

Research Article

Perception of Three-Dimensional Shape From Specular Highlights, Deformations of Shading, and Other Types of Visual Information

J. Farley Norman,¹ James T. Todd,² and Guy A. Orban³

¹Department of Psychology, Western Kentucky University; ²Department of Psychology, Ohio State University; and

³Laboratorium voor Neuro-en Psychofysiologie, Medical School, Katholieke Universiteit Leuven, Leuven, Belgium

ABSTRACT—*There have been numerous computational models developed in an effort to explain how the human visual system analyzes three-dimensional (3D) surface shape from patterns of image shading, but they all share some important limitations. Models that are applicable to individual static images cannot correctly interpret regions that contain specular highlights, and those that are applicable to moving images have difficulties when a surface moves relative to its sources of illumination. Here we describe a psychophysical experiment that measured the sensitivity of human observers to small differences of 3D shape over a wide variety of conditions. The results provide clear evidence that the presence of specular highlights or the motions of a surface relative to its light source do not pose an impediment to perception, but rather, provide powerful sources of information for the perceptual analysis of 3D shape.*

There are many different aspects of the physical environment that can affect the pattern of light intensity within a visual image. Changes in image intensity can sometimes be quite abrupt; this can be the case, for example, at an object's occlusion boundary or at reflectance edges on textured surfaces. Other variations can occur more gradually. For example, when a matte surface scatters light diffusely in all possible directions (see Fig. 1), the luminance in each local region is determined by the relative orientation between the surface and its direction of illumination. This produces a pattern of diffuse shading that varies gradually as a function of surface curvature. For shiny surfaces, light

is reflected within a more limited range of directions, much like the carom of a billiard ball. The luminance of a shiny surface can change very rapidly as a function of its orientation relative to the light source and the point of observation, which produces local regions of high image intensity called specular highlights.

The fact that variations in image intensity can have many different environmental causes is theoretically important because each distinct type of visual feature can behave quite differently as a function of changing viewing conditions. As a consequence, most computational analyses for obtaining three-dimensional (3D) shape from 2D image data have adopted a modular approach in which a single type of visual feature is considered in isolation. For example, numerous algorithms have been developed for computing aspects of 3D shape from smooth occlusion contours (Koenderink, 1984; Koenderink & van Doorn, 1982; Malik, 1987), patterns of texture (e.g., Malik & Rosenholtz, 1997), or gradients of diffuse shading (e.g., Horn & Brooks, 1989; Stewart & Langer, 1997). Specular highlights, in contrast, have received relatively little attention. Most existing algorithms for the analysis of 3D shape from shading are designed explicitly for surfaces with diffuse reflectance functions, and are therefore incapable of correctly interpreting regions of an image that contain specular highlights. Indeed, some researchers have speculated that it may not be possible to compute 3D shape from specular highlights in individual static images (Oren & Nayar, 1997), except perhaps in highly constrained contexts (Savarese & Perona, 2002).

The analysis of 3D structure from visual information is often much easier when multiple images are available because of motion or binocular vision (e.g., Koenderink & van Doorn, 1991; Ullman, 1979). A fundamental assumption that is generally employed in the analysis of multiple images is that visual features must projectively correspond to fixed locations on an object's surface. Although this assumption is

Address correspondence to James T. Todd, Department of Psychology, Ohio State University, Columbus, OH 43210; e-mail: todd.44@osu.edu.

