The visual perception of smoothly curved surfaces from minimal apparent motion sequences

JAMES T. TODD
Brandeis University, Waltham, Massachusetts
and Center for Adaptive Systems, Boston University, Boston, Massachusetts

and

J. FARLEY NORMAN
Brandeis University, Waltham, Massachusetts

A series of four experiments was designed to investigate the minimal amounts of information required to perceive the structure of a smoothly curved surface from its pattern of projected motion. In Experiments 1 and 2, observers estimated the amplitudes of sinusoidally corrugated surfaces relative to their periods. Observers' judgments varied linearly with the depicted surface amplitudes, but the amount of perceived relative depth was systematically overestimated by approximately 30%. The observers' amplitude judgments were also influenced to a lesser extent by the amount of rotary displacement of a surface at each frame transition, and by increasing the length of the apparent motion sequences from two to eight frames. The latter effect of sequence length was quite small, however, accounting for less than 3% of the variance in the observers' judgments. Experiments 3 and 4 examined observers' discrimination thresholds for sinusoidally corrugated surfaces of variable amplitude and for ellipsoid surfaces of variable eccentricity. The results revealed that observers could reliably detect differences of surface structure as small as 5%. The length of the apparent motion sequences had no detectable effect on these tasks, although there were significant effects of angular displacement and surface orientation. These results are considered with respect to the analysis of affine structure from motion proposed by Todd and Bressan (1990).

One of the most extensively investigated phenomena in human vision is the perception of three-dimensional structure from motion, often referred to as the *kinetic depth effect* (Wallach & O'Connell, 1953). A particularly compelling method of demonstrating this phenomenon is to present observers with apparent motion sequences of random-dot surfaces rotating in depth under orthographic projection. When any individual image from such a sequence is viewed in isolation, it appears as nothing more than a pattern of random dots in the picture plane. When several such images are presented in rapid succession, however, the depicted pattern is almost always perceived as a coherent surface rotating rigidly in three-dimensional space.

What type of perceptual mechanism would allow observers to determine the three-dimensional structure of an object solely on the basis of its pattern of projected motion? In attempting to develop a computational analysis of this phenomenon, numerous investigators have noted that an apparent motion sequence of an arbitrary rigid object under orthographic projection must contain a minimum of three distinct views in order to obtain a unique interpretation of the object's three-dimensional structure, and that this defines a theoretical upper bound on the perceptual abilities of ideal observers (e.g., see Bennett, Hoffman, Nicola, & Prakash, 1989; Huang & Lee, 1989; Ullman, 1977, 1979, 1983). During the past several years, however, there has been increasing evidence that this theoretical limit may have surprisingly little relevance to actual human vision. Of particular importance in this regard are the recent findings from several different laboratories that two-frame motion sequences presented in alternation can provide sufficient information to obtain compelling kinetic depth effects and to accurately discriminate between different three-dimensional structures (Braunstein, Hoffman, & Pollick, 1990; Braunstein, Hoffman, Shapiro, Andersen, & Bennett, 1987; Doner, Lappin, & Perfetto, 1984; Lappin, Doner, & Kottas, 1980; Todd, Akerstrom, Reichel, & Hayes, 1988; Todd & Bressan, 1990). Similar results can also be obtained by using longer length sequences of scintillating random-dot surfaces for which no dot is allowed to survive for more than two successive frames (Dosher, Landy, & Sperling, 1990; Norman, 1990; Todd, 1985).

To better appreciate the potential information that is available from two-frame apparent motion sequences, it

is useful to consider a widely overlooked analysis of Ullman's (1977). Ullman noted that any rotary displacement of an object can be decomposed into two components: a rotation about an axis in the picture plane that produces a pattern of parallel image trajectories, and a subsequent rotation about the line of sight that produces a pattern of curl in the image. He also developed a specific algorithm for mathematically transforming any two-frame motion sequence to eliminate the curl component. (See Todd & Bressan, 1990, for more details.) One important property of this algorithm is that it provides a useful test for the rigidity of an object's motion. If elimination of the curl component does not result in a pattern of parallel trajectories under orthographic projection, the depicted object cannot have a possible rigid interpretation (see also the related findings of Bennett et al., 1989; and Koenderink & van Doorn, 1991).

Another useful property of Ullman's (1977) analysis is that it allows us to reduce any rigid displacement of an object to the special case of rotation in depth about an axis in the image plane, which greatly simplifies any subsequent analyses of the object's three-dimensional structure. In this special case, the instantaneous depth of any point relative to the rotation axis is optically specified within an indeterminate scale factor by its projected velocity (see Braunstein & Todd, 1990; Todd & Bressan, 1990). Although this does not allow a *unique* rigid interpretation, it severely constrains the structure of the depicted object to a one-parameter family of rigid interpretations. It is important to note in particular that any variation in the assumed value of the indeterminate depth scaling parameter is mathematically equivalent to subjecting the depicted object to an affine "stretching" transformation along the line of sight (cf. Koenderink & van Doorn, 1991). It follows, therefore, that any structural property of an object that remains invariant under affine transformations can be optically specified from two-frame apparent motion sequences.

For the sake of convenience, we shall refer to such properties as *affine structure*, but it is important not to be misled by this nomenclature. The inherent ambiguity of two-frame apparent motion sequences is restricted to affine stretching transformations along the line of sight, which does not encompass the entire class of possible affine transformations. Thus, two objects that are affine equivalent in the more general sense can still be discriminated from a two-frame apparent motion sequence if they are related by a stretching transformation in any direction that is not parallel to the line of sight.

A wide variety of object properties can be reliably detected solely on the basis of an analysis of affine structure. For example, it is possible with such an analysis to determine the metric length ratio between any pair of parallel line segments; to perform various nominal categorizations, such as distinguishing between planar and nonplanar configurations; and to accurately discriminate structural differences between any pair of objects that cannot be made congruent by an affine stretching transfor-

mation along the line of sight. It is also interesting in this regard that an analysis of affine structure from two-frame sequences is sufficiently powerful to perform most of the existing psychophysical tasks that have been employed previously to study observers' perceptions of structure from motion, including judgments of rigidity or coherence (e.g., Braunstein, 1962; Dosher, Landy, & Sperling, 1989; Green, 1961; Petersik, 1979, 1980; Todd et al., 1988), discriminations of rigid from nonrigid motion (e.g., Braunstein et al., 1990; Cutting, 1987; Doner et al., 1984; Lappin et al., 1980; Petersik, 1987; Todd, 1982), judgments of ordinal depth relations (e.g., Hildreth, Grzywacz, Adelson, & Inada, 1990; Reichel & Todd, 1990), and the discrimination or identification of complex three-dimensional forms (e.g., Braunstein et al., 1987; Cornilleau-Pérès & Droulez, 1989; Dosher et al., 1990; Husain, Treue, & Andersen, 1989; Loomis & Eby, 1988; Sperling, Landy, Dosher, & Perkins, 1989; Todd, 1984, 1985; Todd & Bressan, 1990; Treue, Husain, & Andersen, 1991).

Although the analysis of affine structure from motion may appear at first blush to be surprisingly powerful, it also has some important limitations that need to be highlighted. Knowing the affine structure of an object defines a ratiometric scale of distance in any given direction, but the resulting distance metric is anisotropic—that is to say, it would allow one to determine the relative length ratio of any pair of parallel or collinear line segments (e.g., see Lappin & Fuqua, 1983), but it cannot be used to determine the relative lengths (or angles) of line segments oriented in different directions. The latter judgments would require an analysis of euclidean structure from motion, in which information must be integrated over a minimum of three distinct frames.

There is some evidence to suggest, however, that the human visual system may be relatively insensitive to the higher order relations among three or more frames that are required for an analysis of euclidean structure, and that observers' perceptions of the kinetic depth effect may therefore be primarily determined by a more abstract analysis of affine structure. Todd and Bressan (1990) have recently reported two converging lines of evidence to support this hypothesis: First, they demonstrated that observers' performance on tasks involving judgments of affine structure was an order of magnitude greater than performance on similar tasks requiring an analysis of euclidean structure; second, they also showed that performance did not improve on any of these tasks as the number of distinct frames in an apparent motion sequence was increased from two to eight. Since analyses of euclidean structure require a minimum of three distinct views in order to be effective with arbitrary structural configurations, the latter finding provides especially strong evidence that these analyses may have little relevance to actual human perception.

