The perception of surface curvature from optical motion

VICTOR J. PEROTTI and JAMES T. TODD
Ohio State University, Columbus, Ohio

JOE S. LAPPIN
Vanderbilt University, Nashville, Tennessee

and

FLIP PHILLIPS
Ohio State University, Columbus, Ohio

Observers viewed the optical flow field of a rotating quadric surface patch and were required to match its perceived structure by adjusting the shape of a stereoscopically presented surface. In Experiment 1, the flow fields included rigid object rotations and constant flow fields with patterns of image acceleration that had no possible rigid interpretation. In performing their matches, observers had independent control of two parameters that determined the surface shape. One of these, called the shape characteristic, is defined as the ratio of the two principle curvatures and is independent of object size. The other, called curvedness, is defined as the sum of the squared principle curvatures and depends on the size of the object. Adjustments of shape characteristic were almost perfectly accurate for both motion conditions. Adjustments of curvedness, on the other hand, were systematically overestimated and were not highly correlated with the simulated curvedness of the depicted surface patch. In Experiment 2, the same flow fields were masked with a global pattern of curl, divergence, or shear, which disrupted the first-order spatial derivatives of the image velocity field, while leaving the second-order spatial derivatives invariant. The addition of these masks had only negligible effects on observers' performance. These findings suggest that observers' judgments of three-dimensional surface shape from motion are primarily determined by the second-order spatial derivatives of the instantaneous field of image displacements.

Human observers make use of many different sources of information to perceive the three-dimensional (3-D) structure of their environment. One of the most important of these sources is the changing pattern of retinal stimulation created by the movements of the observer and environmental objects. Wallach and O'Connell (1953) showed that motion information could by itself create an impression of 3-D structure. In their demonstration, an apparently two-dimensional (2-D) form created by the projected shadows of a wireframe figure suddenly appeared to have a compelling 3-D structure when the wireframe was rotated in depth.

Early theoretical analyses of this phenomenon defined the minimal information that would permit an observer to determine the 3-D structure of a configuration of points. To simplify the analysis, these models typically assume that the input is a set of image points moving across discrete frames. For general configurations under orthographic projection, a minimum of three distinct views of

four noncoplanar points are needed to compute the 3-D structure of an object up to a reflection in depth (e.g., see Bennett & Hoffman, 1986; Hoffman & Bennett, 1985, 1986; Hoffman & Flinchbaugh, 1982; Ullman, 1979). For objects observed under polar projection, only two views of five points are needed to specify 3-D structure (Longuet-Higgins & Prazdny, 1980), although, in this case, without additional information to specify an object's distance from the point of observation, the analysis is ambiguous as to the object's exact size.

A large number of psychophysical tests have been performed to determine how well these computational theories apply to human observers. Many of the results have indicated that observers have no trouble perceiving 3-D shapes when presented with only two frames under orthographic projection (see Braunstein, Hoffman, & Pollick, 1990; Braunstein, Hoffman, Shapiro, Andersen, & Bennett, 1987; Lappin, Doner, & Kottas, 1980; Norman & Todd, 1993; Todd, Akerstrom, Reichel, & Hayes, 1988). Moreover, for many different perceptual tasks, there is no significant difference between observers' performance when shown two views and their performance when shown more than two views (see Liter, Braunstein, & Hoffman, 1993; Todd & Bressan, 1990; Todd & Norman, 1991). These results demonstrate that observers are

able to perceive compelling 3-D structure with less information than the minimum amount specified in the computational theories.

In light of these findings, more recent theoretical analyses have examined more closely the available information that can be obtained from only two distinct views of an apparent motion sequence or from the instantaneous field of image velocities for objects observed under orthographic projection (see Bennett, Hoffman, Nicola, & Prakash, 1989; Koenderink & van Doorn, 1991; Todd & Bressan, 1990; Ullman, 1977). These analyses have shown that first-order temporal information can specify the structure of an object up to an infinite family that varies along a single parameter. Despite this ambiguity, however, there are some aspects of 3-D structure that are common to all members of this one-parameter family and can therefore be determined uniquely.

To better understand which aspects of 3-D structure are unambiguously specified from first-order temporal information and which ones are not, it is useful to consider the special case of rotation about a vertical axis in the image plane. Ullman (1977) proved that any rigid 3-D rotation under orthographic projection can be decomposed into two components: a rotation about an axis in the image plane, followed by a rotation about the line of sight. He further proved that the component of rotation about the line of sight can be mathematically eliminated from the global pattern of image motion, such that any 3-D rotation can effectively be reduced to our assumed case. In the analysis presented below, we will restrict our discussion to the instantaneous pattern of velocities for objects observed in continuous motion, although a related analysis for two discrete frames of apparent motion is provided in the Appendix.

Consider a Monge surface, for which the depth (z) of any point can be expressed as a simple function of its horizontal position, x, and its vertical position, y:

$$z = l(x, y). \tag{1}$$

The image velocity field, V, formed by the orthographic projection of this surface as it rotates about a vertical axis, can be expressed as a function of the horizontal and vertical positions:

$$V = v(x, y). \tag{2}$$

These two functions are related by a constant scale factor, ω, which is the rate of the surface's angular rotation:

$$V(x, y) = l(x, y)\omega. \tag{3}$$

The scale factor, ω, is a specific quantity that must be determined in order to get an unambiguous measurement of surface structure from the velocity field. The constant relationship between the image-velocity field and the surface structure holds also for the first and second partial derivatives across space. To simplify the notation, a subscripted letter is used to identify a partial (spatial) derivative in a given direction. Two subscripted letters indicate a second partial derivative in the directions specified. For example, the first partial derivatives of the velocity field are given by

$$V_x = \omega l_x \text{ and } V_y = \omega l_y. \tag{4}$$

The second partial derivatives are

$$V_{xx} = \omega l_{xx}, V_{xy} = \omega l_{xy}, \text{ and } V_{yy} = \omega l_{yy}. \tag{5}$$

Although it is desirable to obtain an unambiguous measure of an object's structure from its projected motion, the existence of an unknown scale factor ω means that the structure is only specified up to a one-parameter family of possible interpretations. However, if we consider instead the ratios of these measures, the scale factor ω would drop out. Ratios of the partial derivatives are unambiguous measures and thus represent *invariant* properties for rotating objects under orthographic projection.

The distinction between ratio and nonratio quantities is a powerful one. This idea allows us to predict which properties in the environment can be unambiguously determined from the velocity field and which cannot. For example, consider the well-known descriptions of surface attitude—slant (σ) and tilt (τ):

$$\tan \sigma = \sqrt{l_x^2 + l_y^2} = \frac{\sqrt{V_x^2 + V_y^2}}{\omega}, \tag{6}$$

$$\tan \tau = \frac{l_x}{l_y} = \frac{V_x}{V_y}. \tag{7}$$

The difference between these two quantities is evident: slant has an inherent ambiguity, as it is reflected in the velocity field, but tilt does not. This means that any estimate of slant is only as good as the observer's estimate for the angular rotation velocity, whereas tilt is specified uniquely by the first spatial derivatives of the velocity field (see Cornilleau-Pérès & Droulez, 1989; Koenderink & van Doorn, 1975, 1977; Te Pas, Kappers, & Koenderink, 1996).