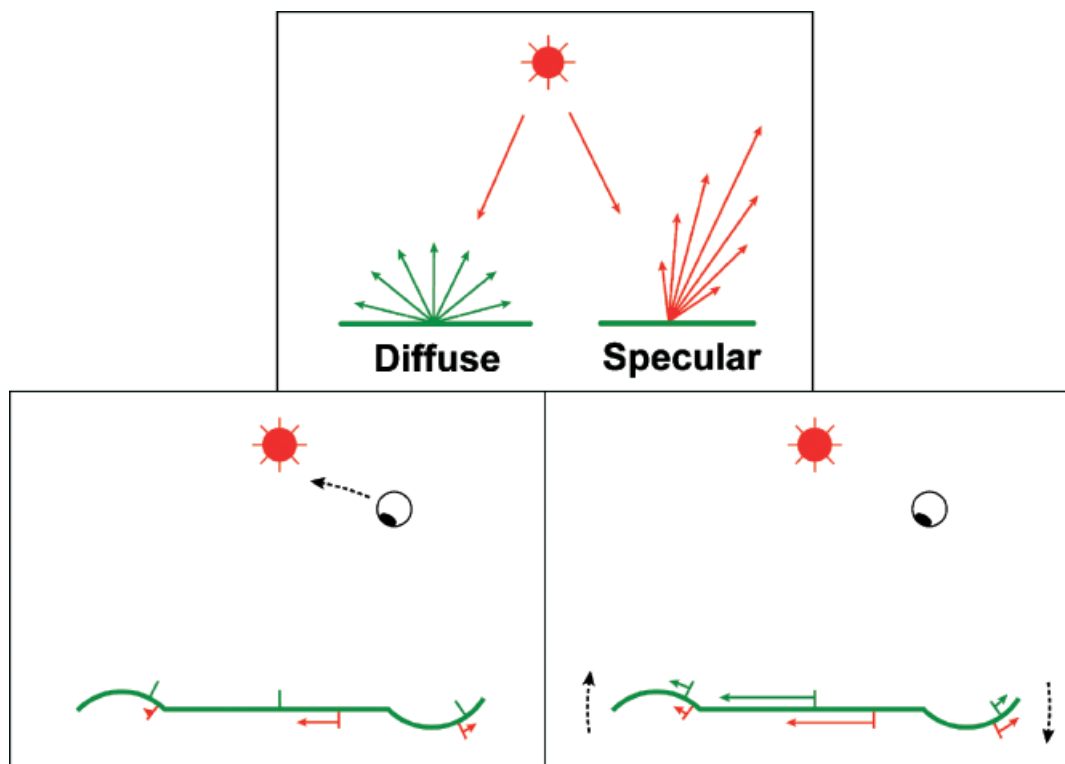


Fig. 1. Diffuse and specular reflections. The upper panel shows how these two types of reflection differ in the pattern of light scattering, as well as in their chromatic structure: Diffuse reflections are the color of the surface, whereas specular reflections are the color of the light source. The lower left panel shows what happens to patterns of shading as an observer moves within a stationary visual scene. The solid green line represents the cross section of a curved surface, the green bars mark the local maxima of diffuse shading, the red bars mark the local maxima of specular shading, and the arrows show how these features are displaced because of motions of the observer. As shown, the local maxima of diffuse shading remain at fixed locations on the object's surface, but the pattern of specular highlights is systematically deformed. The direction of highlight displacement varies with the sign of surface curvature, and the magnitude of displacement is negatively related to the magnitude of curvature. The lower right panel shows what happens to these different types of shading as a surface undergoes a comparable motion relative to a stationary observer and a stationary source of illumination. Whereas local maxima of diffuse shading remain fixed during observer motion, object motion causes them to deform in a manner that is qualitatively similar to the deformations of specular highlights. Highlight displacements from object motion are qualitatively similar to those from observer motion, though the magnitudes of displacement are much larger.

satisfied for the motions or binocular disparities of textured surfaces, it is often strongly violated for other types of visual features. For example, when a smoothly curved object rotates in depth, the locus of surface points that defines its occlusion contour changes continuously over time (Cipolla & Giblin, 1999; Giblin & Weiss, 1987). Gradients of diffuse shading are especially interesting in this context. When an observer moves relative to a fixed scene, the shading at each surface location remains constant. However, when an object moves relative to its sources of illumination, the shading at each point changes continuously (see Fig. 1). As a consequence of this behavior, current techniques for computing 3D shape from deformations of diffuse shading are applicable only for motions of an observer (or camera) within a fixed scene (Horn & Schunck, 1981; Nagel, 1981, 1987). A similar distinction is also applicable to the deformations of specular highlights. Several models have been developed for analyzing the deformations of highlights due to motions of the observer (Blake & Bülthoff, 1990, 1991; Oren & Nayar, 1997; Zisserman, Giblin, & Blake, 1989), but these models have not been generalized for the motions of a surface relative to its sources of illumination.

