Measuring Visual Shape using Computer Graphics Psychophysics

M. S. Langer H. H. Bülthoff
Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany
www.kyb.tuebingen.mpg.de/bu
{michael.langer, heinrich.buelthoff }@tuebingen.mpg.de

Abstract. This paper reviews recent psychophysical methods that have been developed for measuring the perceived shape of objects. We discuss two types of shape ambiguities that exist for many objects – a depth reversal ambiguity and an affine ambiguity. We show that people perceptually resolve these shape ambiguities by making strong prior assumptions the object.

1 Introduction

When we open our eyes and look at the objects around us, we typically feel confident that we can judge the 3-D shapes of the objects we see. The pattern of light that is reflected from these objects depends on several independent factors, however: the shape of the objects, the material of the objects and the light field surrounding the objects. The human visual systems is somehow able to disentangle these factors, and produce a coherent percept of an object from of an image. We would like to understand better how the visual system achieves such coherent percepts. To do so, we must develop methods for measuring what people actually perceive when they look at objects.

In this paper we are concerned mostly with the perception of object shape, rather than perception of lighting or material. The main question we address is how one can experimentally measure perceived shape using psychophysical methods. By "psychophysical," we mean that we treat a person's visual system as a black box: an instrument for measuring some physical property of the world, in this case the shape of objects. Using psychophysics, we would like to characterize this instrument in terms of its biases, noise properties, etc. Psychophysics is to be distinguished from methods that study the neural implementation of perception, such as functional magnetic resonance imaging (fMRI), electro– or magneto–encephalography (EEG or MEG), or single cell electrophysiology.

We invite people "off the street" – so–called "naive observers" – and ask them specific questions about the 3-D shapes that they see when they look at pictures of objects. Naive observers are usually willing to give an hour of their time for such experiments. Our challenge is, given that hour, which objects should we show the observers and which questions should we ask in order to measure the perceived shape? In this paper, we review some of the psychophysical approaches typically taken and some of the findings. We concentrate on methods that have used computer graphics - hence the phrase "computer graphics psychophysics."

Why is this psychophysical research relevant to computer graphics rendering? The basic answer is that, even though observers are remarkably good at perceiving object shape, they nevertheless suffer from certain fundamental limitations which are inherent in the vision problem, namely, a image of an object does not uniquely determine the shape, material, and illumination of that object. We can show using psychophysical methods that observers get around these limitations by making very strong prior assumptions about the object and scene, and stick to these assumptions even in the presence of contradictory image information. These assumptions are remarkably consistent from observer to observer. These results are relevant for computer graphics since, by understanding better the prior assumptions that observers make, we will be better able to render images in a manner that is consistent with these assumptions.

This paper consists of two parts. First, we review several psychophysical methods that have been used to measure perceived shape. Second, we discuss two inherent ambiguities in the perception of object shape and how observers resolve these ambiguities by making prior assumptions about what they are looking at.

2 Measuring perceived shape

2.1 Single point

A common method for measuring perceived shape is to mark a single point on a surface and to ask an observer about the shape of the surface at that point. Is the surface slanted to the right or to the left? Is the surface curved or flat? If it is curved, is it elliptical or hyperbolic [MKK96]? Is the point on a hill or in a valley [LBss]? Such judgments can be made very quickly, typically in one second or so even by naive observers, and so an observer can make such judgments at a rate of several thousand per hour.

2.2 Pair of points

Observers might be shown instead a pair of points and asked to discriminate the depth of these points, *i.e.* to judge which point is further away from the eye [TR89, KvDK96, LBss]. Pairs of points can also be used to measure how well observers can discriminate the relative orientation of two nearby points [TN95, NT96, RTY95].

2.3 Binocular depth probe

One can use a binocular probe to measure perceived depth directly. An observer may be presented with a rendered image monocularly, and a point probe binocularly. The binocular disparity of the probe – that is, the difference in the image positions of the probe in the two eyes — provides a depth cue for the observer [Gre70]. Observers are asked to judge whether the probe is in front of or behind the surface [SB87], or are asked to manipulate the perceived depth of the probe interactively until the probe appears to lie on the surface [BM88]. The binocular disparity value at which the probe appears to lie on the surface is then a direct measure of perceived depth at that point (see also [KKT$^+$96]). By sampling the perceived depth values over the surface, one obtains an estimate of the perceived depth map. This depth map may be thought of as a z–buffer (to borrow computer graphics jargon). It is a depth may that is registered with the intensity image.

