of information about 3D shape.

We have also argued that orientation fields may play a more general role in shape estimation. Under some circumstances, diffuse surfaces produce orientation fields that resemble those produced by specular surfaces. Thus the visual system may not have to separate specular reflections from the underlying surface to use them for shape estimation (although this might be possible anyway). More generally, we suggest that patterns of image orientation are likely to be the crucial “common currency” of shape estimation, which are shared by shading, highlights, and texture.

Finally, we argued that the visual system can use orientation fields to distinguish between textures and reflections. Because textures are compressed by slant while reflections are compressed by curvature, they generally create very different orientation fields. This difference can be used to change a surface from looking matte to glossy. Indeed, when textures and reflections are superimposed, the visual system may be able to use the distinctive orientation fields to separate the two contributions to the image.

In conclusion, the orientation structure of specular reflections appears to be a powerful source of information in visual perception. This information is both more stable and more readily accessible than previous computational work would suggest.

Acknowledgments

This research was supported by National Institutes of Health Grant EY12690-02 to EHA, a Nippon Telegraph and Telephone Corporation grant to the MIT Artificial Intelligence Lab, a contract with Unilever Research, and ONR/MURI contract N00014-01-0625. RWF was also supported by the Max Planck Society.

Commercial relationships: none.

Corresponding author: Roland W. Fleming.

Email: roland.fleming@tuebingen.mpg.de.

Address: Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany.

Footnotes

¹In fact, under perspective projection, there are two distinct processes that compress textures in the image. The first depends on the absolute depth of the surface (i.e., the “zeroth” derivative). The more distant a surface is, the smaller it is in the image, and thus the greater the compression of the texture. The second process is foreshortening, which depends on slant (i.e., the first derivative). There are three reasons for emphasizing the latter process. First, the compression due to distance varies as a function of the inverse tangent of the distance. Thus, the effect is only powerful for surfaces whose undulations in depth are large relative to the viewing distance. Second, the distance effect disappears under orthographic projection and yet we have a vivid impression of shape-from-texture under orthographic projection. Third, the compression due to depth is an *isotropic* scaling of the texture pattern. This shows up as a weak modulation in the spatial frequency content of the image. In contrast, the compression due to slant is by definition *anisotropic*: the texture is only compressed along the direction of slant. This leads to a powerful cue due to the characteristic orientation structure in the image, as discussed below. Previous work (e.g., Li and Zaidi, 2000, 2003) suggests that modulations in image orientation due to surface slant are more important for shape-from-texture than modulations in spatial frequency due to surface distance.

²Note that the second derivative of a surface is different from the intrinsic surface curvature. The curvature is equal in all directions and at every point on the surface of a sphere. What is important for image formation, however, is the rate at which the surface normal changes with respect to the viewer (i.e., the second derivative of the surface). Note also that the directions of maximum and minimum surface curvatures are always orthogonal to one another when measured with respect to the intrinsic coordinates of the surface. However, when projected into the image plane, these directions are only orthogonal when the surface is fronto-parallel. By contrast, the directions of minimum and

maximum second derivative are always orthogonal in the image plane.

³It is well known that images of natural scenes generally have a $1/f$ amplitude spectrum (Field, 1987). In fact the noise can be thought of as a natural image whose phase spectrum has been randomized. The noise has a flat (i.e., uniform) distribution of orientations.

⁴Specifically, we define surface anisotropy as $1 - \sqrt{(k_{\min}^2 / k_{\max}^2)}$, where k_{\min} is the minimum second derivative and k_{\max} is the maximum second derivative. Surface anisotropy is 0 if a local surface patch is equally curved in all directions (e.g., planar or center of a sphere); 1 if it is locally cylindrical, and intermediate if it is locally “egg-shaped.”