Empirical research on the visual perception of 3D shape has generally adopted the same modular approach as in theoretical analyses, using stimuli that contain a single type of visual feature presented in isolation. For example, almost all existing studies on the perception of 3D shape from shading have used uniformly colored surfaces with diffuse reflectance functions. Similarly, most investigations of the perception of 3D shape from motion or binocular disparity have used textured surfaces without any shading that satisfy the assumption of projective correspondence. There are a few exceptions to this trend in which researchers have investigated the perception of 3D shape from the deformations of smooth occlusion boundaries (Norman & Todd, 1994; Norman & Raines, 2002) or the binocular disparities of specular highlights (Blake & Bülthoff, 1991; Todd, Norman, Koenderink, & Kappers, 1997). In general, however, there has been little effort in the field to investigate those stimulus configurations that pose the greatest difficulties for current computational models.

The research described in the present article was designed to investigate the precision of 3D shape discriminations over a much broader range of conditions than has been examined previously. Ob-

servers were presented with two different randomly shaped objects in successive intervals, and they were required to judge whether the global 3D shapes of those objects were the same or different. The objects were depicted with four different types of visual features presented in various combinations: smooth occlusion contours, surface texture, gradients of diffuse shading, and specular highlights (see Fig. 2). The objects could be presented in a stationary pose or undergoing rotation in depth, and they could be observed either monocularly or stereoscopically (example Quicktime 6 videos from each of the motion conditions can be downloaded at www.psy.ohio-state.edu/faculty/todd/JToddMovies.htm). A critical aspect of the experimental design is that the 3D orientation of the depicted object, its direction of illumination, its axis of rotation, and the randomization of its texture were varied randomly across successive intervals, so that the task could not be performed accurately by a direct comparison of the 2D images (see Fig. 3).

METHOD

Stimuli

The stimuli in this experiment depicted randomly shaped objects similar to those used in earlier studies (see Norman & Todd, 1996; Todd & Norman, 1995; Todd et al., 1997). Each object was approximately 9 cm in diameter in any given direction and was defined as a dense mesh of 8,192 triangular polygons. The objects could be rendered with various combinations of shading and texture. Shading was created using the standard OpenGL reflectance model, in which image

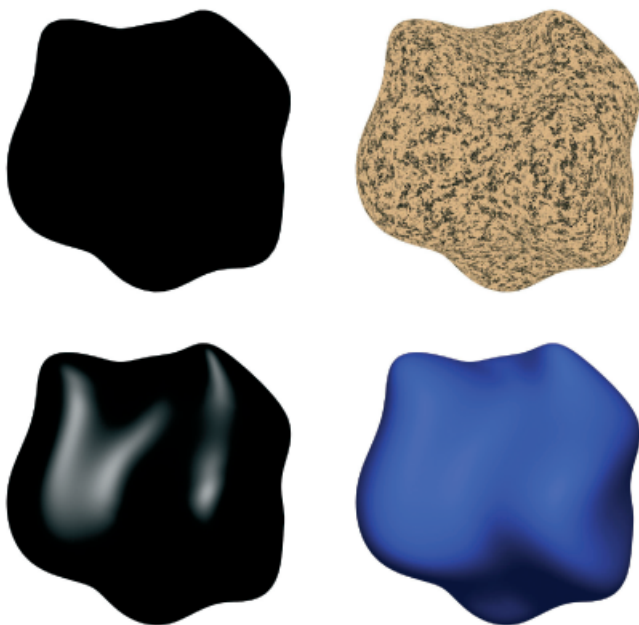


Fig. 2. Images illustrating four different types of visual features from the monocular static conditions in the present study. Moving clockwise from the upper left, the images depict a smooth occlusion contour presented in silhouette, an object with a texture resembling red granite, an object with diffuse shading, and an object with specular highlights. Other stimulus conditions included objects with texture and diffuse shading, objects with diffuse shading and specular highlights, and specular highlights presented in isolation. This last condition was created by presenting the object against a black background. Example videos from each of the motion conditions can be downloaded at www.psy.ohio-state.edu/faculty/todd/JToddMovies.htm.