One limitation of a binocular depth probe is that the probe may perceptually interact with the surface [BM88]. For example, when the depth of the probe is manipulated interactively, the probe can perceptually stick to the surface and stretch or compress the perceived surface in depth as the stereo disparity of the probe is varied. Such interactions between the probe and the image should be avoided if possible.

2.4 Depth gradient probe

Surface depth is an important property of perceived shape. However, depth information is often not directly available in the image. Cues such as texture and shading reveal information about the depth gradient of a surface rather than about the absolute depth [Ste83]. For example, under collimated lighting, shading depends on the local surface normal direction relative to the light field in which the surface is embedded.

One method for measuring the depth gradient on a surface is to show observers a *graphical probe* such as an ellipse. Observers are asked to fit this ellipse to the surface by imagining that the ellipse is a disk that is lying on the surface. The ellipse may be superimposed on the image [SB87] or it may be shown alongside the image [MT86, MLM97]. Using a mouse, the observer may manipulate the aspect ratio and the 2-D orientation of the ellipse until the perceived disk appears to be co-tangent to the surface [KvDK92]. The aspect ratio and orientation of the ellipse then provide a direct measure of the perceived depth gradient at that point. The perceived depth gradients may be sampled over the image and from these samples one can obtain an estimate of the global surface depth map via numerical integration.

While the depth gradient probe is a very useful tool for measuring perceived shape, it too has limitations. One limitation is that the probe itself might not be perceived correctly. That is, one cannot assume that the observer's settings of an ellipse provide a "readout" of the perceived depth gradient at a point. There are two issues here. First, perceiving the orientation of a graphical probe is itself a perceptual problem that the visual system must solve, and there is no reason to assume that every observer solves this problem correctly *e.g.* without systematic biases [MLM97]. Second, the perceived orientation of the probe may interact with the perceived orientation of the image, similar to the interaction found in the stereo probe case. These limitations need to be explored further before we can be certain of how to interpret the data obtained with such probes.

2.5 Global shape probe

An alternative method for measuring global surface shape is to show observers two surfaces and ask them to decide whether the shape of the two surfaces match [BM90, RB00]. The surfaces may be presented simultaneously and rendered with different visual cues, for example, with different albedo patterns or under different lighting. One surface might be shown monocularly and the another binocularly. Finally, one surface might be viewed from the front and compared to a profile slice [TM83, TA87].

3 Ambiguities in shape perception

Now that we have discussed some of the methods commonly used for measuring perceived shape, let us discuss some of the ambiguities in the shape perception problem and some of the strategies that the visual system uses to resolve these ambiguities.

Consider an object with Lambertian reflectance. We allow the albedo **a** to vary from point to point on the surface. Suppose the object is illuminated by a collimated light source in direction **L**. Let **N** denote the unit surface normals. The image **I** may be represented as the product,

$$\mathbf{I} = \mathbf{a}\,\mathbf{L}^T\,\mathbf{N}\,. \tag{1}$$

Observe that the right hand side of Equation (1) has three independent variables, whereas the left side has only a single variable. Given the variables on the right hand side, it is easy to compute the left hand side. This is the graphics problem. The vision problem is harder. Given the left hand side, the visual system tries to compute what it can about the variables on the right hand side.

It should be clear to the reader that it is impossible to solve the vision problem exactly since there are more unknown variables on the right hand side of Equation (1) than there are known values on the left hand side. And yet, the visual system somehow manages to compute something about the right hand side, since otherwise how would the visual system be able to judge object shape? Let us address what the visual system can compute about the right hand side by trying to understand the ambiguities that are present in Equation (1) and how the visual system resolves these ambiguities.

3.1 Ambiguity 1. Depth reversal

One well known ambiguity is that, if the surface depth map $z(x,y)$ is inverted,

$$z(x,y) \ \rightarrow \ -z(x,y)$$

and the light source direction $\mathbf{L} = (l_x, l_y, l_z)$ is reflected about the line of sight

$$(l_x,\ l_y,\ l_z) \ \rightarrow \ (-l_x,\ -l_y,\ l_z)\ ,$$

then the same image **I** is obtained. (We are assuming orthographic projection here.) This depth reversal ambiguity has been known for centuries [Rit86, Bre26].

The depth reversal ambiguity suggests that objects should flip-flop in depth. This is not what we perceive, however, when we look at typical objects. Rather, object shapes are typically perceived to be stable. What strategies does the visual system use to resolve this depth reversal ambiguity?