⁵Specifically, saturation = $1 - \sqrt{(p_{\min}^2 / p_{\max}^2)}$, where p_{\min} is the minimum of the population response, and p_{\max} is the maximum of the population response. Note the similarity between this equation and the definition of surface anisotropy.⁴

⁶It should be noted that image orientations cannot differ by more than 90 deg. This leads to a residual correlation between peak orientations (the hue dimension of the orientation maps), such that even for randomly generated distributions $r^2 = 0.5$. To accommodate for this residual correlation, we can normalize the r^2 scale so that it runs from 0 to 1 instead of 0.5 to 1. We then find that on average pairs of images that contained the same shape rendered in different scenes lead to population peaks that were correlated with a modified r^2 of 0.84. Conversely, pairs of images that consisted of different shapes rendered under the same scene lead to population peaks that were correlated with a modified r^2 of 0.58.

⁷The idea that the visual system can achieve perceptual constancy by making image measurements that remain stable across changes in the viewing conditions has a long tradition, and was advocated particularly strongly by Gibson (1950a, 1979). When available, this is an elegant strategy for visual perception. However, we do not mean to suggest that all problems in vision can be solved in this way, nor that the visual system *never* estimates the light field. We are simply arguing that under our circumstances, the visual system does not need to estimate the illumination to recover certain information about 3D shape from specular reflections.

⁸The authors wish to credit Barton L. Anderson with this observation.

⁹To create the textured surfaces, we generated blocks of homogeneous texture, and carved the 3D surfaces out of these textures. To create the glossy surfaces, we carved a sphere out of each block of texture. We then treated the pattern on this sphere as if it were a standard light probe illuminating a mirrored object (i.e., the pattern was treated as light arriving from an infinite sphere).

References

- Adelson, E. H. (1999). Lightness perception and lightness illusions. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 339-351). Cambridge, MA: MIT Press.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, 26(4), 419-453. [PubMed]
- Beck, J., & Prazdny, S. (1981). Highlights and the perception of glossiness. *Perception and Psychophysics*, 30(4), 407-410. [PubMed]
- Ben-Shahar, O., & Zucker, S. (2001). On the perceptual organization of texture and shading flows: From a geometrical model to coherence computation. In *Proceedings of CVPR*, (pp. 1048-1055), Kawaii, HI.
- Blake, A., & Brelstaff, G. (1988). Geometry from specularities. In *Proceedings of ICCV* (pp. 394-403), Tampa, FL.
- Blake, A., & Bülthoff, H. H. (1990). Does the brain know the physics of specular reflection? *Nature*, 343, 165-168. [PubMed]
- Blake, A., & Bülthoff, H. H. (1991). Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society Series B*, 331, 237-252. [PubMed]
- Blake, A., & Marinos, C. (1990). Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence*, 45, 323-380.
- Breton, P., & Zucker, S. W. (1996). Shadows and shading flow fields. In *Proceedings of CVPR* (pp. 782-789), San Francisco, CA.
- Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture, and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, 33, 919-934. [PubMed]
- Clerc, M., & Mallot, S. (2002). The texture gradient equation for recovering shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 536-549.
- Cumming, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, 33, 827-838. [PubMed]
- Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology*, 113(2), 198-216. [PubMed]
- Debevec, P. E. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proceedings of SIGGRAPH 1998*, 189-198.