Whereas the experiments of Todd and Bressan (1990) were primarily concerned with arbitrary configurations of randomly positioned dots or lines, the research de-

scribed in the present article was designed to address a related set of issues for the visual perception of smoothly curved surfaces. Moving surfaces are different from other more general configurations in that they produce optical flow fields that are smoothly continuous over space (except at occlusion boundaries). Because this is such a common and easily verifiable special case of optical motion, numerous computational models have been designed specifically to exploit the differential structure of smooth flow fields for determining a moving object's three-dimensional form (Aloimonos & Brown, 1989; Hoffman, 1982; Koenderink & van Doorn, 1975, 1977; Lappin, 1990; Longuet-Higgins & Prazdny, 1984; Waxman & Ullman, 1985). Can human observers exploit this property as well? If so, the relative imprecision of perceived euclidean metric structure reported by Todd and Bressan (1990) might not generalize to observers' perceptions of smoothly curved surfaces. In the present series of experiments, we have attempted to explore this hypothesis by using a variety of converging operations, involving both subjective depth ratings and surface discrimination procedures.

EXPERIMENT 1

Method

The experiment was controlled by a Masscomp 5600 computer with a Lex-90 graphics system. The stimuli were presented within a 33 × 26 cm rectangular region of the display screen with a spatial resolution of 1,280 × 1,024 pixels. The displays were observed through a monocular viewing hood at a distance of 76 cm, such that each pixel spanned a visual angle of approximately 1.15′. All responses were recorded by pressing keys on the computer keyboard.

Each stimulus was designed to simulate an orthographic projection of a sinusoidally corrugated surface (with vertically aligned corrugations) composed of 10,000 discrete points that were distributed randomly in both the horizontal and the vertical dimensions. The spatial period of the sinusoidal corrugations was 320 pixels—that is, they had a spatial frequency of 0.163 cycles per degree. The peak-to-trough amplitude of the corrugations was varied across trials, so that the ratio of amplitude/period had five possible values of 0.19, 0.38, 0.56, 0.75, or 0.94. The observer's task on each trial was to estimate the ratio of perceived surface amplitude relative to the period in increments of 0.1.

To create a visual impression of three-dimensional structure, each surface was presented in apparent motion, rotating in depth about a horizontal axis that was parallel to the direction of maximum surface curvature. The length of the apparent motion sequence was varied across trials from a set of possible sequence lengths of two, three, four, six, or eight distinct frames. On the first frame of every sequence, the depicted surface was oriented at a 20° angle relative to the vertical, which was increased by 1.5° at each subsequent frame transition. Each sequence was presented in continuous oscillation—first in one direction, then in reverse—until an appropriate response was recorded.

A particularly important consequence of orienting the axis of rotation so that it was orthogonal to the sinusoidal corrugations of each surface is that the resulting motion patterns produced no systematic spatial gradients of texture density that might have provided a potentially confounding source of static information about three-dimensional structure. Consider, for example, the pair of dot patterns presented in Figure 1, which include Frames 1 and 8 of an eight-frame sequence for a surface with the largest possible amplitude/period ratio of 0.96. Note in the figure that the individual static images
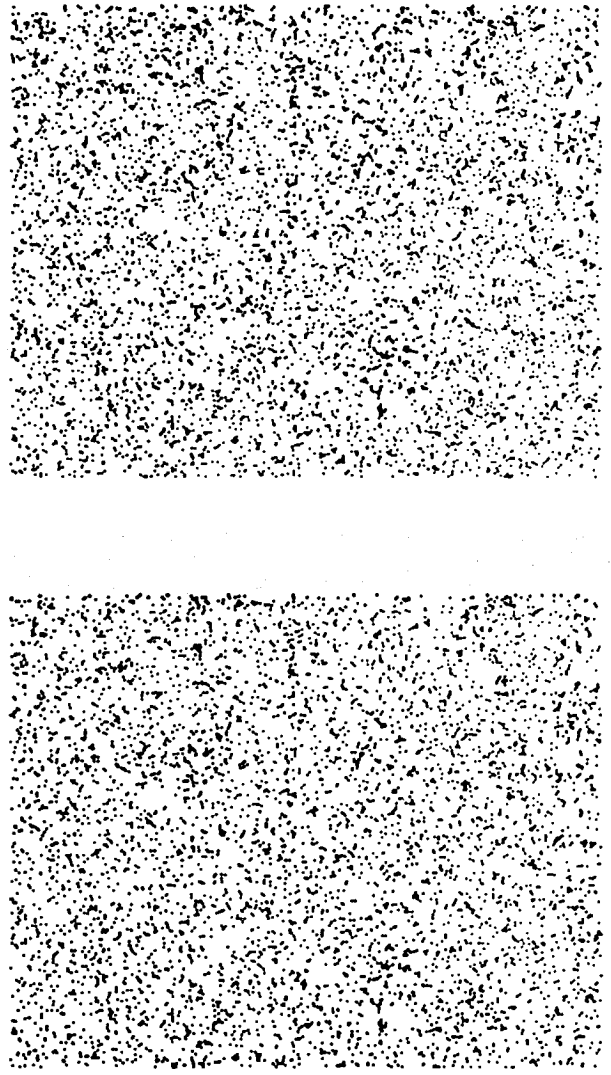

Figure 1. A pair of dot patterns from an apparent motion sequence in Experiment 1, depicting a sinusoidally corrugated surface with an amplitude/period ratio of 0.96. Note that each individual image contains a homogeneous pattern of random noise with no available information about the depicted three-dimensional structure. When the same images are presented as part of an apparent motion sequence, however, they appear quite clearly as a sinusoidally corrugated surface. A similar effect can also be obtained by turning the page sideways and viewing the images stereoscopically.

appear as homogeneous patterns of random noise. If the same set of images were viewed in apparent motion, however, it would produce a compelling impression of a sinusoidally corrugated surface in depth. (A similar effect can also be observed by turning the page sideways and viewing the patterns stereoscopically.)

A preliminary experiment, based on the earlier findings of Todd et al. (1988) and Todd and Bressan (1990), was performed to determine the optimal timing parameters for each possible sequence length. Three naive observers compared the relative salience of perceived structure from motion for similar displays with different stimulus onset asynchronies (SOAs), which could be varied in increments of 1/60 sec as constrained by the 60-Hz raster refresh rate. For each display, the observers were asked to judge whether the perceived salience of the depicted surface could be improved by
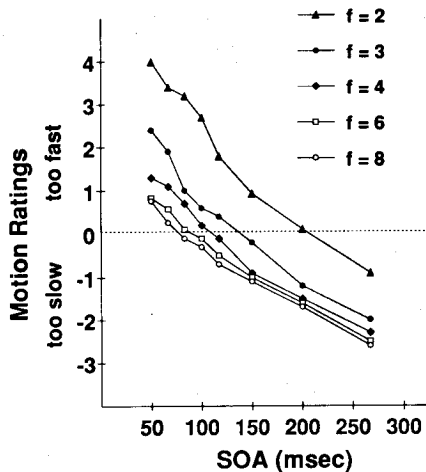
Figure 2. The perceived salience of structure from motion for 3 naive observers as a function of the frame-to-frame stimulus onset asynchrony in an investigation preliminary to the present series of experiments. The curve parameter f indicates the number of distinct frames in an apparent motion sequence.

speeding up or slowing down its rate of oscillation. They were asked to respond using a 9-point scale from −4 to 4, where 0 represented an optimal rate of apparent motion, and positive (or negative) values were used to designate the extent to which the oscillation appeared too fast (or too slow). The results of this preliminary experiment are shown in Figure 2. As is evident from the figure, the optimal SOAs varied systematically with sequence length, which was also confirmed by the observers' subjective reports. Longer length sequences with slow presentation rates often appeared discontinuous, whereas shorter length sequences with high presentation rates tended to appear nonrigid (see Todd et al., 1988). Indeed, at the most extreme values of the latter category (i.e., a two-frame sequence with a 50-msec SOA), the resulting displays do not elicit a compelling kinetic depth effect. For sequence lengths of two, three, four, six, and eight distinct frames, the estimated optimal SOA values were 200, 150, 117, 100, and 83 msec, respectively, which produced oscillation rates of 2.50, 1.67, 1.43, 1.00, and 0.86 Hz (cf. Todd & Bressan, 1990). These values were employed for all of the experiments in the present series, with an interstimulus interval of 0 msec.