One strategy is to use information in the image that is not captured by Eq. (1). For example, cast shadows are not captured by Eq. (1) and can be used to determine the direction of the source and thereby resolve the depth reversal ambiguity [BBC84, EKK93]. A second source of information is binocular stereo. Even the sign of stereo disparity is sufficient for resolve the depth reversal ambiguity [HB93]. Familiarity with the object is an important factor. A hollow mask of a face will typically be seen incorrectly as a convex face [Luc16].

A second strategy is to make prior assumptions about the object or scene. For example, the visual system can make assumptions about the lighting. One well-known assumption is that the light source is above the line of sight rather than below the light of sight [Bre26, Rit86]. The assumption is a natural one since the sun is typically above the line of sight. Another assumption about the lighting, which applies for animation sequences, is that the light source is stationary [KMK97]. Other assumptions concern the surface geometry. For example, it has been shown that a floor orientations are preferred over ceiling orientation, that is, the visual system prefers to interpret the surface as if it is viewed from above rather than from below [RT90]. Again this is a natural assumption. We tend to see more floors than ceilings and we tend to see the tops of objects more than the bottoms of objects. A second assumption about surface geometry is that the surface bounds a globally convex object, *i.e.* the surface is the boundary of a solid object rather than the interior of a hollow mould or shell. This too is a natural assumption. Most objects seen in isolation are solid rather than shell-like.

Which of the many strategies do observers use to resolve the depth reversal ambiguity? Let us present an example of how we can address this question using computer graphics psychophysics. Figure 1 shows two rendered images of smooth bumpy surfaces. The surfaces have uniform Lambertian reflectance and are rendered using RADIANCE [War94]. One of the surfaces is a concave hemi-cylinder and the other is a convex hemi-cylinder. Although the depth reversal ambiguity applies to these surfaces in theory, the reader will note that the surfaces do not flip-flop in depth. Rather, there is a compelling sense of the shape of the surface, in particular, which points are local hills and which points are the local valleys. One can imagine answering questions about the depth gradient or the curvature at isolated points in the image. Although the depth reversal ambiguity implies that one should *not* be able to answer such questions, it is clear that we are able to answer them. What strategies does the visual system use to solve this shape perception problem?

Let us describe an experiment we recently carried out to answer this question. The experiment used many surfaces similar to those in Figure 1. Half of the surfaces were concave hemi-cylinders and the other half were convex hemi-cylinders. Each surface was rendered under two different lighting conditions. The source was either slightly above the line of sight or was slightly below the line of sight. Images were presented on a monitor in a dark room and viewed monocularly on a CRT from a distance of 80 cm. This provided the correct perspective. A small black probe was superimposed on each image and the observer was asked to judge whether the probe was on a local hill or in a local valley, one of these answers being correct in each case. Each observer made 512 of such judgments. To test the possible assumptions that observers were making to resolve the depth reversal ambiguity, we used three independent conditions. (1.) The light source was either above the line of sight or below it. (2.) The probe was either above or below the horizontal mid-line of the object, and hence was either on a floor–like or a ceiling–like region of the surface. (3.) The surface was either globally convex (a solid) or globally concave (a hollow shell). We balanced the conditions such that each of the eight combinations ($2 \times 2 \times 2$) was tested 64 times.

Percent correct scores are shown in Figure 1. We see that each of the three assumptions played a role. Observers scored highest if the light was from above, if the surface had a floor orientation near the probe, and if the surface was globally convex. Performance fell off as each of these assumptions was violated. It was interesting to note that performance overall was at chance (48 $\%$ correct). This indicates that shadows and perspective information were not used in the task. Rather, observers performed the task as if this information was not present in the image.

3.2 Ambiguity 2. Affine invariance

A second general ambiguity was discovered recently by researchers in computer vision [BKY97]. Consider the following affine transformation of a surface depth map,

$$z(x,y) \ \rightarrow \ \lambda \, z(x,y) \ + \ \mu \, x \ + \ \nu \, y \ , \tag{2}$$

where $\lambda > 0$. This transformation corresponds to a depth scaling plus an additive slanted plane. Depth is compressed if $\lambda \in (0,1)$ and expanded if $\lambda > 1$, and the μ and ν variables define the slant of the added depth plane.