In the same manner in which the attitude of surfaces in the environment are related to the first spatial derivatives, the surface shape is related to the second spatial derivatives of the image-velocity field. To better explain this relationship, we will use the concepts of differential geometry to describe any 3-D shape. At a given surface location, the local shape is defined by the *normal curvatures*, or the curvature of the surface in every direction in the tangent plane. A concise definition of the local surface shape is the minimum and maximum normal curvature called the *principal curvatures*: $\kappa_{\min}$ and $\kappa_{\max}$ (see Figure 1). The directions in the tangent plane of the principal curvatures are called the *principal directions* and are always perpendicular to one another.

We shall now consider several important shape properties and whether they are unambiguously measurable from the image-velocity field. For example, the angle (α) of rotation about the line of sight that is needed to align
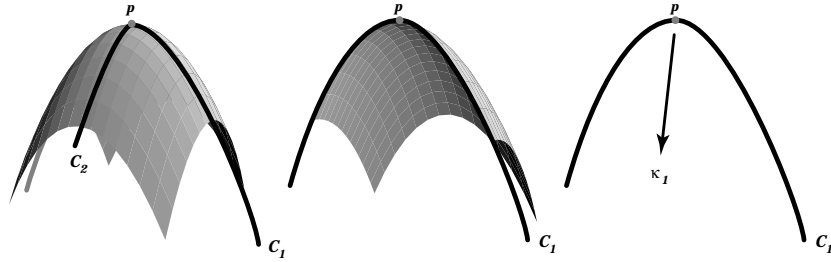
Figure 1. The principal directions, C_1 and C_2, and principal curvature, κ_1 are shown at a surface point p. From "Perception of Local Three-Dimensional Shape," by F. Phillips and J. T. Todd, 1996, *Journal of Experimental Psychology: Human Perception & Performance*, *22*, p. 931. Copyright 1996 by the American Psychological Association. Adapted with permission.

the principal curvatures with the horizontal and vertical axes can be unambiguously determined:

$$\cot(2\alpha) = \frac{1}{2}\left(\frac{l_{xx} - l_{yy}}{l_{xy}}\right) = \frac{V_{xx} - V_{yy}}{2V_{xy}}. \tag{8}$$

The magnitudes of these realigned principal curvatures are defined by the following equations:

$$\kappa_x = \frac{1}{\sqrt{1 + l_x^2 + l_y^2}}\left(\frac{l_{xx}}{1 + l_x^2}\right) = \frac{\omega^2 V_{xx}}{\left(\sqrt{\omega^2 + V_x^2 + V_y^2}\right)\left(\omega^2 + V_x^2\right)}, \tag{9}$$

$$\kappa_y = \frac{1}{\sqrt{1 + l_x^2 + l_y^2}}\left(\frac{l_{yy}}{1 + l_y^2}\right) = \frac{\omega^2 V_{yy}}{\left(\sqrt{\omega^2 + V_x^2 + V_y^2}\right)\left(\omega^2 + V_y^2\right)}. \tag{10}$$

Note that these equations both contain the scale factor ω, so that the principal curvatures are not uniquely specified by the instantaneous velocity field. There are, however, some invariant relationships between the principal curvatures that can be determined under appropriate conditions.

Consider, for example, the ratio of the two principal curvatures, which is sometimes referred to as the *shape characteristic* (Mamassian, 1993; see Figure 2):

$$\frac{\kappa_x}{\kappa_y} = \frac{l_{xx}}{l_{yy}}\left(\frac{1 + l_y^2}{1 + l_x^2}\right) = \frac{V_{xx}}{V_{yy}}\left(\frac{\omega^2 + V_y^2}{\omega^2 + V_x^2}\right). \tag{11}$$

For cases where the first spatial derivatives of the velocity field are zero (i.e., where the tangent plane to the surface is frontoparallel) or where they are equal to one another, the shape characteristic is unambiguously measurable. This is not strictly the case for points outside these regions, though the resulting error is negligible over a relatively broad range of surface slants (see Dijkstra, Snoeren, & Gielen, 1994).

Another commonly used relationship among the two principal curvatures is called *curvedness* and is defined as the square root of their sum of squares (Koenderink,

1990). Because this is a sum rather than a ratio, the scale factor ω is not removed, so that the curvedness portion of 3-D shape cannot be uniquely specified by the instantaneous velocity field. It could, however, be perceptually specified if the visual system were sensitive to higher order temporal derivatives, such as image accelerations.

It is especially interesting to note, in this context, that a similar distinction between the shape characteristic and curvedness is also applicable to the perceptual analysis of structure from motion for objects that are observed under polar projection. It is important to keep in mind that an instantaneous velocity field under polar projection allows a one-parameter family of possible interpretations that are related to one another by a homogeneous scaling transformation. Thus, it is similar to orthographic projection, in that the shape characteristic of a surface patch is uniquely specified but not its curvedness.

Although the analysis presented above requires some preliminary transformations in order to achieve a specific coordinate system and a specific rotation axis, a more general formulation without these requirements can be found in Koenderink and van Doorn (1992) or Dijkstra et al., (1994). In the latter paper, several simulations are also presented to show that estimations of the shape characteristic are only minimally perturbed if the tangent plane of a surface patch is slanted in depth or if it is viewed under polar perspective (see also Lappin, Ahlström, Craft, & Tschantz, 1995).

The research described in the present article was designed to compare the predictions of this analysis with the perceptual performance of human observers. In Experiment 1, we examined the importance of higher order temporal derivatives for the visual perception of surface curvature by comparing performance for two different types of displays—rigid object rotations and constant flow fields for which the pattern of image accelerations were mathematically inconsistent with any rigid object rotation. In Experiment 2, we investigated the importance of higher order spatial derivatives in the image velocity field by comparing performance for moving surfaces presented both with and without a first-order masking field.
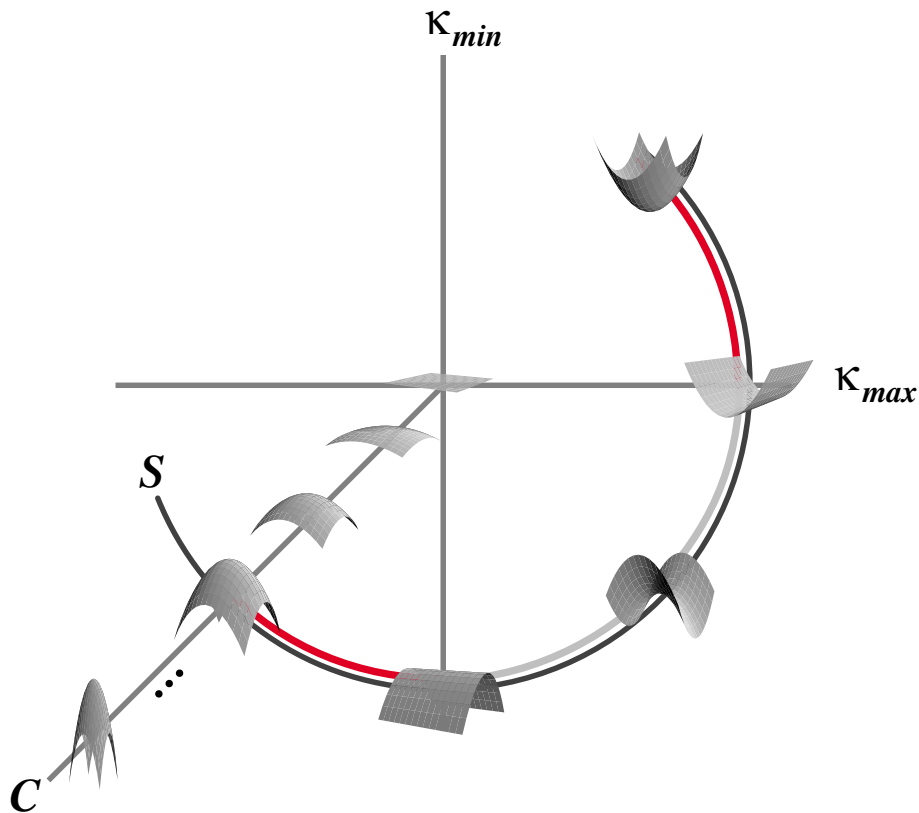
Figure 2. If the two principal curvatures ($\kappa_{\min}$ and $\kappa_{\max}$) of a surface patch are plotted in Cartesian coordinates, the curvedness (C) and shape characteristic (S) can be obtained using a transformation to polar coordinates. From "Perception of Local Three-Dimensional Shape," by F. Phillips and J. T. Todd, 1996, *Journal of Experimental Psychology: Human Perception & Performance*, *22*, p. 933. Copyright 1996 by the American Psychological Association. Adapted with permission.