The design of Experiment 1 included all possible combinations of the different stimulus manipulations described above, resulting in 25 distinct conditions (5 sequence lengths × 5 amplitude/period ratios). Five naive observers evaluated the displays over two experimental sessions. Each session contained four blocks of trials, and each stimulus condition was presented once per block. The first block of each session was considered practice and was excluded from all subsequent data analyses. All of the observers reported that the displays appeared quite clearly as moving surfaces in three-dimensional space, and they all expressed considerable confidence in the accuracy of their judgments. No feedback was provided about their performance.

At the beginning of each experimental session, the observers were explicitly warned that the displays might sometimes appear to reverse in depth spontaneously. Because the displays were generated under parallel projection, the overall direction of slant for the depicted surfaces was mathematically indeterminate, and it was reasonable to expect, therefore, that the perceived orientations might be perceptually multistable. Although this should not have been relevant to the experimental task, it is interesting that all of the observers reported that spontaneous reversals in the perceived direction of surface orientation were remarkably rare. As has recently been in-

vestigated by Reichel and Todd (1990), there is a strong bias to perceive the slant of a depicted surface in such a way that the lower regions of a display appear to be closer in depth.

Results and Discussion

The combined results of all 5 observers are presented in Figure 3. An analysis of variance for these data revealed that the simulated amplitude/period ratios had a significant effect on the observers' judgments [$F(4,16) = 213.3, p < .01$], accounting for over 84% of the total sum of squares. The observers' judgments were also influenced to some extent by the length of the apparent motion sequences [$F(4,16) = 6.5, p < .01$], but this effect was quite small, accounting for only 2% of the total sum of squares.

In a further attempt to assess the accuracy of the observers' perceptions of three-dimensional structure, we measured the linear correlation between the judged amplitude/period ratios in each condition with the actual values used to generate the displays. The analysis revealed a clear linear relation ($r = .98$), with an intercept of 0.036 and a slope of 1.31. That is to say, the results demonstrated that the observers' judgments varied linearly with the depicted three-dimensional structure, but that the perceived amplitudes of the surface corrugations were over 30% larger than what was actually specified by the geometry of the displays.

Let us now consider the available sources of optical information from which the observers' judgments of three-dimensional structure could potentially have been determined. One important finding that is especially relevant in addressing this issue is that there were only negligible improvements in performance as the number of distinct frames in the apparent motion sequences was increased
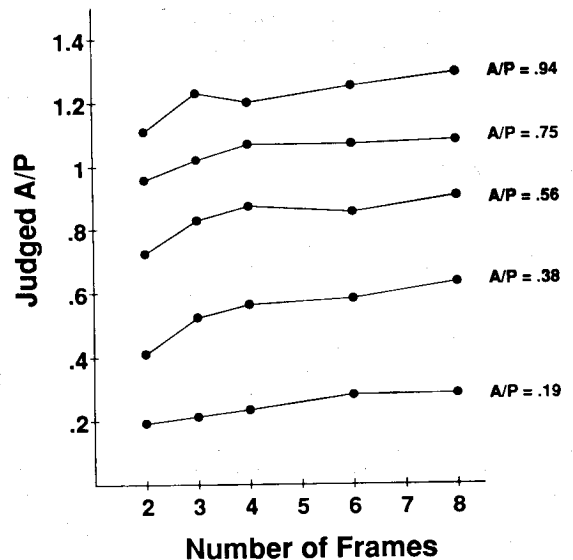

Figure 3. The perceived amplitude/period ratios of 5 naive observers in Experiment 1, as a function of the number of distinct frames in an apparent motion sequence. The curve parameter A/P indicates the actual amplitude/period ratio of the depicted surface.

from two to eight. It seems reasonable to conclude, on the basis of this finding, that the perceptually relevant information for the observers' judgments was fully present within the first-order field of element displacements between a single pair of random-dot images.

For small displacements of a surface rotating about an axis in the picture plane under parallel projection, the optical motion of any point (i) is defined by the following equation:

$$D_i = \epsilon Z_i, \qquad (1)$$

where Z_i is its position in depth relative to the rotation axis, D_i is its projected displacement in the picture plane, and ϵ is the angular displacement of the surface in three-dimensional space (see Braunstein & Todd, 1990; Todd & Bressan, 1990). For the displays used in the present experiment, the instantaneous depth of each point relative to the rotation axis was defined as follows:

$$Z_i = Y_i \cot(\alpha) - a \cos(\phi X_i)/\sin(\alpha), \qquad (2)$$

where X_i and Y_i are the horizontal and vertical position coordinates in the picture plane, ϕ is the spatial frequency of the surface corrugations in depth, a is the amplitude of those corrugations, and α is the slant of the surface relative to the line of sight. Combining these two equations, the field of projected element displacements is given by:

$$D_i = \epsilon[Y_i \cot(\alpha) - a \cos(\phi X_i)/\sin(\alpha)]. \qquad (3)$$

One possible strategy by which the observers could have performed this task would be to estimate some default value of ϵ, and to use that estimate in Equation 1 to determine the relative depth of each element in the display. If the value of ϵ is estimated incorrectly, this strategy will produce a predictable pattern of errors. The possible family of surfaces consistent with Equation 3 are all sinusoidally corrugated with an invariant relationship between their amplitudes and slants. If ϵ' is the assumed value of ϵ, and if all other measurements and computations are performed with perfect accuracy, the perceived angle (α') of the surface relative to the line of sight would be:

$$\alpha' = \tan^{-1}[(\epsilon' \tan(\alpha)/\epsilon)], \qquad (4)$$

and the perceived surface amplitude (a') would be:

$$a' = a \cos(\alpha')/\cos(\alpha). \qquad (5)$$

Note that this strategy could lead to either underestimations or overestimations of surface amplitude, depending on the particular relation between ϵ' and ϵ. When $\epsilon' > \epsilon$, the perceived relative depth of a moving object would be systematically underestimated, as has been observed in previous studies by Loomis and Eby (1988) and Todd (1984). When $\epsilon' < \epsilon$, on the other hand, the perceived relative depth would be systematically overestimated, as occurred in the present experiment and in other related studies by Tittle, Braunstein, and Liter (1990) and Todd and Bressan (1990). If this is indeed the strategy by which observers estimate the amplitude/period ratio of a sinusoidally corrugated surface, it ought to be possible to systematically alter these judgments by manipulating the size of the rotary displacement ϵ used to generate an apparent motion sequence. Experiment 2 of the present series was designed to investigate this prediction.

EXPERIMENT 2

Method

The apparatus and general procedure were identical to those used in Experiment 1. Observers judged the amplitude/period ratios of sinusoidally corrugated surfaces rotating in depth about a horizontal axis. The only difference with respect to the earlier experiment involved the possible values of various parameters used to generate the displays. For these displays, there were only two possible amplitude/period ratios, 0.38 or 0.75, and three possible sequence lengths, containing two, four, or eight distinct frames. The rotary displacement of the surface at each frame transition was varied across trials from possible values of 1°, 1.5°, or 2°.

All possible combinations of these different stimulus parameters were used, so the resulting experimental design had 18 distinct conditions (3 sequence lengths × 2 amplitude/period ratios × 3 angular displacements). These were presented without response feedback to 5 naive observers over two experimental sessions. As in the previous experiment, each session contained four blocks of trials, the first of which was considered practice and was excluded from all subsequent data analyses.