When the depth map is transformed as above, the unit surface normals undergo a transformation point-wise,

$$\mathbf{N} \ \rightarrow \ \frac{1}{\| \, \mathbf{G} \, \mathbf{N} \, \|} \, \mathbf{G} \, \mathbf{N}$$

where

$$\mathbf{G} = \left(\begin{array}{ccc} \lambda & 0 & -\mu \\ 0 & \lambda & -\nu \\ 0 & 0 & 1 \end{array} \right) .$$

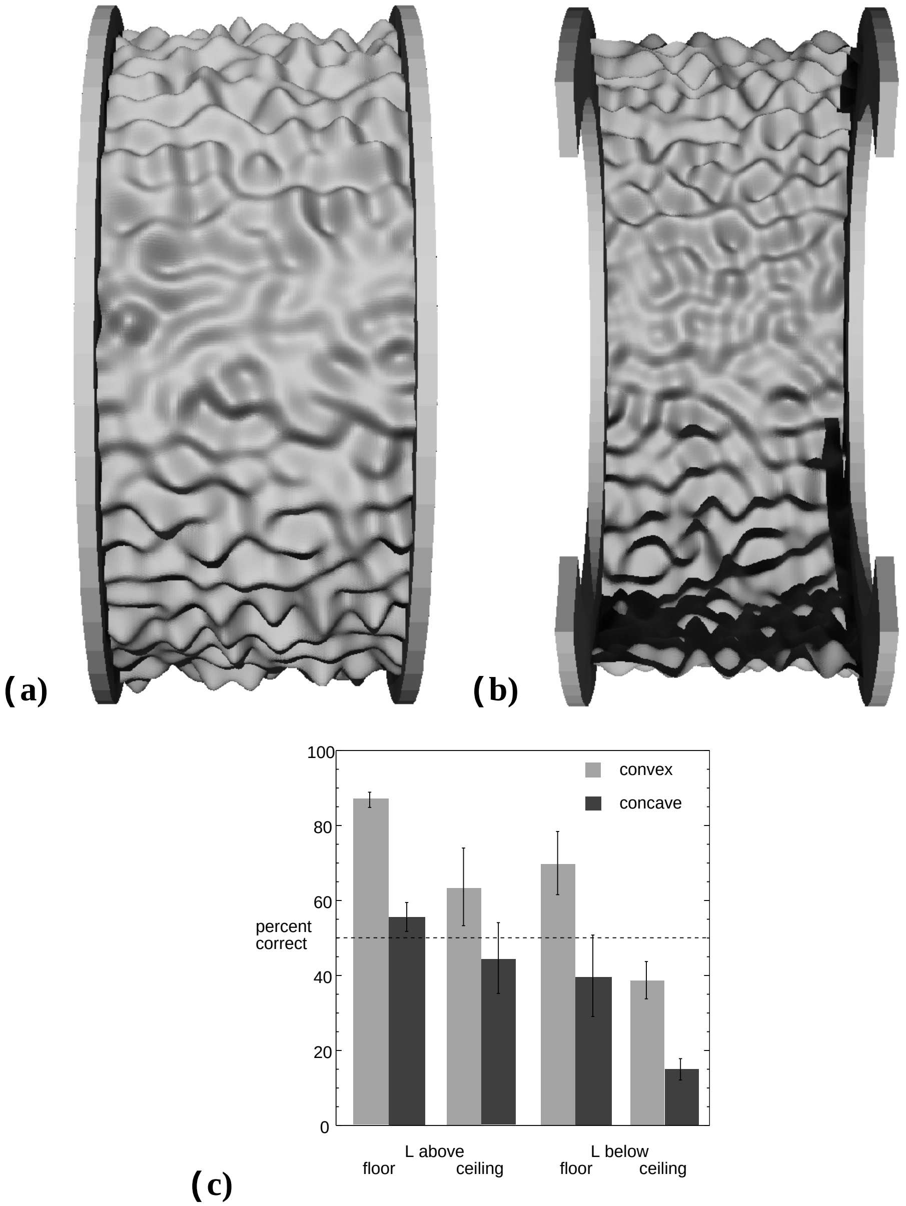

Fig. 1. Two surfaces are shown (a.) a globally convex surface illuminated from above (b.) a globally concave surface illuminated from below. Nine observers judged whether isolated points on many such surfaces were “on a hill” or “in a valley.” (c.) Percent correct scores were higher when the light source is from above, when the surface is convex, and when the point lies on a floor–like region of the surface. Error bars show the standard error of the mean of observer’s scores for each of the eight combinations of conditions.