EXPERIMENT 1

Method

The basic procedure employed in this experiment was to have observers first view a motion stimulus depicting a randomly chosen 3-D shape and then adjust a stereoscopically presented stimulus until it appeared to have the same shape. The motion stimuli were either regular rigid object rotations or displays with the second and higher temporal derivatives that were inconsistent with any rigid object rotation. A comparison of observers' adjusted shapes in these two conditions allowed an estimation of the extent to which observers utilized the image accelerations or higher order relations among three or more views.

Stimuli. The stimuli depicted paraboloid quadric surfaces, each having the form

$$z = \frac{(\kappa_1 x^2 + \kappa_2 y^2)}{2}, \qquad (12)$$

where κ_1 and κ_2 represent the principal curvatures at the origin. Examples of this kind of surface can be seen in Figure 2. The depicted surfaces were rotated in the image plane so that the principal curvatures were oriented at a 45º angle to the horizontal and vertical. This manipulation was performed in order to nullify the effects of the known anisotropies in the horizontal and vertical directions (e.g., see Todd, Tittle, & Norman, 1995).

In half of the displays, the pattern of image accelerations was inconsistent with any rigid object rotation. The technique for creating

these stimuli was the same as that used in Perotti, Todd, and Norman (1996). To understand how this was done, consider how one would define the image-displacement field for an orthographically projected, rigidly moving 3-D object. The projected displacement field, D, depends on three factors: the horizontal position, x; the vertical position, y; and the time, t, at which the displacement is measured:

$$D = f(x, y, t). \qquad (13)$$

For one rare situation in nature, the projected displacements at each image location do not change over time; thus, they create a constant displacement flow field. This situation occurs when a surface of revolution rotates about its axis of symmetry (see also Lappin & Ahlström, 1994). In general, however, a constant displacement flow field will not be consistent with any rigid object rotation.

The motion of an image point in a constant (displacement) flow field, D, depends only on its horizontal and vertical position, x and y:

$$D = f(x, y). \qquad (14)$$

Each 2-D position in the image is associated with a particular displacement that remains fixed over time. To create a compelling stimulus, the movement of a set of randomly placed points is determined by these displacements. When a point arrives at a particular 2-D location, it is assigned the displacement associated with that position, and it moves accordingly on the next frame transition. The important property of constant flow fields is that the first temporal derivatives of the points' positions are consistent with a rigid object motion, but their image accelerations are not. Thus, if observers rely
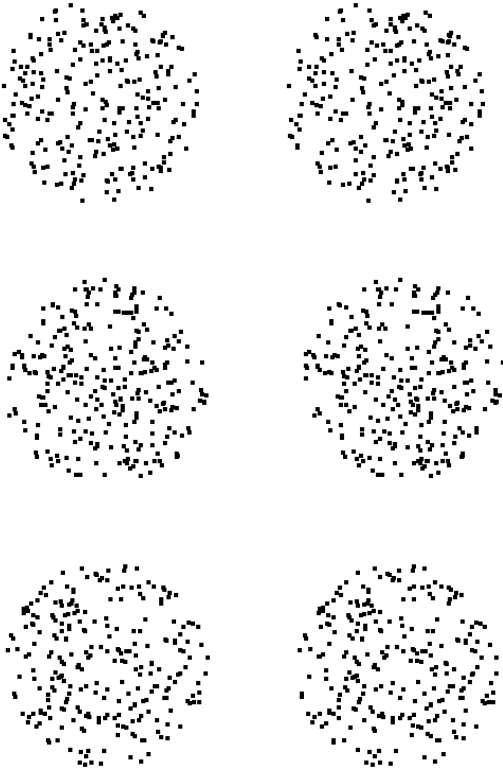
Figure 3. Stereograms illustrating the type of surfaces used in Experiment 1. From the top: an elliptic bump, a parabolic cylinder, and a hyperbolic saddle shape are depicted.

on the higher order temporal derivatives, the 3-D shapes that they reconstruct from the constant flow field stimuli should be much less accurate than those that they reconstruct from the rigid object motion stimuli. However, if observers can only utilize the first temporal derivatives, then the constant flow field and the rigid object rotation stimuli should yield equally accurate 3-D shapes.

The stereograms in Figure 3 give a schematic depiction of the quadric surfaces that were used in the experiment. The surfaces were generated from the equations described below, where x and y are the horizontal and vertical image positions at which a displacement is computed, and κ_1 and κ_2 are constants chosen at random from the range $[-3.0, 3.0]$ cm^{-1}. For the rigid rotations and the stereoscopic matching stimuli, the z depth values of the points (in pixels) were generated, using the following equations:

$$z = 30\left[(\kappa_1 + \kappa_2)\left(\frac{x}{100}\right)^2 + (\kappa_1 + \kappa_2)\left(\frac{y}{100}\right)^2 + 2(\kappa_2 - \kappa_1)\frac{xy}{10,000}\right],$$

(15)

$$z = 30\left[\begin{array}{c}(\kappa_{1adj} + \kappa_{2adj})\left(\frac{x}{100}\right)^2 + (\kappa_{1adj} + \kappa_{2adj})\left(\frac{y}{100}\right)^2 \\ + 2(\kappa_{2adj} - \kappa_{1adj})\frac{xy}{10,000}\end{array}\right].$$

(16)

For the stereoscopic matching stimuli, the subscript *adj* indicates that the values of κ_1 and κ_2 were interactively adjusted by the observer. A similar equation was also used to generate the image displacements (expressed in pixels per frame) in the constant flow field conditions:

$$D(x,y) = \frac{(\kappa_1 + \kappa_2)\left(\frac{x}{100}\right)^2 + (\kappa_1 + \kappa_2)\left(\frac{y}{100}\right)^2 + 2(\kappa_2 - \kappa_1)\frac{xy}{10,000}}{6}.$$

(17)

The rigid stimuli rotated about a vertical axis between $-22.5°$ and $+22.5°$ from frontoparallel at a rate of 0.75° per frame. Thus, it took 60 total frames to rotate the rigid surface through its full extent of motion. The surface continued to rotate between the two extrema until the observer indicated that a judgment had been made. Constant flow fields also alternated back and forth through 60 unique frames, with a flow field that was identical to the rigid rotations in their frontoparallel orientation. In both cases, the initial direction of rotation was chosen at random.