Results and Discussion

The combined results of all 5 observers are presented in Figure 4. An analysis of variance for these data revealed that the observers' judgments were primarily de-
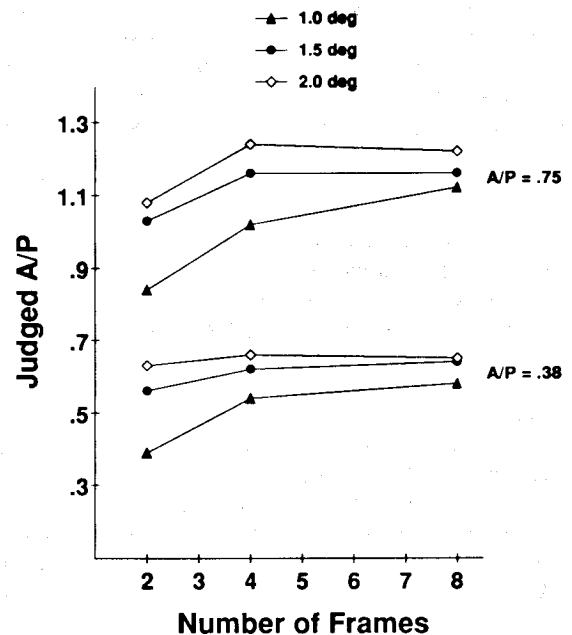

Figure 4. The perceived amplitude/period ratios of 5 naive observers in Experiment 2, as a function of the number of distinct frames in an apparent motion sequence. The curve parameter A/P indicates the actual amplitude/period ratio of the depicted surface. The different possible angular displacements of 1.0°, 1.5°, and 2.0° are represented by filled triangles, filled circles, and open diamonds, respectively.

termined by the simulated amplitude/period ratios [$F(1,4)$ = 100.9, $p < .01$], which accounted for over 64% of the total sum of squares. There were also significant effects of angular displacement [$F(2,8) = 26.1, p < .01$] and sequence length [$F(2,8) = 11.9, p < .01$], whose relative proportions of the total sum of squares were 5% and 3%, respectively. There were no significant interactions.

Our main purpose in performing this experiment was to test the hypothesis that observers' judgments would be based on a simple linear mapping between projected velocity and perceived depth relative to the rotation axis. If that were the case, then doubling the range of projected element displacements by increasing the rotary displacement of a surface from 1° to 2° (see Equation 1) should have produced a 90% increase of perceived surface amplitude (see Equations 4 and 5). As is clearly evident in Figure 4, this prediction was not confirmed. Although there was a statistically significant effect of rotary displacement, the effect was much smaller than would have been predicted on the basis of a simple linear mapping hypothesis. Indeed, as the rotary displacement was increased from 1° to 2°, the corresponding increase in the observers' judgments was only 23%. When the simulated amplitude/period ratios were doubled from 0.38 to 0.75, in contrast, the observers' judgments increased by over 87%.

These findings provide strong evidence that observers' perceptions of three-dimensional structure in these displays are determined, at least in part, by some analysis of the optical flow fields that is more complex than a simple linear mapping between projected displacement and perceived depth. Whatever this analysis is, it seems to provide some degree of constancy over changes in an object's angular displacement (or velocity). There are several possible strategies by which this could be achieved. For example, one reasonable approach would be to always select an interpretation of an object that minimizes some measure of structural complexity (see, e.g., Aloimonos & Brown, 1989). Because the set of possible rigid interpretations remains invariant over small variations in an object's rotary displacement, any strategy of this type would be expected to produce perfect displacement constancy. Another possible strategy would be to rescale the field of projected element displacements by some global measure that varies proportionally with the magnitude of an object's rotary displacement in three-dimensional space. For example, Loomis and Eby (1988) have suggested that perceived relative depth may be determined by image shear divided by the average projected velocity. Similarly, Tittle et al. (1990) have proposed that image displacements can be rescaled by the overall magnitude of compression.

It remains to be demonstrated by future research whether any of these strategies can adequately account for observers' perceptions of rotating sinusoidal surfaces. The primary difficulty in this regard is that displacement constancy is imperfect—that is to say, when the simulated angular displacements were varied in the present experi-

ment, there was a corresponding gain of perceived surface amplitude of approximately 23%. Precisely how this occurs remains unclear. What we can conclude with more confidence, however, is that the mechanism responsible for this partial constancy seems to be restricted to an analysis of the first-order field of projected element displacements, and that it cannot make use of other potential constraints that are available in principle from integrating information over three or more successive views.

EXPERIMENT 3

In the two experiments described above, the magnitude of perceived depth was measured in a numerical estimation task. Although it is clear from the results of these experiments that observers are capable of discriminating differences in amplitude of sinusoidally corrugated surfaces, the procedure was not sufficiently powerful to determine the precise limits of their performance. Experiment 3 was designed, therefore, to measure observers' amplitude discrimination thresholds for variable-length apparent motion sequences with variable angular displacement.

Method

The apparatus and stimuli were identical to those used in the previous experiments, but the procedure was quite different. A method of constant stimuli was employed, in which observers judged whether the depicted amplitude/period ratio in each display was larger or smaller than an implicit standard of 0.75. Six different test stimuli were employed with possible amplitude/period ratios that differed from the standard by −8.3%, −5%, −1.7%, 1.7%, 5%, or 8.3%, respectively.

In addition to the variations in surface amplitude, the displays were presented with two possible sequence lengths of two or eight distinct frames, and two possible angular displacements at each frame transition of 1.5° or 2°. All possible combinations of these different stimulus parameters were used, so the resulting experimental design had 24 distinct conditions (6 amplitudes × 2 sequence lengths × 2 angular displacements). Five practiced observers evaluated the displays over five experimental sessions, and each condition was presented 10 times per session in a random sequence. At the beginning of each session, the observers were given 24 trials of practice to make it possible for them to adapt themselves to the implicit standard. Response feedback was provided after every trial throughout each experimental session.

Results and Discussion

Figure 5 shows the percentage of trials in which observers responded "larger," as a function of the percentage difference between each possible test stimulus and the implicit standard. To achieve perfect performance on this task, an ideal observer would have to always respond "larger" whenever there was a positive difference between the test stimulus and the standard, and to always respond "smaller" whenever there was a negative difference. As is evident from the figure, the actual observers were far from perfect, and their performance varied systematically with the size of the difference they were required to detect.
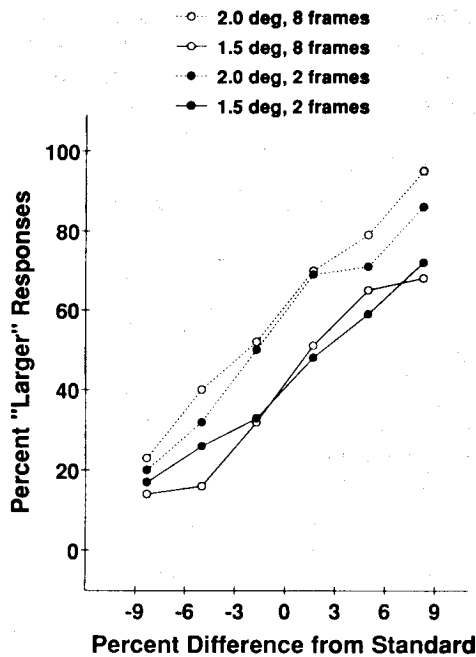
Figure 5. The percentage of "larger" responses for 5 practiced observers in Experiment 3, as a function of the percentage difference between the depicted amplitude of a test stimulus and the implicit standard. The two- and eight-frame apparent motion sequences are represented by filled and open circles, while the 1.5° and 2.0° angular displacements are represented by solid and dotted lines.

An analysis of variance of these data revealed significant effects of surface amplitude [$F(5,20) = 88.36, p < .01$] and angular displacement [$F(1,4) = 22.6, p < .01$]. The effect of sequence length was not statistically significant as it was in the two previous experiments. We also computed an overall discrimination threshold for this task by converting the data to percent correct and collapsing over variations in angular displacement and sequence length. The results of this analysis revealed a mean discrimination threshold of 6.7%. That is to say, the observers were able to detect a 6.7% variation of surface amplitude relative to the standard with 75% accuracy.