If we also consider a transformation of the light source vector,

$$\mathbf{L} \;\rightarrow\; (\mathbf{G}^{-1})^T \mathbf{L}$$

and a transformation of the albedo

$$\mathbf{a} \;\rightarrow\; \| \, \mathbf{G} \, \mathbf{N} \, \| \, \mathbf{a}$$

then it is easy to show that these transformed variables yield exactly the same image **I** as the original object. Remarkably, this affine transformation of the object and lighting leaves the shadowed regions of the surface unchanged (see [BKY97] for a proof). This implies that the affine ambiguity holds for an arbitrary sum of collimated light sources, not just for a single collimated source. To summarize, we have that for any surface seen under orthographic projection, there is a family of affine related surfaces that produce exactly the same images.

This affine ambiguity appears to be related to a recent psychophysical finding that used depth gradient probes to recover global depth maps by integration. It was found that, although the depth maps of different observers were dissimilar in an absolute Euclidean sense, they were very similar in an affine sense [KvDK92]. One observer's depth map could be fit to another observer's depth map quite well by an affine transformation such as Eq. (2).

How do observers choose among the family of affine-related surfaces when perceiving the surface shape? Several possibilities come to mind. If the object is a familiar shape such as a human figure then observers might perceive a shape that is consistent with the many other human figures that have been seen before. Such prior information about shape may be learned from observer movement and stereo vision. For example, it has been demonstrated using a computational model that knowledge of 200 face shapes can be used to reconstruct the shape of a new face, from only a single image [BV99]. Other prior assumptions about shape might be used as well. Bilateral symmetry may be preferred, especially if the object is an animal. Observers might also prefer surfaces in which the albedo variation is minimal. A surface whose albedo is constant might be preferred over one in which the albedo is varying.

One strategy for which there is considerable evidence is that observers have a bias to see darker points as further away. Such a bias for dark-means-deep has been observed in several studies that examined local shape perception [CK97, LBss]. The bias seems to extend from local shape to global shape perception as well. One study that reconstructed a depth map from local depth gradients found that the overall slant of the surface varied by ±4 degrees as the light source direction was moved. When the source was above and to the left, the upper–left part of the surface appeared closer to the observer than when the light source was from the lower–right, in which case the lower right part of the surface appeared closer to the observer [KvDCL96]. The dark–means–deep bias is a natural one to make, in the sense that indentations of a surface tend to be dark because they tend to lie in shadow [LZ94] whereas protrusions tend to be bright because they are fully illuminated. Further studies are needed of course to explore other strategies observers use for resolving the affine ambiguity.

4 Discussion

When talking about *the shape* that an observer perceives when looking at an object, one needs to keep several issues in mind. The retina does not sample all visual directions uniformly but rather the sampling density is greatest near the line of sight. Thus for a single glance at an object, observers process some image regions much more than others. Observers can compensate for this limitation to some extent by making eye movements to explore the various parts of an image. However, one should not get the idea that there is a single high resolution z-buffer in the brain in which observers piece together the local shapes computed from each glance. Rather, as observers explore an image with eye movements, much of the information about surface shape that is computed in one glance is lost in the next. The depth maps that are computed by the experimenter from a set of measurements of perceived depth gradients (recall Sec. 2.4) should not be taken literally as a readout of the brain's z-buffer. Rather, they should be regarded as a way of studying how perceived shape varies *e.g.* as a function of the observer or as a function of an independent scene variable such as light source direction.

These issues raise a host of questions. What strategies do observers use to actively explore an image with eye movements? How are perceived shapes (or surface materials) retained and integrated from one eye movement to the next? Such questions can be addressed, in principle, by tracking eye movements and by changing the images in a systematic way in real time as an observer actively explores an image. Such studies are at the forefront of computer graphics psychophysics and are just now getting underway. In the coming years we will surely see several exciting new approaches to measuring perceived shape. These approaches will take us closer to our goal of understanding what observers perceive when they look at graphically rendered images. They should also provide key insights into how we can render images so that observers see what we want them to see, rather than what their brains want them to see.

References

BBC84. K. Berbaum, T. Bever, and C.S. Chung. Extending the perception of shape from known to unknown shading. *Perception*, 13:479–488, 1984.

BKY97. P. Belhumeur, D. Kriegman, and A. Yuille. The bas-relief ambiguity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1060–1066, San Juan, Puerto Rico, June 1997.

BM88. H. Bülthoff and H. Mallot. Interaction of depth modules: Stereo and shading. *Journal of the Optical Society of America*, 5:1749–1758, 1988.