Both the motion and the stereoscopic stimuli were composed of 2,000 points, whose positions were defined in a Cartesian coordinate system with an origin at the center of the display screen, which had a horizontal and vertical resolution of 1,280 × 1,024 pixels. Thus, the horizontal position of each element ranged from -640.0 to 640.0 pixels, and the vertical positions ranged from -512.0 to 512.0 pixels. The constant flow field patterns were viewed through a circular aperture that was 380 pixels in diameter, with the video screen a uniform black outside this area. For the rigid rotation conditions, a similar aperture was not used, because it greatly limited the available range of curvedness that was possible. Instead, the rigid rotation stimuli initially projected to a 380-pixel-diameter circle, but during the rotation had a deforming boundary consistent with the simulated shape.

Procedure. Each observer monocularly viewed a motion stimulus depicting the selected 3-D shape for as long as desired and then toggled (using a mouse button) to a binocularly viewed stereoscopic adjustment stimulus. Observers manipulated the stereoscopic stimuli using horizontal and vertical movement of the mouse to adjust the two principal curvatures of the depicted quadric surface. They were instructed to switch back and forth between both displays until the apparent 3-D shapes were matched as closely as possible. Once this was achieved, a trial was terminated by pressing a separate mouse button. Rigid rotation and constant flow field stimuli were presented in separate blocks, each containing 100 trials.

Apparatus. The displays were presented on a Silicon Graphics VGXT workstation. The refresh rate for all displays was 60 Hz. All displays were viewed through Crystal Eyes LCD shuttered glasses, at a viewing distance of 57.3 cm. The motion displays were viewed with an eyepatch over one eye, whereas the stereoscopically presented displays were viewed with both eyes uncovered.

Observers. Three of the authors (V.P., J.T., and F.P.) participated as observers, all of whom had corrected-to-normal vision.

Results and Discussion

All of the observers reported that the task seemed quite natural and that they were able to adjust the stereoscopic displays to closely match the apparent 3-D shapes of the motion test surfaces. They also reported that there was little apparent difference between the rigid rotations and constant flow fields, and that both types of display appeared perfectly rigid, as has been reported previously by Perotti et al. (1996).

Two factors complicated the analysis of the data. The first was an inherent ambiguity between the direction of rotation and the sign of curvature for objects displayed under orthographic projection. For example, a motion pattern of a bump rotating in one direction is identical to the motion pattern depicting a dimple rotating in the op-
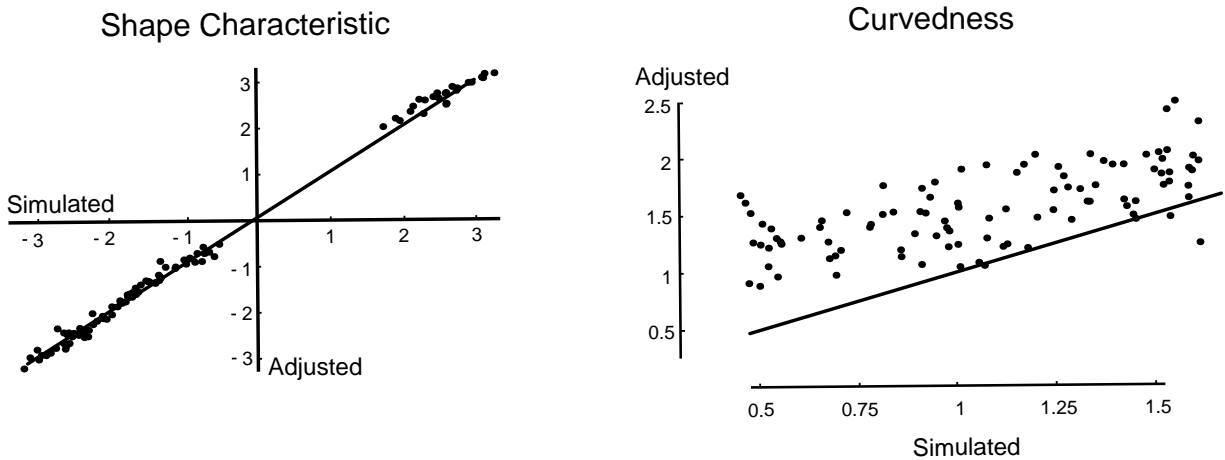
Shape Characteristic

Curvedness

Figure 4. Simulated versus adjusted shape characteristic and curvedness for observer J.T. in the constant flow field condition of Experiment 1. The diagonal lines indicate perfect reconstruction performance. Note the superior performance for judgments of shape characteristic in comparison with those of curvedness.

posite direction. In order to resolve this ambiguity, the one alternative that was closest to the adjusted shape was considered to be the test stimulus.

A second potential problem involved scaling the shape characteristic. As suggested by Koenderink (1990), we transformed this to a circular measure by computing the arctangent of the ratio of the adjusted values for κ_1 and κ_2. These angular data were then analyzed by performing a linear regression between the adjusted and simulated shapes along the circumference of a circle. A separate linear regression was performed for the curvedness component of the observers' adjustments, which was calculated on each trial as the square root of the sum of squares of κ_1 and κ_2.

To provide a typical example of the overall pattern of performance, Figure 4 shows a scatter plot of the data for observer J.T. in the constant flow field condition. The data for the other observers are summarized in Table 1, which gives the R^2 values for adjusted versus simulated shape characteristic and curvedness in all conditions. It is clear from these results that the observers were remarkably accurate in their judgments of the shape characteristic. Indeed, the adjusted values of this component in the stereo-

scopic displays were almost perfectly correlated with their simulated values in the motion displays. Across all observers and conditions, the average R^2 for the shape characteristic correlations was .994, the average slope of the best fitting regression line was 1.00, and the average intercept was 0.03. It is important to keep in mind that the shape characteristic is unambiguously specified in both the constant flow fields and rigid rotations, so it is not theoretically surprising that the observers produced equivalent performance in both conditions.

It is interesting to note in Figure 4 that there is a noticeable gap in the central region of the shape characteristic scatter plot. This phenomenon arises because the sign of relief is mathematically ambiguous for moving surfaces under orthographic projection. For hyperbolic (i.e., saddle-shaped) surface patches in the present experiment, both possible interpretations of surface relief were perceived with equal frequency, but for the elliptic patches, the observers were strongly biased to see the surfaces as convex, relative to the point of observation.