One important issue to consider in evaluating these results is the potential use of two-dimensional velocity cues for performing the required discriminations. Sperling et al. (1989) have recently warned against the use of structure-from-motion tasks for which accurate performance could potentially be achieved by measuring projected velocities in just one or two areas of a display (see, however, Braunstein & Todd, 1990). Could observers have employed such a strategy in Experiment 3? It is important to keep in mind with respect to this issue that the manipulations of angular displacement in our displays produced much larger variations in projected velocity than did the manipulations of surface amplitude, which the observers were instructed to discriminate. Thus, if their judgments had been based instead on a simple comparison of two-dimensional image velocities, the manipula-

tions of angular displacement should have had the largest effect on performance. Note, however, that this is not what occurred in the actual experiment. Although there were significant effects of angular displacement, they accounted for only 11% of the total sum of squares, whereas the effects of surface amplitude accounted for over 77%. It is clear from this finding that the observers' judgments could not have been based on a simple comparison of the two-dimensional image velocities.

From a theoretical point of view, one of the most important aspects of these results is that performance did not improve significantly as the number of distinct frames in an apparent motion sequence was increased from two to eight. When considered in conjunction with the similar pattern of results obtained in Experiments 1 and 2 using a magnitude-estimation task and the earlier discrimination studies of Braunstein et al. (1990), Braunstein et al. (1987), and Todd and Bressan (1990) using random configurations of dots or lines, these findings provide strong converging evidence that the perceptual analysis of structure from motion by actual human observers may be restricted to the use of first-order velocity (or displacement) measures that are available within two-frame apparent motion sequences. Moreover, because it is not mathematically possible to uniquely determine the euclidean metric structure of an arbitrary object from a two-frame apparent motion sequence, an additional implication of these results is that perceived three-dimensional structure from parallel projected motion must be inherently ambiguous up to an affine stretching transformation along the line of sight.

It is important to recognize in this regard that although the surfaces employed in the present experiments were affine equivalent in the more general sense of the term, they were not affine equivalent along the line of sight, and they were therefore discriminable from two-frame apparent motion sequences. To appreciate better why this is so, it is useful to reconsider the structure of the field of projected element displacements at each frame transition, as described by Equation 3.

$$D_i = \epsilon[Y_i \cot(\alpha) - a \cos(\phi X_i)/\sin(\alpha)].$$

The parameter α in this equation defines the overall surface orientation relative to the line of sight. For most values of α, including the range between 56° and 70° used in the present experiments, the optical effects of varying surface amplitude (a) are clearly different from those produced from comparable variations in the amount of angular displacement ϵ. There is one degenerate orientation, however, for which this is not the case. When the surface corrugations are all oriented in parallel with the picture plane (i.e., when $\alpha = 90°$), Equation 3 is reduced to:

$$D_i = \epsilon a \cos(\phi X_i). \tag{6}$$

Because all possible variations of surface structure defined by this reduced equation are affine equivalent along the line of sight, the optical effects of varying a or ϵ in that

case would be mathematically (and perceptually) indistinguishable.

Thus, from the perspective of an analysis of affine structure from motion, the ability of observers to discriminate variations of surface structure from comparable variations in angular displacement is critically dependent on the specific orientations from which a surface is viewed. It follows from this conclusion that some pairs of frames within a multiple-frame apparent motion sequence may be more perceptually informative than others for performing a discrimination task. As an object rotates continuously in depth, it may sometimes pass through a degenerate orientation for which it is affine equivalent along the line of sight with some other specific comparison object. As it continues to rotate, however, it will eventually arrive at a new orientation for which this particular structural ambiguity will no longer be present.

To demonstrate the generality of this point, it is useful to consider an alternative structural configuration from the sinusoidally corrugated surfaces employed in the present experiments. Equation 7 defines a family of ellipsoid surfaces, which can all be generated by stretching a unit sphere along the z-axis by a variable proportion q:

$$X^2 + Y^2 + Z^2/(1+q)^2 = 1. \qquad (7)$$

Suppose that a particular surface within this family is viewed under parallel projection from a position P_0 such that the line of sight forms an angle θ with respect to the z-axis (see Figure 6). If the surface is rotated about the y-axis by a small angular displacement ϵ, the optical motion of any point i is defined by the following quadratic relation:

$$(D_i/\epsilon)^2[\sin^2(\theta)+\cos^2(\theta)/(1+q)^2]$$
$$+ (D_i/\epsilon)\{2X_i \sin(\theta) \cos(\theta)[1-1/(1+q)^2]\}$$
$$+ [X_i^2 \cos^2(\theta)+X_i^2 \sin^2(\theta)/(1+q)^2+Y_i^2-1] = 0, \quad (8)$$

where X_i and Y_i are its horizontal and vertical position coordinates, and D_i is its projected displacement in the image plane. Note in this equation that for most possible values of θ, the effects of q and ϵ on the overall pattern of optical flow are quite different. The degenerate orientation for which the possible variations of surface structure are all affine equivalent along the line of sight occurs when $\theta = 0$. Equation 8 can be reduced in that case to:

$$D_i^2 = \epsilon^2(1+q)^2 (1-X_i^2-Y_i^2). \qquad (9)$$

It should be evident from this equation that any variations of ϵ and q will have identical effects on the instantaneous flow pattern, and cannot therefore be distinguished within a two-frame apparent motion sequence.

This last example is especially instructive for interpreting the results of a recent experiment by Norman (1990). Norman measured observers' discrimination thresholds for rotating random-dot ellipsoids similar to those described above, using apparent motion sequences composed of either 2 or 15 distinct frames. The results of this experiment revealed what appeared to be a large effect of sequence length, in that the 2-frame displays produced much higher discrimination thresholds than did the 15-frame displays (see also Loomis & Eby, 1988). Could this be a relevant counterexample to the conclusions of Todd and Bressan (1990) concerning observers' insensitivity to higher order relations among 3 or more frames of an apparent motion sequence? Before drawing any premature conclusions, however, there is one important aspect of Norman's experimental design that needs to be highlighted. For the 2-frame displays employed in this study, the depicted ellipsoids were always presented at a degenerate orientation for which the possible variations of surface structure were all affine equivalent along the line of sight, and could not therefore be reliably discriminated solely on the basis of an analysis of affine structure from motion. For the 15-frame displays, in contrast, the ellipsoids were rotated up to 35° away from this degenerate orientation, so that some pairs of frames within the overall sequence would have provided sufficient information about differences in affine structure to allow accurate performance on the required discrimination task. In other words, what might appear at first blush to be a clear counterexample to the theory of affine structure from motion proposed by Todd and Bressan (1990) might turn out instead to provide strong confirmation for one of the fundamental predictions of that theory.

One possible method of testing this hypothesis would be to include an additional two-frame condition in which each depicted object would be viewed at an angle of 35° relative to the degenerate orientation. If the number of distinct frames in an apparent motion sequence is a critical parameter for the observer's performance, then this new two-frame condition should yield significantly higher discrimination thresholds than those obtained with longer length sequences. If, on the other hand, the depicted orientation of a surface is the critical parameter, this new two-frame condition should produce levels of performance that are roughly comparable to those obtained with multiframe displays. Experiment 4 was designed to test these predictions.
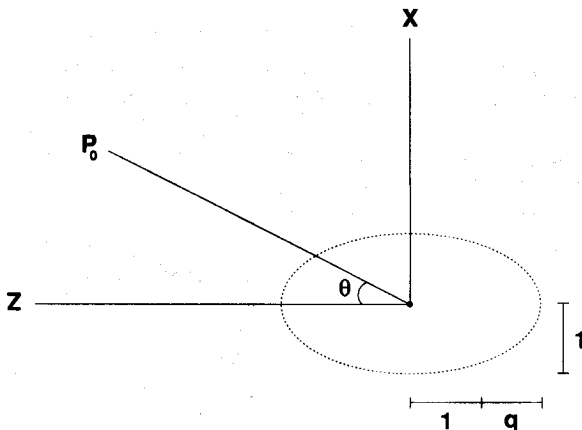

Figure 6. An ellipsoid surface that differs from a unit sphere by a variable proportion q. The surface is viewed from a position P_0 at an angle θ from the axis of variable eccentricity.