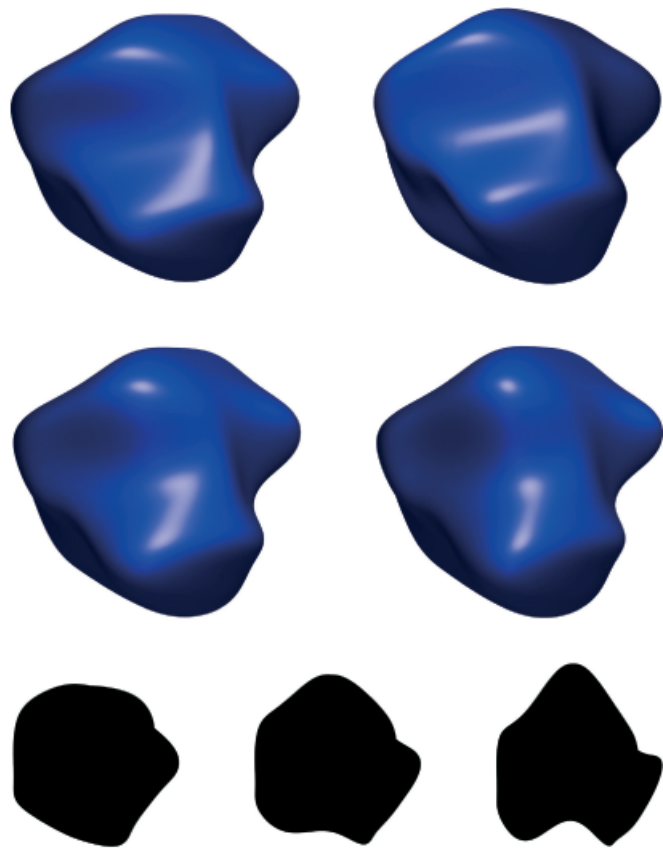


Fig. 3. Some possible stimulus objects from the monocular static condition with shading, highlights, and occlusions. The images shown in the top row depict objects with identical shapes, but with different orientations and directions of illumination. The two images in the middle row show this same base object with sinusoidal perturbations having amplitudes of 1 cm (left) and 2 cm (right). The three images in the bottom row provide a clearer perspective of how the perturbations altered the overall patterns of the three-dimensional shapes by showing how an object would appear in silhouette if viewed from above so that the direction of displacement is parallel to the image plane. The image on the lower left shows an untransformed base object, and the images in the middle and on the right show perturbations of this object with amplitudes of 2 cm and 4 cm, respectively. A perturbation of 2 cm was the approximate threshold in the most difficult conditions, when the occlusion contour was presented with static texture or with no other sources of information.

intensity is determined as an additive sum of ambient, diffuse, and specular components. The objects were illuminated at a fixed slant of 30° and a tilt that varied randomly over a 360° range. The texture patterns were generated from an image of red granite. Each polygon in the triangular mesh was positioned at random within this image to define the polygon's individual texture pattern. This ensured that the pattern of texture on the depicted surface was statistically homogeneous and isotropic. In the moving conditions, the objects oscillated back and forth in depth over a 56° range at a rate of $87.5^\circ/\text{s}$. The slant of the rotation axis varied randomly across trials over a range from 0° to 30° .

The stimuli were rendered in real time on an Apple Power Macintosh Dual-Processor G4 with OpenGL and hardware graphics acceleration (Nexus 128, ATI Technologies, Inc., Markham, Ontario, Canada). The displays were presented at a viewing distance of 1 m on a Mitsubishi Diamond Plus 200 22-in. flat-screen monitor with a

spatial resolution of 1280×1024 pixels. The observers wore CrystalEyes2 LCD shuttered glasses (Stereographics, Inc., San Rafael, CA), which alternated stereoscopic images in the left and right eyes at a 60-Hz refresh rate. In the monocular conditions, observers wore a patch over one eye.