Curvedness adjustments were much less accurate than those for shape characteristic, which is consistent with the hypothesis that the perception of 3-D structure from motion is primarily determined by the first-order pattern of projected image velocities. In contrast to the shape characteristic, curvedness requires an analysis of higher order relations among three or more views and cannot be uniquely specified by a pattern of image velocities or displacements. To the extent that observers are able to make use of this information, we would expect to see differences between the rigid rotations and constant flow fields, because their patterns of image acceleration are quite different. To compute a correlation for the constant flow fields, the simulated curvedness was calculated using the values of κ_1 and κ_2 from Equation 17 that were used to generate the field of image displacements. Because the patterns of image acceleration in that condition are in-

Table 1
The R^2 Values for the Linear Correlations Between
Simulated and Adjusted Shape Characteristic and
Curvedness for All Observers and Conditions in Experiment 1

Observer	Motion Stimulus	R^2 for Shape Characteristic	R^2 for Curvedness
F.P.	constant flow field	.9927	.1886
	rigid rotation	.9901	.3741
J.T.T.	constant flow field	.9964	.4398
	rigid rotation	.9972	.7246
V.J.P.	constant flow field	.9972	.2083
	rigid rotation	.9975	.5521
Average	constant flow field	.9954	.2789
	rigid rotation	.9949	.5503

compatible with any rigid body motion, it is perhaps not surprising that the correlations between simulated and adjusted curvedness produced a mean R^2 of only .27.

An interesting question to consider is how the observers perceived any curvedness at all from the constant flow field stimuli, since the information to specify curvedness was not reliable in those displays. One possibility is that the observers used some sort of heuristic analysis to estimate a unique structure. For example, Liter et al. (1993) and Todd and Norman (1991) have suggested that perceived depth may be scaled by the overall range of projected displacements in a moving display. If that were the case, perceived depth should increase proportionally with the angular velocity of an object's rotation. However, Todd and Norman and Perotti, Todd, Tittle, and Norman (1994) have found little difference in observers' performance when angular velocity is systematically varied.

Although there was potential information about curvedness in the higher order temporal relations of the rigid rotation displays, there was not much improvement in the observers' performance. In order to compare the reliability of adjusted curvedness for the rigid rotation and constant flow field conditions, an F test was performed for each observer on the ratio of residual variance from the linear correlations. For one observer, V.P., the R^2 values that were obtained in the two conditions were significantly different [$F(99,99) = 2.05568, p < .05$], but for the other two observers they were not. We also performed a Fischer's protected least significant difference test (PLSD) to compare the average signed error in the different conditions (see Figure 5). For two of the observers, V.P. and F.P., the magnitude of adjusted curvedness was signifi-

cantly greater in the constant flow field displays ($p < .001$), though this effect did not occur for observer J.T. These small differences in performance between the rigid rotation and constant flow field conditions could be due to observers' limited sensitivity to the pattern of image accelerations (see also Norman & Todd, 1993; Perotti et al., 1996), or alternatively, they could be due to the fact that the constant flow field displays had fixed circular boundaries, whereas those for the rigid rotations deformed over time.

Although it may seem theoretically reasonable to attribute errors of adjusted curvedness to distortions that occur in the perceptual analysis of structure from motion, that is not necessarily the case. It is also possible that the errors were due to observers' perceptions of the stereoscopic matching stimuli. With appropriate information about interocular distance and the state of convergence, it would be theoretically possible to accurately compute 3-D structure from binocular disparity, but the available evidence does not suggest that human observers are capable of that (e.g., see Norman, Todd, Perotti, & Tittle, 1996; Tittle, Todd, Perotti, & Norman, 1995). Several studies have shown that either motion or stereo provides sufficient information to accurately discriminate the shape characteristics of quadric surfaces (de Vries, Kappers, & Koenderink, 1994; Phillips & Todd, 1996; van Damme, Oosterhoff, & van de Grind, 1994), but this does not appear to be true for the scale-dependent property of curvedness, as indicated by the low R^2 values shown in Table 1. This also suggests, moreover, that the errors occurring in these two modalities are not linearly related to one another.

EXPERIMENT 2

Experiment 2 was designed to evaluate how the spatial relationships among image velocities are used to determine 3-D structure. Recall that our analysis specifies that the second spatial derivatives of the velocity field allow an unambiguous measurement of the surface shape characteristic. To test this claim, the observers in Experiment 2 adjusted 3-D shapes after viewing displays with the first spatial derivatives of the motion stimulus masked. Figure 6 schematically illustrates the three masking fields that were used to perturb the motion patterns. These masking fields alter the first spatial derivatives of the velocity field, but they leave the higher spatial derivatives intact. Thus, if the perception of 3-D shape is based primarily on the second-order spatial derivatives, adjusted shape should not be affected by the masks at all.

Method

The apparatus and procedure were identical to those used in Experiment 1. Each motion stimulus depicted a constant flow field or a rigid object rotation identical to those used in Experiment 1, which was additively combined with a first-order motion mask. A masked display was created by adding curl, divergence, or shear to the projected motions of each dot, as described by Equations 15 and 17.
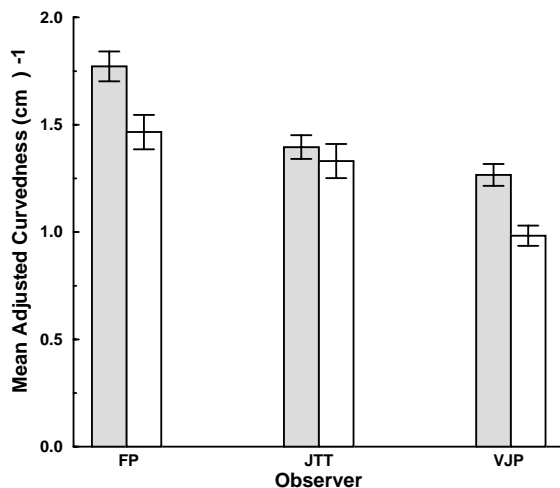

Figure 5. The average error of observers' curvedness adjustments for the different conditions of Experiment 1. The shaded bars indicate the mean difference between adjusted and simulated curvedness for the constant flow field condition; the light bars represent the same measurement for the rigid rotation condition. Error bars indicate the 95% confidence intervals for each condition.
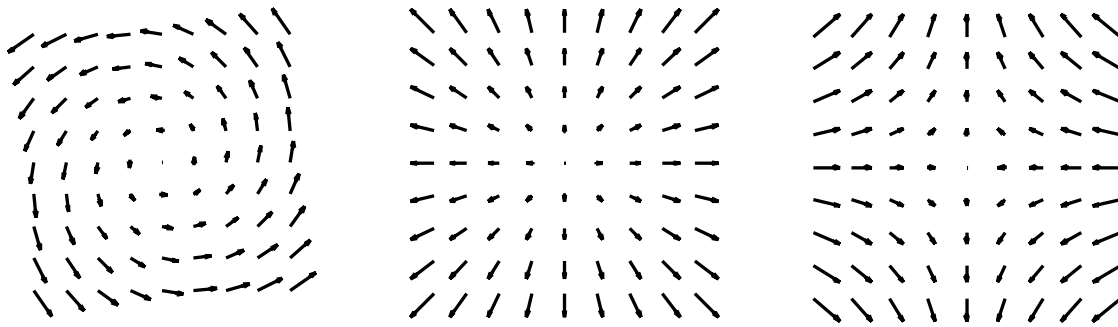
Figure 6. The three motion patterns used to mask the first spatial derivatives of the velocity field in Experiment 2.