EXPERIMENT 4

Method

The basic procedure employed in this study was designed to conform as closely as possible to the earlier experiment by Norman (1990) described above. The displays were generated from a "standard" sphere with a radius of 116 pixels (3 cm). For each display, this sphere would be covered with 900 randomly positioned dots, which would then be expanded or compressed along the z-axis by a variable proportion q as defined by Equation 7. The resulting ellipsoid surfaces were then presented in apparent motion, rotating in depth about a vertical axis with their occluded dots removed. The observer's task on each trial was to indicate whether the three-dimensional structure of the depicted surface appeared "expanded" or "compressed," by pressing an appropriate key on the computer keyboard. Response feedback was presented after each trial.

There were three different experimental conditions in which the orientations of the depicted ellipsoids and the length of the apparent motion sequences were systematically manipulated. These will be referred to as the eight-frame condition, the two-frame slanted condition, and the two-frame straight condition.

For the eight-frame condition, each depicted ellipsoid appeared initially at the degenerate orientation for which the possible variations of surface structure were all affine equivalent along the line of sight. It then rotated to the left with a frame-to-frame angular displacement of either 4° or 6°, which was determined at random on each trial. As in our previous experiments, these eight-frame displays were presented with an SOA of 83 msec and an inter-stimulus interval of 0 msec. Each apparent motion sequence was presented in continuous oscillation—first in one direction, then in reverse—until an appropriate response was recorded. The difficulty of the discrimination task was systematically manipulated by adjusting the proportion, q, with which the standard sphere was deformed to produce expanded or compressed ellipsoids. The absolute value of q remained constant within each block of trials, but it was varied across blocks from possible values of ± 0.065, ± 0.050, ± 0.035, or ± 0.020.

For the two-frame slanted condition, the displays were identical to those described above except that the apparent motion sequences consisted of only two distinct frames presented in continuous alternation with an SOA of 200 msec. The ellipsoid depicted in the first frame of every sequence was always presented at an angle of 35°
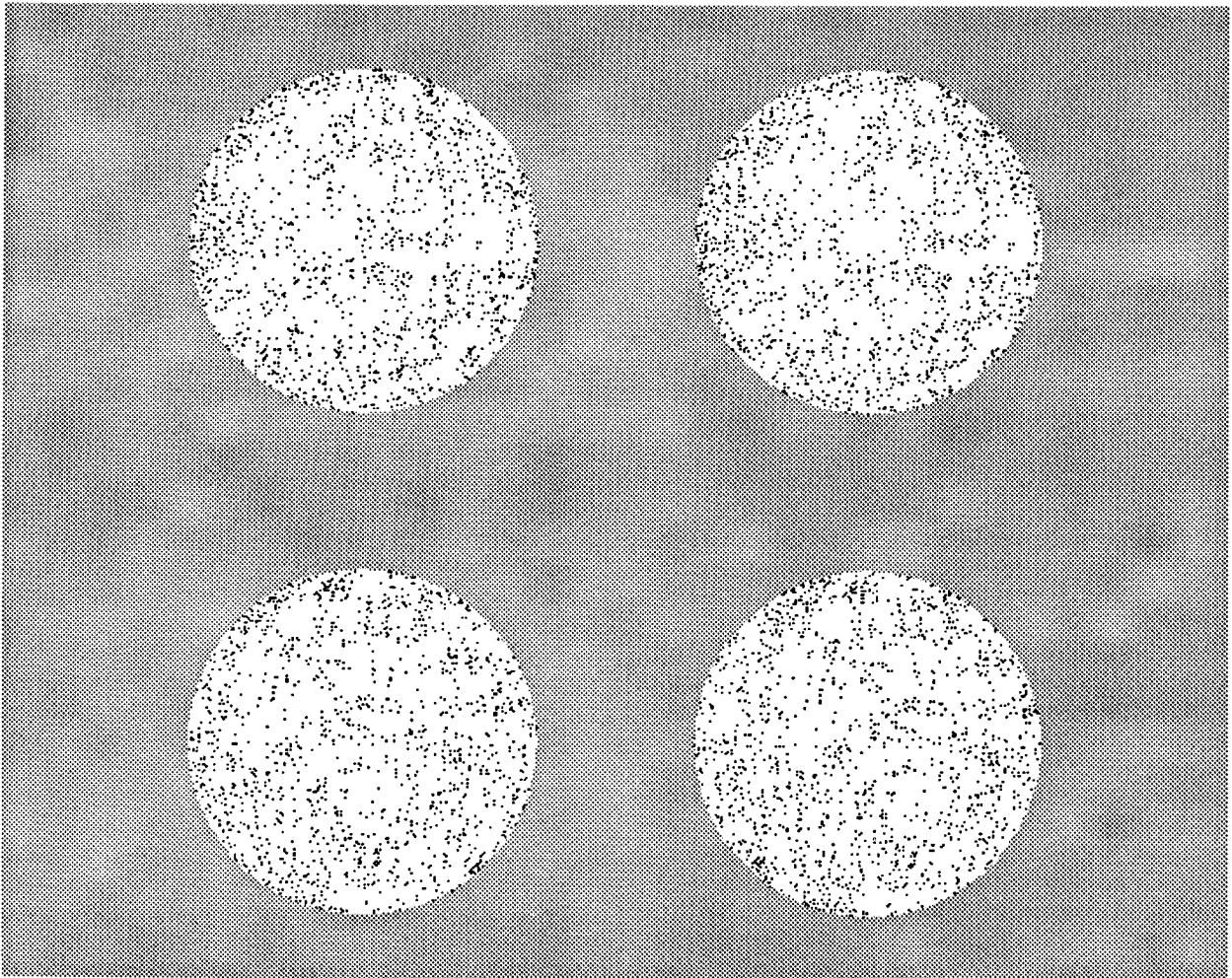

Figure 7. Two pairs of dot patterns similar to those used in the two-frame slanted condition of Experiment 4. The upper pair of images depicts a 0.065 compressed ellipsoid, while the lower pair of images depicts a 0.065 expanded ellipsoid. When presented in apparent motion, these displays can produce a compelling impression of a smoothly curved surface. A similar effect can also be obtained by viewing the images stereoscopically.

from its degenerate orientation (i.e., the average of the most extreme orientations for the eight-frame displays).

For the two-frame straight condition, in contrast, the displays were the same as those employed in the two-frame slanted condition, except that each depicted ellipsoid was presented at its degenerate orientation for which the possible variations of surface structure were affine equivalent along the line of sight. It quickly became apparent during pilot studies for this experiment that these two-frame straight displays were much more difficult to discriminate than those employed in the other conditions. Thus, in an effort to obtain a more informative measure of the limits of the observers' performance, the range of variation between the expanded and compressed ellipsoids was increased for this condition to include possible q values of ±0.20, ±0.15, ±0.10, or ±0.05.

To eliminate any possible confounding sources of information from the changing shape of an ellipsoid's occlusion boundary (see Andersen & Cortese, 1989; Loomis & Eby, 1989; Todd, 1985), all of the displays were presented behind a simulated opaque blue occluding surface through a circular aperture with a radius of 112 pixels. The size of this aperture was just small enough to occlude the boundaries of the depicted ellipsoid surfaces in all possible conditions. The resulting experimental displays appeared in each frame of an apparent motion sequence as a random configuration of white dots on a black circle within a homogeneous blue background. Figure 7 shows two pairs of images depicting a 0.065 compressed ellipsoid and a 0.065 expanded ellipsoid similar to those used in the two-frame slanted condition. When either of these pairs is presented in continuous alternation, it produces a compelling impression of a smoothly curved surface rotating in depth behind a circular aperture. A similar effect can also be obtained by viewing the surfaces stereoscopically.