The complete experimental design included 28 different conditions formed by the orthogonal combination of three variables: 2 stereo conditions (stereoscopic vs. monocular presentations) \times 2 motion conditions (3D rotation in depth vs. static) \times 7 combinations of image features (occlusions only; texture and occlusions; texture, diffuse shading, and occlusions; diffuse shading and occlusions; diffuse shading, specular highlights, and occlusions; specular highlights and occlusions; and specular highlights only).

Procedure

On each trial, an observer was presented with two objects in successive 1.2-s intervals and was required to judge whether the global 3D shapes of those objects were the same or different. The variations of 3D shape on different-shape trials were created by adding a vertically oriented sinusoidal corrugation to the original random shape. That is, each vertex was displaced in depth, such that the relative magnitude of displacement among different vertices varied as a sinusoidal function of their horizontal positions (see Fig. 3). The period of this sinusoidal perturbation was 5 cm, and its amplitude was systematically adjusted using an adaptive PEST (parameter estimation by sequential testing) staircase procedure (Taylor & Creelman, 1967) in order to determine in each condition a threshold at which the observer's responses were 80% accurate. Various manipulations prevented subjects from basing their responses on 2D image structure: The relative 3D slant of objects presented in successive intervals varied randomly over a 9° range; in the motion conditions, the slant of the axes of rotation in successive intervals varied randomly over a 9° range; in the shading conditions, the illumination tilt in successive intervals varied randomly over a 40° range; and, finally, in the texture conditions, the objects presented in successive intervals had different randomizations of texture. These different manipulations occurred simultaneously whenever the appropriate stimulus attributes were present.

The illustrations in the top row of Figure 3 depict a typical base object in the shading-plus-highlights condition, with different orientations and directions of illumination. The two illustrations immediately below show transformed versions of this same object with different amplitudes of sinusoidal perturbation. The three black illustrations at the bottom of the figure show how an untransformed base object and two perturbations of it would appear if viewed from above, so that the overall pattern of 3D shape change is more clearly visible.

Five different observers participated in the experiment, 2 of the authors (J.F.N. and J.T.T.) and 3 other observers who were naive. The naive observers were given no information about any details of the experimental design or the precise nature of the shape changes they were required to detect. Each subject received a random sequence of the 28 possible display conditions in separate blocks of trials. After all of these conditions had been completed, the same sequence was repeated again in reverse order.

RESULTS AND DISCUSSION

The average shape-discrimination thresholds of the 5 observers are presented in Figure 4. Because performance was generally compa-

rable for both stereo and monocular objects presented with motion and for static stereo objects, the thresholds in those conditions have been collapsed into a single category that is labeled in the figure as "multiple images." It is clear from these data that the shape-discrimination thresholds varied dramatically over a fourfold range across the different conditions. As is evident in the figure, performance was lowest for the single-image displays with texture and no shading and the displays that contained occlusion contours presented in isolation. An analysis of variance using orthogonal comparisons revealed that those conditions produced significantly higher thresholds than the remaining conditions, $F(1, 108) = 855.25$, $p < .001$; this difference accounted for more than 92% of the between-display variance. Among the remaining conditions, there was also a significant reduction in performance for the single-image displays with diffuse shading and no specular highlights, $F(1, 108) = 46.57$, $p < .001$; this difference accounted for another 5% of the variance. No other orthogonal comparisons were statistically significant. That is, the multiple-image displays with shading or texture and all of the displays with specular highlights produced comparable levels of shape-discrimination accuracy.

One important issue in evaluating these results is the extent to which successful performance could have been achieved on the sole basis of changes in 2D image structure without the perceptual analysis of 3D shape. In an effort to confirm whether the randomization procedures designed to prevent such a strategy were successful, we calculated the number of changed pixels across the two intervals on same-shape and different-shape trials for a random sample of objects with threshold perturbation magnitudes in each of the seven single-image conditions. In almost all cases, the average number of changed pixels for same-shape trials was within 2% of the average number of changed pixels for different-shape trials. The one salient exception was the contour-only condition, for which the average number of changed pixels was 14% smaller on same-shape trials than on different-shape trials. It is important to keep in mind, however, that the contour-only condition produced the lowest levels of performance, which suggests quite strongly that the observers' judgments could not have been based on a simple comparison of 2D image structures.