The mask displacement for a given point was a function of its horizontal position, x, and its vertical position, y. For curl,

$$x' = x \cos \theta + y \sin \theta, y' = y \cos \theta - x \sin \theta; \qquad (18)$$

for divergence,

$$x' = x\alpha, y' = y\alpha; \qquad (19)$$

and for shear,

$$x' = x\alpha, y' = \frac{y}{\alpha}. \qquad (20)$$

The orientation of the shear component was determined at random for each trial. The image rotation angle, θ, was set to 0.18° per frame, and the scale factor, α, was set to 1.00325. These values were chosen using a set of Monte Carlo simulations. For each masking condition, the total motion energy was determined by summing the squared magnitude of all points' image displacements. Different values of θ and α were tried until this total motion energy was roughly equal for all three masking conditions and the total motion energy in each mask was equivalent to the unmasked motion energy. The rigid rotations and constant flow fields were presented in separate sessions, each containing three blocks of 100 trials for the different mask conditions. Two of the authors, V.P. and J.T., again participated as observers.

Results and Discussion

In describing their impressions of the phenomena, both observers noted that each masking pattern had a distinct effect on the apparent 3-D motion of the depicted surfaces. For the curl displays, the objects appeared to be rotating in depth about a moving axis that was precess-

ing in the image plane. For the divergence displays, the objects appeared to be rotating back and forth about a vertical axis, while simultaneously translating back and forth in depth. For the shear displays, the objects appeared to be rotating in depth about a fixed axis in the image plane that varied at random across trials.

Table 2 shows the R^2 values that were obtained, for each observer in each condition, from the linear correlations between adjusted and simulated shape characteristic and curvedness. As in Experiment 1, the observers' judgments of the shape characteristic were almost perfectly accurate. Averaged over conditions and observers, the mean (R^2) for the shape characteristic was .995; the average slope of the best fitting regression line was .99; and the average intercept was .04. For judgments of curvedness, in contrast, the values of R^2 were much lower.

Additional analyses were performed to compare the R^2 values and mean adjusted errors among the different conditions. The ratio of residual variances from the linear correlation revealed that the R^2 values for both observers were significantly different for the rigid rotation and constant flow field displays [J.T., $F(299,299) = 1.5869$, $p < .05$; V.P., $F(299,299) = 1.61591$, $p < .05$]. They also differed significantly in their mean adjusted errors as revealed by a Fischer's PLSD test ($p < .0001$). A comparison of the R^2 values among the various types of motion masks and the unmasked displays of Experiment 1 re-

Table 2
The R^2 Values for the Linear Correlations Between Simulated and Adjusted
Shape Characteristic and Curvedness for All Observers and Conditions in Experiment 2

| Observer	Motion Stimulus	Image Rotation		Shear		Divergence	
		Shape Characteristic	Curvedness	Shape Characteristic	Curvedness	Shape Characteristic	Curvedness
V.J.P.	Constant flow field	.9954	.2506	.9898	.3303	.9947	.1595
	Rigid rotation	.9970	.5466	.9906	.3802	.9983	.4941
J.T.T.	Constant flow field	.9970	.5276	.9955	.5063	.9972	.4098
	Rigid rotation	.9969	.6619	.9955	.5608	.9970	.6340
Average	Constant flow field	.9962	.3891	.9926	.4183	.9960	.2846
	Rigid rotation	.9970	.6043	.9930	.4705	.9976	.5640
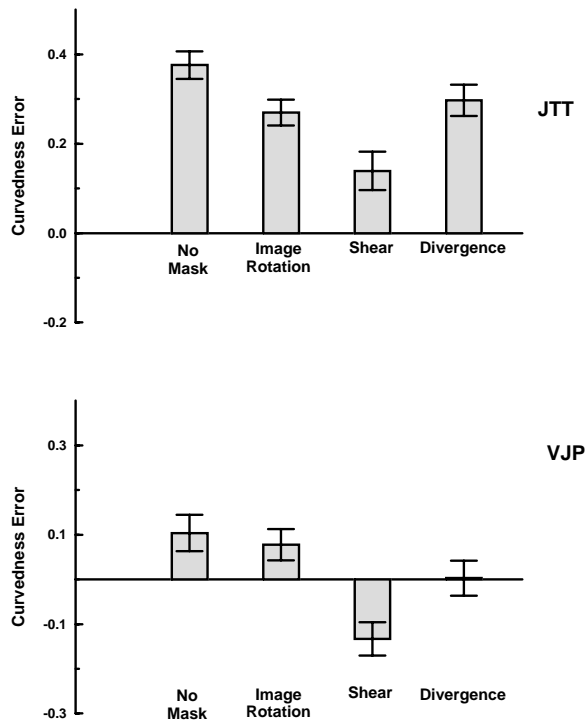
Figure 7. The average error of observers' curvedness adjustments for the different motion masks of Experiment 2 and the unmasked displays of Experiment 1. Error bars indicate the 95% confidence intervals for each condition.

vealed no systematic effects, though several of the pairwise comparisons were statistically significant. There was, however, a clear pattern of results for the mean signed errors of the observers' curvedness adjustments (see Figure 7). For both observers, the addition of a motion mask significantly lowered the magnitude of perceived curvedness (PLSD, $p < .0001$), and the effect of shear was significantly larger than the effects of rotation or divergence (PLSD, $p < .0001$).

It is important to keep in mind that these particular masking patterns were designed to disrupt the first-order spatial derivatives of the image velocity field while leaving the second-order spatial derivatives invariant. Thus, the fact that these masks had virtually no effect at all on observers' judgments of the shape characteristic (see also Lappin & Craft, 1997) provides strong empirical support for the theoretical proposals of Dijkstra et al. (1994) and Koenderink and van Doorn (1992) that second-order spatial derivatives of the instantaneous velocity field are a primary source of visual information for the scale-independent aspects of 3-D shape. It is also interesting to note, in this regard, that the addition of masking fields in the present experiment produced patterns of motion whose overall range of image velocity were roughly twice as large as those in the unmasked displays of Experiment 1. Because this change significantly reduced the magnitude of observers' curvedness adjustments, it seems

reasonable to conclude that they could not have been based on a heuristic analysis in which perceived depth is determined by the magnitude of relative image motion (see Liter, et al., 1993; Todd & Norman, 1991).

GENERAL DISCUSSION

When an object rotates in depth under orthographic projection, there is sufficient information within the pattern of optical flow to compute its 3-D structure up to a reflection in depth. In order to actually perform this computation, however, it is necessary to measure the instantaneous acceleration of each point or the higher order relations among three or more discrete views for objects observed in apparent motion. In the case of continuous rotation about a vertical axis, the depth z of any given point, relative to the rotation axis, is specified by

$$z = -V\sqrt{\frac{x}{A}}, \qquad (21)$$

where x is its horizontal position relative to the axis of rotation, and V and A are its projected velocity and acceleration. It is interesting to note, in this case, that the ratio x/A specifies the frequency of rotation, which must be the same for all points on a rigidly moving object.