The displays were evaluated by 3 highly practiced observers, including both authors and 1 naive subject who was unaware of the theoretical issues being investigated or any specific details of how the displays were generated. Each observer participated in six experimental sessions, composed of four 60-trial blocks of a single experimental condition. The difficulty of the required discriminations was varied across blocks in order of increasing difficulty. Thus, for the two-frame straight condition, a session would begin with a block of 60 trials with possible q values of ±0.20, and it would end with a block in which the q values were reduced to ±0.05. Similarly, for the two-frame slanted and eight-frame conditions, the sessions would begin with possible q values of ±0.065 and would end with those values reduced to ±0.02. The order of the six experimental sessions was determined at random for each observer.

Results and Discussion

Figure 8 shows the combined percentage of "expanded" responses for all 3 observers in the two-frame straight condition as a function of each surface's percentage distortion from the standard sphere. The data are plotted separately for both possible values of angular displacement. It is important to keep in mind that to achieve perfect performance on this task, an observer would have to respond "expanded" for all positive differences from the standard and respond "compressed" for all negative differences. Note in the figure that for most of the displays the observers' performance was barely above chance. Although they were highly practiced and received immediate feedback after every trial, they could not attain a 75% threshold of accuracy even when the structural variations to be detected were as large as $\pm20\%$ (see also Norman, 1990). Indeed, the structural variations of the depicted
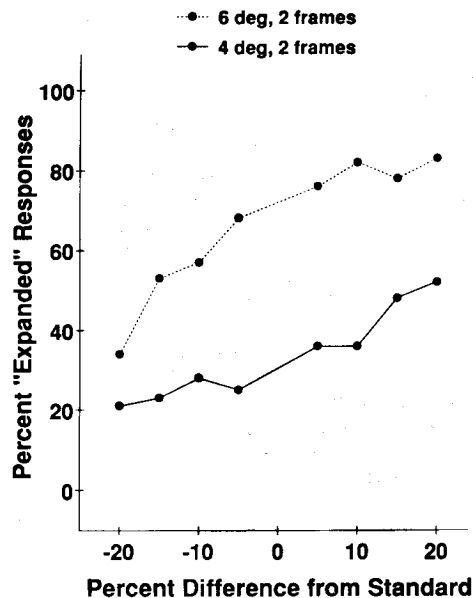

Figure 8. The percentage of "expanded" responses for 3 practiced observers in the two-frame straight condition of Experiment 4, as a function of the percentage difference between the depicted eccentricity of a test stimulus and a standard sphere. The 4° and 6° angular displacements are represented by solid and dotted lines, respectively.

ellipsoids accounted for only 28% of the total sum of squares, whereas the variations in angular displacement accounted for over 46%. These results should not be surprising, in light of the fact that the only possible variations of surface structure in the two-frame straight condition were all affine equivalent along the line of sight, and that the optical effects produced by these variations were mathematically indistinguishable from those produced by comparable variations in angular displacement.

The results obtained from the two-frame slanted and eight-frame conditions are presented together in Figure 9. Several important aspects of these data deserve to be highlighted. First, note that there were no significant effects of sequence length when the extreme orientations of the ellipsoids depicted in the two-frame and eight-frame displays were equated. Second, the overall level of accuracy for these displays improved by almost an order of magnitude relative to the two-frame straight condition. A structural variation of only $\pm3.5\%$ for objects viewed at a slant was more perceptually salient than a structural variation of $\pm20\%$ for objects viewed at a degenerate orientation. Finally, in the two-frame slanted and eight-frame conditions, the observers had no difficulty in discriminating the effects of surface structure from angular displacement as they did in the two-frame straight condition. The manipulations of surface structure accounted for over 86% of the total sum of squares, whereas the much larger manipulations of angular displacement accounted for less than 3%. Since the largest differences between the patterns of velocity in these displays were produced by the manipula-

··o·· 6 deg, 8 frames
—o— 4 deg, 8 frames
··•·· 6 deg, 2 frames
—•— 4 deg, 2 frames

Percent "Expanded" Responses

100

80

60

40

20

0

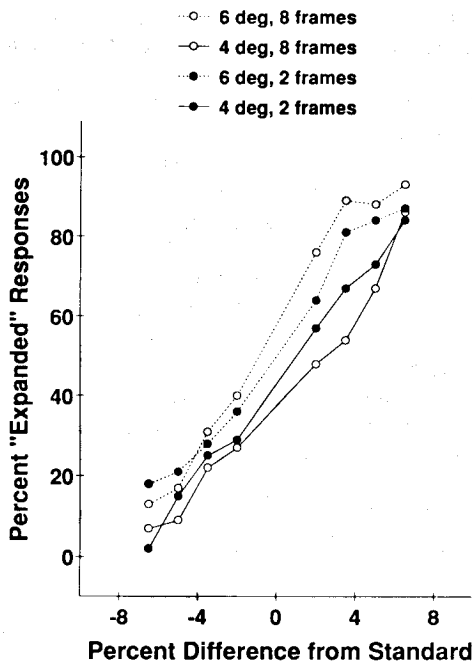-8   -4   0   4   8

Percent Difference from Standard

Figure 9. The percentage of "expanded" responses for 3 practiced observers in the two-frame slanted and eight-frame conditions of Experiment 4, as a function of the percentage difference between the depicted eccentricity of a test stimulus and a standard sphere. The two-frame slanted and eight-frame conditions are represented by filled and open circles, while the 4° and 6° angular displacements are represented by solid and dotted lines.

tions of angular displacement, this last finding shows clearly that the observers' judgments could not have been based on a simple direct comparison of the two-dimensional image velocities.

When considered as a whole, the results of this experiment provide strong support for the theoretical analysis of affine structure from motion proposed by Todd and Bressan (1990). According to this theory, the relevant information for perceiving structure from motion is restricted to first-order measures of velocity (or displacement) that are available within two-frame apparent motion sequences, and, as a consequence of this restriction, observers should be incapable of discriminating variations of surface structure that are affine equivalent along the line of sight. It is important to recognize that the perceptual ambiguities predicted by this analysis are directionally anisotropic, in such a way that any pair of distinct objects should always be discriminable except when viewed from a degenerate orientation. The present experiment was designed specifically to test this prediction, and the results provided a clear confirmation of its psychological validity.

It is also interesting in this regard that the significant effects of object orientation could easily have been mistaken for an effect of sequence length if the two-frame slanted condition had not been included in the experimental design (cf. Loomis & Eby, 1988; Norman, 1990). As

the objects depicted in the eight-frame displays rotated farther and farther from their degenerate orientations, the variations in their structure must have gradually become more and more discriminable over time, relative to the displays in the two-frame straight condition. However, in order to demonstrate a true effect of sequence length, it is necessary to show that a multiframe display is capable of producing more accurate performance than would be possible for any single pair of images presented in isolation. Although the eight-frame condition of the present experiment produced much higher performance than did the two-frame displays of objects depicted at a degenerate orientation, this difference was eliminated in the two-frame slanted condition by controlling which pairs of images from the longer length sequences would be used to generate the displays.

GENERAL DISCUSSION

Our primary conclusion from the present experiments and the related work of Todd and Bressan (1990) is that when all other relevant variables are equated, the length of an apparent motion sequence has little or no effect on the abilities of human observers to perceive structure from motion. Although some reports in the literature may appear to conflict with this conclusion, we believe that most of these discrepancies can be attributed to other confounding variables that have been allowed to covary with sequence length. One such variable, whose effects were demonstrated in Experiment 4, is the specific orientation at which an object is depicted within a two-frame apparent motion sequence. If it is viewed at a degenerate orientation from which all possible structural variations are affine equivalent along the line of sight, the accuracy of observers' judgments may be severely impaired.