The most theoretically surprising aspect of these results is that the highest levels of performance were achieved for the displays that contained specular highlights—even when no other sources of information were available. The perceptual information provided by highlights is most likely based on the fact that specular reflections diminish quite rapidly as a function of surface orientation. Thus, the extent of a highlight in any given direction is negatively related to the magnitude of curvature in that direction. For example, in the lower left object in Figure 2, the extension of the highlights indicates the presence of two vertically oriented ridges. Similar information is provided by the deformations of highlights when objects are observed stereoscopically or in motion. As was first noted by Koenderink and van Doorn (1980), highlights cling to regions of high curvature: Their relative displacements in different local regions are negatively related to the magnitudes of curvature in those regions, and the directions of their displacements are determined by the sign of surface curvature (see also Blake & Bülthoff, 1990, 1991; Oren & Nayar, 1997; Zisserman et al., 1989). These earlier analyses were restricted to motions of an observer within a fixed visual scene, but the overall pattern of highlight deformations is qualitatively similar when an object moves relative to its sources of illumination (see Fig. 1).

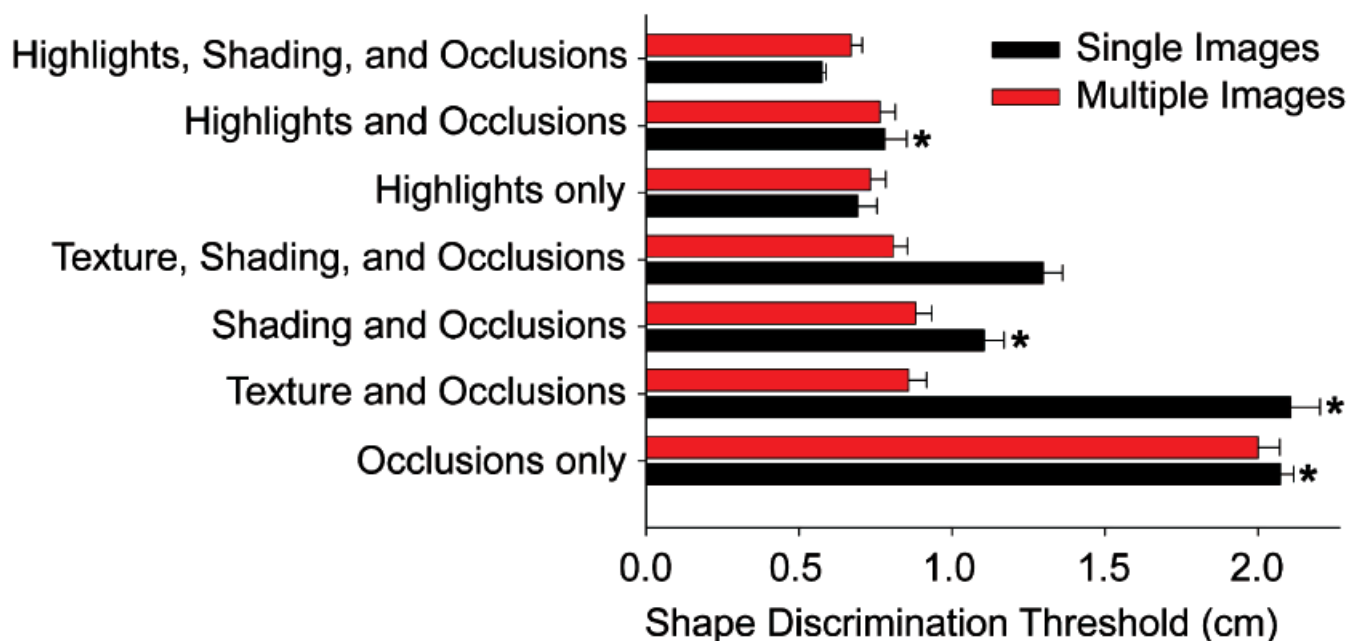


Fig. 4. The average shape-discrimination thresholds of the 5 observers for the seven possible combinations of visual features employed in the present experiment. The results obtained for both stereo and monocular objects presented with motion and for static stereo objects have been collapsed into a single “multiple images” category. Thus, each single-image threshold is an average of 10 PEST (parameter estimation by sequential testing) staircases, and each multiple-image threshold is an average of 30 PEST staircases. The four conditions illustrated in Figure 2 are marked with asterisks. Error bars indicate the standard errors of the mean.