If a visual system were incapable of measuring how the velocities (or displacements) of points change over time, a projected pattern of motion would only be sufficient to determine an object's structure up to a one-parameter affine transformation (see Koenderink, 1990; Todd & Bressan, 1990). The best that could be accomplished in that case would be to somehow guess the value of the unknown parameter, perhaps by assuming some ecological constraints on the overall range of possible object depths or rates of rotation. Any inaccuracies in this estimate would cause the perceived structure of an object to be systematically distorted by an affine stretching transformation.

There is a growing amount of evidence to suggest that this type of distortion is a typical characteristic of our visual perceptions of structure from motion. There have been numerous investigations reported in the literature, in which observers have judged the perceived depths of rotating objects, such as cylinders, ellipsoids, dihedral angles, or random configurations of points in a volume (e.g., see Liter & Braunstein, 1997; Liter et al., 1993; Tittle et al., 1995; Todd, 1984; Todd & Norman, 1991). In virtually all of these studies, there have been large constant errors in the magnitude of perceived depth scaling, though the size and direction of this effect can vary considerably over different viewing conditions.

Such findings suggest that human observers have relatively poor sensitivity to optical accelerations, so that the visual perception of structure from motion must be determined primarily by the pattern of image velocities (or displacements). Additional evidence to support this conclusion has been obtained by comparing the accuracy of observers' judgments for apparent motion sequences

with varying numbers of discrete frames. In general, there are few or no improvements in the accuracy of perceived structure as the number of distinct frames in a motion sequence is increased beyond two—provided that all other parameters are optimized at each sequence length to produce the best possible performance (e.g., see Liter et al., 1993; Todd & Bressan, 1990; Todd & Norman, 1991).

Another useful procedure for addressing this issue involves motion patterns with higher order structural deformations that are inherently undetectable from the first-order relations between individual pairs of views. This includes any pattern with parallel image trajectories for which all moving elements do not have the same ratio x/A (see Equation 21). Previous research has shown that different types of higher order structural deformations are not all perceptually equivalent. Although some such displays appear perfectly rigid, there are others that are easily identified as nonrigid, thus indicating that observers can obtain at least some useful information from changes over time in the pattern of image velocities or displacements (Liter & Braunstein, 1997; Norman & Todd, 1993; Perotti et al., 1996).

One important goal of the present research was to examine the effects of higher order structural deformations on observers' judgments of 3-D structure. In order to address this issue, rigid rotations of quadric surface patches were compared with constant flow field displays whose patterns of optical acceleration had no possible rigid interpretation. The results revealed that the rigid rotations produced slightly higher correlations between simulated and judged curvedness than did the constant flow field displays, but that observers were equally accurate in both conditions for their judgments of the shape characteristic. These findings are similar to those obtained by Liter and Braunstein (1997) for higher order structural deformations of a rotating dihedral angle. Observers' judgments of rigidity for these displays were significantly reduced relative to rigid rotations, but both types of motion produced comparable results for judgments of the magnitude of the depicted dihedral angle or of the extent of its rotation.

Although observers may be relatively poor at extracting information about 3-D structure from changes in image velocity over time, this does not appear to generalize to changes in image velocity over space. It has long been recognized that first-order spatial changes in velocity (i.e., curl, divergence, and shear) provide important information for the control of locomotion and the detection of impending collisions (see, e.g., Gibson, 1950; Koenderink & van Doorn, 1975, 1977; Todd, 1981). More recent analyses have demonstrated that second-order spatial changes in velocity can provide potentially useful information about certain aspects of surface curvature (Dijkstra et al., 1994; Droulez & Cornilleau-Pérès, 1990; Koenderink & van Doorn, 1992). Under appropriate conditions, these changes can specify the ratio of the two principal curvatures (i.e., the shape characteristic) and their directions, though it cannot specify other properties, such as curvedness.

There have been a number of empirical studies reported in the literature that have examined observers' sensitivity to the 3-D curvature of moving surfaces (see Cornilleau-Pérès & Droulez, 1989; Lappin et al., 1995; Norman & Lappin, 1992; van Damme et al., 1994). The present experiments were designed to extend this research by employing a response task with two degrees of freedom, so that we could separate the distinct components of shape characteristic and curvedness, and by including displays with first-order motion masks in an effort to restrict the relevant information to second-order spatial changes in velocity. The results obtained were highly consistent with the mathematical analyses proposed by Dijkstra et al. (1994) and by Koenderink and van Doorn (1992). Observers were almost perfectly accurate in their judgments of the shape characteristic, which is unambiguously specified by the instantaneous velocity field, but they produced large constant errors in their judgments of curvedness, which is not. Moreover, this performance was only minimally affected by altering the displays with a first-order motion mask, which suggests that the relevant information for performing these judgments involves second-order spatial changes in the pattern of image velocities or displacements.

An interesting question that arises from this research is why changes in image velocity over time should be more difficult to process than are changes in velocity over space. One possible explanation for this distinction is that mechanisms for measuring the accelerations of individual moving points would be highly sensitive to noise. Because visually sensitive neurons can only respond to fixed locations in space (i.e., their receptive fields), they are not well suited for tracking individual elements over multiple locations. The detection of spatial changes, in contrast, involves relationships among different moving elements. Thus, they can be measured by comparing the simultaneous outputs of different motion detectors, without having to worry about whether they were stimulated by the same identifiable feature.

REFERENCES

Bennett, B., & Hoffman, D. (1986). The computation of structure from fixed axis motion: Nonrigid structures. *Biological Cybernetics*, **51**, 293-300.

Bennett, B., Hoffman, D., Nicola, J., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America*, **6**, 1052-1069.

Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1990). Discriminating rigid from nonrigid motion: Minimum points and views. *Perception & Psychophysics*, **47**, 205-214.

Braunstein, M. L., Hoffman, D. D., Shapiro, L. R., Andersen, G. J., & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 335-343.

Cornilleau-Pérès, V., & Droulez, J. (1989). Visual perception of

surface curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics*, **46**, 351-364.

DE VRIES, S. C., KAPPERS, A. M. L., & KOENDERINK, J. J. (1994). Influence of surface attitude and curvature scaling on discrimination of binocularly presented curved surfaces. *Vision Research*, **34**, 2409-2423.

DIJKSTRA, T. M. H., SNOEREN, P. R., & GIELEN, C. C. A. M. (1994). Extraction of three-dimensional shape from optic flow: A geometric approach. *Journal of the Optical Society of America A*, **11**, 2184-2196.

DROULEZ, J., & CORNILLEAU-PÉRÈS, V. (1990). Visual perception of surface curvature: The spin variation and its psychological implications. *Biological Cybernetics*, **62**, 211-224.

GIBSON, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

HOFFMAN, D., & BENNETT, B. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A*, **2**, 242-249.

HOFFMAN, D., & BENNETT, B. (1986). The computation of structure from fixed axis motion: Rigid structures. *Biological Cybernetics*, **54**, 71-83.

HOFFMAN, D. D., & FLINCHBAUGH, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, **42**, 195-204.

KOENDERINK, J. J. (1990). *Solid shape*. Cambridge, MA: MIT Press.