- Debevec, P. E., Hawkins, T., Tchou, C., Duiker, H. -P., Sarokin, W., & Sagar, M. (2000). Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000*, 145-156.
- De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22, 531-544. [PubMed]
- Dror, R. O. (2002). *Surface reflectance recognition and real-world illumination statistics* (AI Lab Technical Report No. AITR-2002-009). Cambridge, MA: MIT Artificial Intelligence Laboratory. [Article]
- Dror, R. O., Leung, T., Willsky, A. S., & Adelson, E. H. (2001). Statistics of real-world illumination. In *Proceedings of CVPR*, 2, 164-171.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12), 2379-2394. [PubMed]
- Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5), 347-368, <http://journalofvision.org/3/5/3/>, doi:10.1167/3.5.3. [PubMed][Article]
- Gibson, J. J. (1950a). *The perception of the visual world*. Boston: Houghton Mifflin.
- Gibson, J. J. (1950b). The perception of visual surfaces. *American Journal of Psychology*, 63, 367-384.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Hartung, B., & Kersten, D. (2002). Distinguishing shiny from matte [Abstract]. *Journal of Vision*, 2(7), 551a, <http://journalofvision.org/2/7/551/>, doi:10.1167/2.7.551.
- Hartung, B., & Kersten, D. (2003). How does the perception of shape interact with the perception of shiny material? [Abstract] *Journal of Vision*, 3(9), 59a, <http://journalofvision.org/3/9/59/>, doi:10.1167/3.9.59.
- Heider, G. M. (1933). New studies in transparency, form and color. *Psychologische Forschung*, 17, 13-56.
- Hill, H., & Bruce V. (1993). Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion. *Perception*, 22, 887-897. [PubMed]
- Hill, H., & Bruce, V. (1994). A comparison between the hollow-face and 'hollow-potato' illusions. *Perception*, 23, 1335-1337. [PubMed]
- Helmholtz, H. (1962). *Helmholtz's treatise on physiological optics*. New York: Dover. (Original work published in 1867)
- Howard, I. (1983). Occluding edges in apparent reversal of convexity and concavity. *Perception*, 12, 85-86. [PubMed]
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology London*, 148, 574-591. [PubMed]
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology London*, 160, 106-154. [PubMed]
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology London*, 195, 215-243. [PubMed]
- Huggins, P. S., Chen, H. F., Belhumeur, P. N., & Zucker, S. W. (2001). Finding folds: On the appearance and identification of occlusion, in CVPR'01. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2, 718-725.
- Kardos, L. (1934). Ding und Schatten: Eine experimentelle Untersuchung. In *Zeitschrift für Psychologie*. Leipzig: Barth.
- Klinker, G. J., Shafer, S. A., & Kanade, T. (1998). The measurement of highlights in color images. *International Journal of Computer Vision*, 2, 7-32.
- Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, 13, 321-330. [PubMed]
- Koenderink, J. J., & van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Optica Acta*, 27(7), 981-996.
- Koenderink, J. J., & van Doorn, A. J. (2003). Shape and shading. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1090-1105). Cambridge: MIT Press.
- Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1992). Surface perception in pictures. *Perception and Psychophysics*, 52, 487-496. [PubMed]
- Koffka, K. (1935). *Principles of Gestalt psychology*. Cleveland: Harcourt, Brace and World.
- Langer M. S., & Bülthoff, H. H. (2001). A prior for global convexity in local shape from shading. *Perception*, 30(4), 403-410. [PubMed]
- Li, A., & Zaidi, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Research*, 40(2), 217-242. [PubMed]
- Li, A., & Zaidi, Q. (2003). Observer strategies in perception of 3-D shape from isotropic textures: Developable surfaces. *Vision Research*, 43, 2741-2758. [PubMed]
- Malik, J., & Rosenholtz, R. (1997). Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 23, 149-168.

- Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research*, 38, 2817-2832. [\[PubMed\]](#)
- Mamassian, P., & Kersten, D. (1993). Surface orientation and illumination direction from shading. *Investigative Ophthalmology & Visual Science*, 34, 1082.
- Mamassian, P., & Kersten, D. (1996). Illumination, shading and the perception of local orientation. *Vision Research*, 36, 2351-2367. [\[PubMed\]](#)
- Metelli, F. (1974). The perception of transparency. *Scientific American*, 230(4), 90-98. [\[PubMed\]](#)
- Mingolla, E., & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, 53, 137-151. [\[PubMed\]](#)
- Norman, J. F., Todd, J. T., & Orban, G. A. (2004). Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science*, 15(8), 565-570. [\[PubMed\]](#)
- Oren, M., & Nayer, S. K. (1996). A theory of specular surface Geometry. *International Journal of Computer Vision*, 24, 105-124.
- Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317, 314-319. [\[PubMed\]](#)
- Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, 331(14), 163-166. [\[PubMed\]](#)
- Ramachandran, V. S. (1990). Perceiving shape from shading. In I. Rock (Ed.), *The perceptual world* (pp. 127-138). New York: W. H. Freeman. [\[PubMed\]](#)
- Rosenholtz, R., & Malik, J. (1997). Surface orientation from texture: isotropy or homogeneity (or both)? *Vision Research*, 37, 2283-2293. [\[PubMed\]](#)
- Savarese, S., Li, F. F., & Perona, P. (2003). Can we see the shape of a mirror? [\[Abstract\]](#) *Journal of Vision*, 3(9), 74a, <http://journalofvision.org/3/9/74/>, doi:10.1167/3.9.74.
- Savarese, S., Li, F. F., & Perona, P. (in press). What do reflections tell us about the shape of a mirror? *Proceedings of First Symposium on Applied Perception in Graphics and Visualization*. New York, NY: ACM Press.
- Savarese, S., & Perona, P. (2001). Local analysis for 3d reconstruction of specular surfaces. *Proceedings of CVPR*, 2, 738-745.
- Savarese, S., & Perona, P. (2002). Local analysis for 3d reconstruction of specular surface. Part II. In A. Heyden, G. Sparr, M. Nielsen, & P. Johansen (Eds.), *Computer Vision - ECCV 2002, 7th European Conference on Computer Vision* (pp. 759-774). Berlin: Springer-Verlag.
- Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *Journal of Neurophysiology*, 39, 1320-1333. [\[PubMed\]](#)
- Simoncelli, E. P., & Freeman, W. T. (1995). The steerable pyramid: A flexible architecture for multi-scale derivative computation. *IEEE Second Int'l Conference on Image Processing*, 3, 444-447.
- Simoncelli, E. P., Freeman, W. T., W. T., Adelson, E. H., & Heeger, D. J. (1992). Shiftable multi-scale transforms [or, "What's Wrong with Orthonormal Wavelets"]. *IEEE Transactions on Information Theory, Special Issue on Wavelets*, 38(2), 587-607.
- Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review*, 109(3), 492-519. [\[PubMed\]](#)
- Stevens, K. A. (1981). The information content of texture gradients. *Biological Cybernetics*, 42, 95-105. [\[PubMed\]](#)
- Stevens, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, 46, 183-195. [\[PubMed\]](#)
- Super, B., & Bovic, A. (1995). Shape from texture using local spectral moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, 333-343.
- Symons, L. A., Cuddy, F., & Humphrey, K. (2000). Orientation tuning of shape from shading. *Perception and Psychophysics*, 62, 557-568. [\[PubMed\]](#)
- Todd, J. T., & Akerstrom, R. A. (1987). The perception of three dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2), 242-255. [\[PubMed\]](#)
- Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illuminant from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 583-595.
- Todd, J. T., Norman, J. F., Koenderink, J. J., & Kappers, A. M. L. (1997). Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception*, 26, 807-822. [\[PubMed\]](#)
- Todd, J. T., Oomes A. H., Koenderink, J. J., & Kappers, A. M. L. (2004). The perception of doubly curved surfaces from anisotropic textures. *Psychological Science*, 15(1), 40-46. [\[PubMed\]](#)
- Ullman, S. (1976). On visual detection of light sources. *Biological Cybernetics*, 21, 205-211. [\[PubMed\]](#)
- Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. *Proceedings of SIGGRAPH 1994*, 459-472.
- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17, 17-45.

- Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology*. New York: Holt, Rinehart, and Winston.
- Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research*, 41, 2581-2600. [PubMed]
- Zaidi, Q., & Li, A. (2002). Limitations on shape information provided by texture cues. *Vision Research*, 42(7), 815-835. [PubMed]