It is interesting to note that adding motion or binocular disparity to the displays significantly improved performance for just three of the possible combinations of image features. These improvements with multiple images were greatest for the displays that contained random-noise textures. This result would be expected on the basis of current theory, because current computational analyses of 3D structure from motion can produce correct interpretations only for surfaces that are textured (e.g., Koenderink & van Doorn, 1991; Ullman, 1979). A more theoretically surprising finding is that performance was significantly improved when the diffusely shaded objects were presented in motion. Although there are some algorithms for determining 3D shape from optical deformations of diffuse shading (Horn & Schunck, 1981; Nagel, 1981, 1987), these algorithms are all designed for motions of an observer within a fixed visual environment, and would therefore produce erroneous results for objects that move relative to their sources of illumination, as in the present experiment. Our results provide strong evidence, however, that these deformations of diffuse shading provide useful information for human perception. One possible source of that information is that local maxima of diffuse shading deform in a manner that is qualitatively similar to the deformations of specular highlights (see Fig. 1): That is, their directions of motion vary with the sign of surface curvature, and the magnitudes of their displacements are negatively related to the magnitude of curvature.

Although there were no effects of motion for the three conditions that included specular highlights, this was most likely due to a ceiling effect, given that performance was so high for the static monocular presentations of those displays. During their debriefing sessions, all of the observers reported that the moving surfaces with specular highlights all appeared to be rigidly rotating in depth—even when no other sources of information were available. This perception of rigid motion is theoretically quite remarkable, because there are no known methods of analysis

that could correctly interpret these displays without any prior knowledge about the position of the camera or the direction of illumination.

The ability of human observers to accurately detect small variations in 3D shape from specular highlights or deformations of shading cannot be explained by existing computational models for determining 3D structure from visual information. One possible explanation why perceptual judgments are so surprisingly robust is that they may be based on a weaker type of data structure than is typically employed by most computational models. Because patterns of image shading are inherently ambiguous (Belhumeur, Kriegman, & Yuille, 1999), it is not mathematically possible to obtain a unique metric interpretation of an observed scene without incorporating additional constraints. There is a growing amount of evidence to suggest, however, that human perception is often based on more qualitative aspects of 3D structure, such as affine, ordinal, or topological relations (Todd & Norman, 2003; Todd & Reichel, 1989), and there is also evidence to indicate that these qualitative aspects of structure may be encoded by neurons within the shape-processing regions of the visual cortex (Janssen, Vogels, & Orban, 2000).

What is the information by which these qualitative aspects of 3D structure are perceptually specified? In an influential early article, Koenderink and van Doorn (1976) provided a formal analysis of how the qualitative structures of smoothly curved surfaces can be determined from the topological arrangement of a special set of features that include local depth extrema as well as discontinuities and changes in the sign of curvature along occlusion contours. Over small changes in viewing direction, the topological structure of these features generally remains quite stable, though it is also possible for this structure to change abruptly, such that new features can suddenly appear or disappear. These transitions are highly constrained, however, and they can occur in only a few possible ways that have been

exhaustively enumerated (see also Cipolla & Giblin, 1999). Thus, if human observers had knowledge of those constraints through experience or evolution, they might be able to predict with reasonable accuracy the types of image changes that are likely to occur because of variations in viewing direction, and to distinguish those changes from others that may result from an overall distortion of 3D shape.

Subsequent research has attempted to extend this type of analysis to include the behavior of specular highlights (Blake & Bülthoff, 1990, 1991; Koenderink & van Doorn, 1980; Oren & Nayar, 1997; Zisserman et al., 1989). This research has shown, for example, that the appearance or disappearance of specular points always occurs in pairs at points on a surface that have no curvature in one direction. In light of the fact that human observers can identify the rigid motions of surfaces from deformations of highlights, it is likely to be the case that these deformations are sufficiently constrained to distinguish them from nonrigid shape changes. Although the precise nature of these constraints has yet to be elaborated, the remarkable performance of observers in the present experiment suggests this may be a fruitful area for future theoretical analyses.