KOENDERINK, J. J., & VAN DOORN, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Opt. Acta*, **22**, 773-791.

KOENDERINK, J. J., & VAN DOORN, A. J. (1977). How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In G. Hauske & F. Butenandt (Eds.), *Kybernetik* (pp. 224-247). Munich: Oldenburg.

KOENDERINK, J. J., & VAN DOORN, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 377-385.

KOENDERINK, J. J., & VAN DOORN, A. J. (1992). Second-order optic flow. *Journal of the Optical Society of America A*, **9**, 530-538.

LAPPIN, J. S., & AHLSTRÖM, U. B. (1994). On the scaling of visual space from motion—in response to Pizlo and Salach-Golyska. *Perception & Psychophysics*, **55**, 235-242.

LAPPIN, J. S., AHLSTRÖM, U. B., CRAFT, W. D., & TSCHANTZ, S. T. (1995). Spatial primitives for seeing three-dimensional shape from motion. In T. Papathomas (Ed.), *Early vision and beyond* (pp. 145-153). Cambridge, MA: MIT Press.

LAPPIN, J. S., & CRAFT, W. D. (1997). Definition and detection of binocular disparity. *Vision Research*, **37**, 2953-2974.

LAPPIN, J. S., DONER, J. F., & KOTTAS, B. L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science*, **209**, 717-719.

LITER, J. C., & BRAUNSTEIN, M. L. (1997). *The relationship of vertical and horizontal velocity gradients in the perception of shape, rotation and rigidity*. Manuscript submitted for publication.

LITER, J. C., BRAUNSTEIN, M. L., & HOFFMAN, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception*, **22**, 1441-1465.

LONGUET-HIGGINS, H. C., & PRAZDNY, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London: Series B*, **208**, 385-387.

MAMASSIAN, P. (1993). Illuminance critical points on generic smooth surfaces. In B. C. Vemuri (Ed.), *Geometric methods in computer vision II: Proceedings of SPIE* (Vol. 2031, pp. 124-133). Bellingham: International Society for Optical Engineering.

NORMAN, J. F., & LAPPIN, J. S. (1992). The detection of surface curvatures defined by optical motion. *Perception & Psychophysics*, **51**, 386-396.

NORMAN, J. F., & TODD, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics*, **53**, 279-291.

NORMAN, J. F., TODD, J. T., PEROTTI, V. J., & TITTLE, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 173-186.

PEROTTI, V. J., TODD, J. T., & NORMAN, J. F. (1996). The visual perception of rigid motion from constant flow fields. *Perception & Psychophysics*, **58**, 666-679.

PEROTTI, V. J., TODD, J. T., TITTLE, J. S., & NORMAN, J. F. (1994). The perception of 3D form from instantaneous flow components. *Investigative Ophthalmology & Visual Science*, **35**, 1317.

PHILLIPS, F., & TODD, J. T. (1996). Perception of local three-dimensional shape. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 930-944.

TE PAS, S., KAPPERS, A. M. L., & KOENDERINK, J. J. (1996). Detection of first-order structure in optic flow fields. *Vision Research*, **36**, 259-270.

TITTLE, J. S., TODD, J. T., PEROTTI, V. J., & NORMAN, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 663-678.

TODD, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 795-810.

TODD, J. T. (1984). The perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 173-186.

TODD, J. T., AKERSTROM, R. A., REICHEL, F. D., & HAYES, W. (1988). Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors. *Perception & Psychophysics*, **43**, 179-188.

TODD, J. T., & BRESSAN, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, **48**, 419-430.

TODD, J. T., & NORMAN, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics*, **50**, 509-523.

TODD, J. T., TITTLE, J. S., & NORMAN, J. F. (1995). Distortions of three-dimensional space in the perceptual analysis of motion and stereo. *Perception*, **24**, 75-86.

ULLMAN, S. (1977). *The interpretation of visual motion*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

VAN DAMME, W. J. M., OOSTERHOFF, F. H., & VAN DE GRIND, W. A. (1994). Discrimination of 3-D shape and 3-D curvature from motion in active vision. *Perception & Psychophysics*, **55**, 340-349.

WALLACH, H., & O'CONNELL, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217.

APPENDIX
Displacement Field for Discrete Rotations Under Parallel Projection

For the traditional Cartesian space $\Re^3$ (with the z-axis denoting depth), the horizontal position (x) of a point rotating about a vertical axis by some angle (θ) is

$$x' = x \cos \theta + z \sin \theta. \tag{A1}$$

This new position (x') can be expressed as the old position (x) plus some displacement in x (Δx):

$$x + \Delta x = x \cos \theta + z \sin \theta. \tag{A2}$$

The displacement field, D, can thus be specified as

$$D = \Delta x = x(\cos \theta - 1) + z \sin \theta. \tag{A3}$$

Substituting the surface equation, $z = l(x,y)$:

$$D = x(\cos \theta - 1) + l(x,y) \sin \theta. \tag{A4}$$

Adopting the notation that subscripts indicate spatial derivatives in the specified direction, the first spatial derivatives of the displacement field are

$$D_x = (\cos \theta - 1) + l_x \sin \theta \text{ and } D_y = l_y \sin \theta. \tag{A5}$$

And the second spatial derivatives of the displacement field are

$$D_{xx} = l_{xx} \sin \theta, D_{xy} = l_{xy} \sin \theta, \text{ and } D_{yy} = l_{yy} \sin \theta. \tag{A6}$$

The tilt component of the surface attitude is unambiguously specified for small rotations:

$$\tan \tau = \frac{l_x}{l_y} = \frac{D_x - (\cos \theta - 1)}{D_y} . \tag{A7}$$

But the slant component of the surface attitude is not

$$\tan \sigma = \sqrt{l_x^2 + l_y^2} = \csc \theta \sqrt{D_y^2 + (1 - D_x + \cos \theta}. \tag{A8}$$

The magnitude of the surface curvatures in the horizontal and vertical directions cannot be unambiguously determined from the velocity field:

$$\kappa_x = \frac{1}{\sqrt{1 + l_x^2 + l_y^2}} \left(\frac{l_{xx}}{1 + l_x^2} \right) = \frac{(\sin \theta)^2 V_{xx}}{\left[(\sin \theta)^2 + V_x^2 \right] \sqrt{(\sin \theta)^2 + V_x^2 + V_y^2}},$$

$$\tag{A9}$$

$$\kappa_y = \frac{1}{\sqrt{1 + l_x^2 + l_y^2}} \left(\frac{l_{yy}}{1 + l_y^2} \right) = \frac{(\sin \theta)^2 V_{yy}}{\left[(\sin \theta)^2 + V_y^2 \right] \sqrt{(\sin \theta)^2 + V_x^2 + V_y^2}},$$

$$\tag{A10}$$

However, the ratio of these is an unambiguous measure of surface shape for the case where the first spatial derivatives of zero are equal to one another:

$$\frac{\kappa_x}{\kappa_y} = \frac{l_{xx}}{l_{yy}} \left(\frac{1 + l_y^2}{1 + l_x^2} \right) = \frac{D_{xx}}{D_{yy}} \left(\frac{(\sin \theta)^2 + D_y^2}{(\sin \theta)^2 + D_x^2} \right). \tag{A11}$$