BM90. H. Bülthoff and H. Mallot. *AI and the eye*, chapter Integration of stereo, shading, and texture, pages 119–146. Wiley, 1990.

Bre26. D. Brewster. On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena. *Edinburgh Journal of Science*, 4:99–108, 1826.

BV99. V. Blanz and T. Vetter. Morphable model for the synthesis of 3d faces. In *SIGGRAPH Conference Proceedings*, pages 187–194, 1999.

CK97. C. G. Christou and J. J. Koenderink. Light source dependence in shape from shading. *Vision Research*, 37(11):1441–1449, 1997.

EKK93. R. Erens, A. Kappers, and J. J. Koenderink. Perception of local shape from shading. *Perception and Psychophysics*, 54(2):145–156, 1993.

Gre70. R. Gregory. *The Intelligent Eye*. MacGraw-Hill, New York, 1970.

HB93. H. Hill and V. Bruce. Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion. *Perception*, 22:887–897, 1993.

KKT^{+}96. J. J. Koenderink, A. M. L. Kappers, J. T. Todd, J. F. Norman, and F. Philips. Surface range and attitude probing in stereoscopically presented dynamic scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 22:869–878, 1996.

KMK97. D. Kersten, P. Mamassian, and D. C. Knill. Moving cast shadows induce apparent motion in depth. *Perception*, 26(2):171–192, 1997.

KvDCL96. J. J. Koenderink, A. J. van Doorn, C. Christou, and J. S. Lappin. Perturbation study of shading in pictures. *Perception*, 25(9):1009–1026, 1996.

KvDK92. J. Koenderink, A. van Doorn, and A. Kappers. Surface perception in pictures. *Perception and Psychophysics*, 52(5):487–496, 1992.

KvDK96. J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers. Pictorial surface attitude and local depth comparisons. *Perception and Psychophysics*, 1996.

LBss. M. S. Langer and H. H. Bülthoff. Depth discrimination from shading under diffuse lighting. *Perception*, 2000 (in press).

Luc16. M. Luckiesh. *Light and shade and their applications*. Van Nostrand, 1st edition edition, 1916.

LZ94. M. Langer and S. Zucker. Shape-from-shading on a cloudy day. *Journal of the Optical Society of America A*, 11(2):467–478, 1994.

MKK96. P. Mamassian, D. Kersten, and D. Knill. Categorical local-shape perception. *Perception*, 25:95–107, 1996.

MLM97. P. Mamassian, M. S. Landy, and L. T. Maloney. Global shape and surface orientation. *Investigative Ophthalmology and Visual Science*, 1997.

MT86. E. Mingolla and J. Todd. Perception of solid shape from shading. *Biological Cybernetics*, 53:137–151, 1986.

NT96. J. F. Norman and J. T. Todd. The discriminability of local surface structure. *Perception*, 25:381–398, 1996.

RB00. J. C. Rodger and R. A. Browse. Choosing rendering parameters for effective communication of 3d shape. *IEEE Computer Graphics and Applications*, 2000.

Rit86. D. Rittenhouse. Explanation of an optical deception. *Transactions of the American Philosophical Society*, 2:37–43, 1786.

RT90. F. R. Reichel and J. T. Todd. Perceived depth inversion of smoothly curved surfaces due to image orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3):653–664, 1990.

RTY95. F. D. Reichel, J. T. Todd, and E. Yilmaz. Visual discrimination of local surface depth and orientation. *Perception and Psychophysics*, 57(8):1233–1240, 1995.

SB87. K. Stevens and A. Brookes. Probing depth in monocular images. *Biological Cybernetics*, 56:355–366, 1987.

Ste83. K. Stevens. Slant-tilt: the visual encoding of surface orientation. *Biological Cybernetics*, 46:183–195, 1983.

TA87. J. Todd and R. Akerstrom. Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2):242–255, 1987.

TM83. J. Todd and E. Mingolla. Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, 9(4):583–595, 1983.

TN95. J. T. Todd and J. F. Norman. The visual discrimination of relative surface orientation. *Perception*, 24:855–866, 1995.

TR89. J. Todd and F. Reichel. Ordinal structure in the visual perception and cognition of smoothly curved surface. *Psychological Review*, 96(4):643–657, 1989.

War94. G. J. Ward. The radiance lighting simulation and rendering system. *Computer Graphics*, pages 459–472, July 1994.