Acknowledgments—James Todd's participation in this research was supported by grants from the National Institutes of Health (R01-Ey12432) and National Science Foundation (BCS-0079277).

REFERENCES

- Belhumeur, P.N., Kriegman, D.J., & Yuille, A.L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1), 33–44.
- Blake, A., & Bülthoff, H.H. (1990). Does the brain know the physics of specular reflection? *Nature*, 343, 165–168.
- Blake, A., & Bülthoff, H.H. (1991). Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society of London B*, 331, 237–252.
- Cipolla, R., & Giblin, P. (1999). *Visual motion of curves and surfaces*. Cambridge, England: Cambridge University Press.
- Giblin, P., & Weiss, R. (1987, June). *Reconstruction of surfaces from profiles*. Paper presented at the First International Conference on Computer Vision, London.
- Horn, B.K.P., & Brooks, M.J. (1989). *Shape from shading*. Cambridge, MA: MIT Press.
- Horn, B.K.P., & Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17, 185–203.
- Janssen, P., Vogels, R., & Orban, G.A. (2000). Three-dimensional shape coding in inferior temporal cortex. *Neuron*, 27, 385–397.
- Koenderink, J.J. (1984). What does the occluding contour tell us about solid shape? *Perception*, 13, 321–330.
- Koenderink, J.J., & van Doorn, A.J. (1976). The singularities of the visual mapping. *Biological Cybernetics*, 24, 51–59.
- Koenderink, J.J., & van Doorn, A.J. (1980). Photometric invariants related to solid shape. *Optica Acta*, 27, 981–996.
- Koenderink, J.J., & van Doorn, A.J. (1982). The shape of smooth objects and the way contours end. *Perception*, 11(2), 129–137.
- Koenderink, J.J., & van Doorn, A.J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, 8, 377–385.
- Malik, J. (1987). Interpreting line drawings of curved objects. *International Journal of Computer Vision*, 1, 73–103.
- Malik, J., & Rosenholtz, R. (1997). Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 23, 149–168.
- Nagel, H.H. (1981, August). *On the derivation of 3D rigid point configurations from image sequences*. Paper presented at the IEEE Conference on Pattern Recognition and Image Processing, Dallas, TX.
- Nagel, H.H. (1987). On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33, 299–324.
- Norman, J.F., & Raines, S.R. (2002). The perception and discrimination of local 3-D surface structure from deforming and disparate boundary contours. *Perception & Psychophysics*, 64, 1145–1159.
- Norman, J.F., & Todd, J.T. (1994). The perception of rigid motion in depth from the optical deformations of shadows and occlusion boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 343–356.
- Norman, J.F., & Todd, J.T. (1996). The discriminability of local surface structure. *Perception*, 25, 381–398.
- Oren, M., & Nayar, S.K. (1997). A theory of specular surface geometry. *International Journal of Computer Vision*, 24, 105–124.
- Savarese, S., & Perona, P. (2002). Local analysis for 3D reconstruction of specular surfaces - 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, Germany: Springer-Verlag.
- Stewart, A.J., & Langer, M.S. (1997). Towards accurate recovery of shape from shading under diffuse lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 1020–1025.
- Taylor, M.M., & Creelman, C.D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 41, 782–787.
- Todd, J.T., & Norman, J.F. (1995). The visual discrimination of relative surface orientation. *Perception*, 24, 855–866.
- Todd, J.T., & Norman, J.F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception & Psychophysics*, 65, 134–148.
- 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, 806–822.
- Todd, J.T., & Reichel, F.D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, 96, 643–657.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Zisserman, A., Giblin, P., & Blake, A. (1989). The information available to a moving observer from specularities. *Image and Video Computing*, 7, 38–42.

(RECEIVED 5/28/03; REVISION ACCEPTED 7/20/03)