Another important variable to consider in this context is stimulus duration. The perceptual analysis of two-frame apparent motion sequences need not occur instantaneously. If observers require some minimal amount of time to sample and process the available information, performance might be expected to suffer when a display is terminated after too brief an interval. There is some evidence to suggest that temporal duration can have a large effect on the accuracy of observers' judgments about three-dimensional form. When a two-frame sequence is immediately terminated (Hildreth et al., 1990) or followed by a blank interval (Loomis & Eby, 1988) after a single presentation, performance is significantly reduced relative to longer length displays. This should not be confused, however, with a true effect of sequence length. Indeed, there is growing evidence—including the results of the present experiments and the earlier studies of Braunstein et al. (1990), Braunstein et al. (1987), Todd et al. (1988), and Todd and Bressan (1990)—that the effects of sequence length can be effectively eliminated by presenting each display in continuous alternation so that observers can sample the available information for as long as is necessary for them to arrive at a maximally accurate judgment.

Still another potentially confounding variable can arise when one investigates the temporal buildup of perceived structure from motion by limiting the lifetimes of individual feature points (Dosher et al., 1990; Husain et al., 1989; Norman, 1990; Todd, 1985; Treue et al., 1991). To manipulate element lifetimes within an extended apparent motion sequence, some proportion of a depicted object's visible features can be randomly repositioned at each frame transition. As this proportion is gradually increased, the amount of perceived depth in a display will eventually become attenuated. One possible interpretation of this result is that the visual perception of structure from motion can be facilitated by allowing observers to track individual features over multiple frames. The problem with this conclusion is that as element lifetimes are decreased, there is a corresponding increase in the relative proportion of uncorrelated noise. Our own observations have indicated that when the level of noise is equated for all displays, there is no noticeable perceptual effect of any increase beyond two in the number of distinct frames for which each element is allowed to survive. A similar finding has also been reported by Dosher et al. (1990).

Why would observers fail to exploit the available information about euclidean metric structure within multiple-frame apparent motion sequences? One likely explanation for this theoretically surprising result is that human observers have poor sensitivity to higher order temporal derivatives of an object's projected motion. Previous research has shown, for example, that observers perform poorly on tasks such as the discrimination of scalar acceleration, in which the relative displacements of a moving point must be compared over multiple time intervals (e.g., see Gottsdanker, 1952, 1955; Gottsdanker, Frick, & Lockard, 1961; Todd, 1981).

If it is indeed true that the perceptual analysis of structure from motion is restricted to the use of first-order temporal derivatives (or displacements), as we have argued above, then it ought to be possible to confirm this limitation through a variety of converging operations. The mathematical analysis of Todd and Bressan (1990) provides some highly specific predictions about which tasks are theoretically possible from two-frame apparent motion sequences, and which ones are not. In the discussion that follows, we will examine these predictions in considerable detail for several generic tasks that have previously been employed to measure observers' perceptions of three-dimensional form.

Discriminations of Length or Angle

The primary limitation of restricting an analysis of structure from motion to the available information within two-frame apparent motion sequences is that there is no general solution for determining the relative lengths or angles between line segments oriented in different directions. Todd and Bressan (1990) have recently demonstrated that a similar limitation may also exist for human perception. Observers in their experiments were asked to discriminate the relative three-dimensional lengths or

angles between moving line segments, whose relative orientations were carefully controlled so that performance above chance could not be achieved solely on the basis of the projected lengths or projected angles depicted in each display. The results on both tasks were perfectly consistent with what would be expected on the basis of the available information within two-frame apparent motion sequences. That is to say, the overall level of performance at discriminating variations of euclidean metric structure was quite poor, and there were no appreciable improvements as the number of frames in the apparent motion sequences was increased from two to eight.

There are some exceptions to the general limitations described above. Indeed, there are at least two specialized circumstances in which it is theoretically possible to determine an object's euclidean metric structure from a two-frame apparent motion sequence. One interesting exception, which has been analyzed by Hoffman and Flinchbaugh (1982), occurs for objects viewed under orthographic projection whose motion is confined to a fixed plane (see also Lappin, 1990). A similar exception has also been demonstrated by Longuet-Higgins (1981) for arbitrary configurations under strong polar perspective. With respect to the psychological validity of these analyses, a recent experiment by Lappin and Love (in press) has provided some evidence that a planarity constraint can be exploited by actual human observers, but the effects of perspective on the visual perception of euclidean metric structure have not been systematically investigated.

One type of length discrimination that is always theoretically possible from two-frame apparent motion sequences includes judgments of relative depth. Note in Equation 1 that when an object rotates about an axis in the picture plane, the projected displacement of each point is linearly related to its position in depth relative to the axis of rotation. Similar information is also available for objects rotating about a slanted axis, though the analysis in that case is more complex, requiring the removal of the component of rotation about the line of sight (see Todd & Bressan, 1990; Ullman, 1977). There is little evidence to suggest the extent to which human observers are able to make use of this information. Hildreth et al. (1990) have recently demonstrated that observers are sensitive to ordinal depth relations among elements rotating about a slanted axis under orthographic projection, but no one has yet examined observer sensitivity to relative depth intervals in this context (cf. Lappin & Fuqua, 1983).

Discriminations of Objects or Surfaces

A particularly common procedure for studying the visual perception of structure from motion is to measure an observer's ability to discriminate objects or surfaces (e.g., see Braunstein & Andersen, 1984; Braunstein et al., 1987; Dosher et al., 1990; Husain et al., 1989; Loomis & Eby, 1988; Sperling et al., 1989; Todd, 1984, 1985; Treue et al., 1991). As a general rule, any object-discrimination task should be theoretically possible with two-frame apparent motion sequences unless two specific conditions

are satisfied: (1) the variations of object structure to be discriminated are all affine equivalent along the line of sight; and (2) they are presented in apparent motion with varying amounts of angular displacement.

Experiment 4 of the present series was designed to examine the psychological validity of this theoretical limitation for actual human perception. The results confirmed that structural variations that are affine equivalent along the line of sight cannot be discriminated from comparable variations in the amount of angular displacement (or velocity) at each frame transition. It is important to keep in mind, however, that both of these conditions must be satisfied simultaneously for performance to be impaired. For example, if all of the depicted objects to be discriminated are rotated about an axis in the picture plane with the same frame-to-frame angular displacement, as in Todd (1984), then observers' discriminations can be quite accurate. Similarly, if the objects to be discriminated are not affine equivalent along the line of sight, then the effects of angular displacement will be greatly diminished.

Detection of Planar Surfaces

From the available information within two-frame apparent motion sequences, it is theoretically possible to detect any structural property of an object that remains invariant under arbitrary affine transformations. One such property is surface flatness (i.e., if a planar surface patch is stretched in any direction, its flatness remains unchanged). In order to detect whether a configuration of moving elements under orthographic projection is coplanar, it is useful first to remove the component of rotation about the line of sight to produce a pattern of parallel image trajectories. Within this transformed pattern, any subset of points must be coplanar if there exists a set of constant coefficients a, b, c, and d, such that for any point i:

$$aX_i + bY_i + cD_i + d = 0, \qquad (10)$$

where X_i and Y_i are its instantaneous position coordinates in the image plane, and D_i is its transformed projected displacement.

The available evidence indicates that human observers are remarkably sensitive to small deviations from a perfectly flat surface (see Cornilleau-Pérès & Droulez, 1989; Norman, 1990; Todd & Bressan, 1990). Moreover, in the studies that have manipulated sequence length, there have been no appreciable improvements in performance as the number of distinct frames in an apparent motion sequence is increased beyond two. The latter finding provides strong evidence that the perceptually relevant information for discriminating planar from nonplanar surfaces is fully available within the first-order field of projected displacements, and that observers' performance on these tasks cannot therefore be based on an analysis of euclidean metric structure.

Magnitude Estimations

Another popular task for studying the perception of structure from motion involves magnitude estimations of

depth, slant, or curvature (e.g., see Braunstein & Andersen, 1984; Loomis & Eby, 1988, 1989; Tittle et al., 1990; Todd, 1984, 1985). Experiments 1 and 2 of the present series are typical examples of this paradigm. Observers in these studies made numerical ratings of the amplitudes of sinusoidal surfaces relative to their periods. The results indicated that observers can accurately discriminate variations of surface amplitude, that performance is only minimally affected by sequence length, and that the perceived amplitudes are systematically overestimated. The last of these three findings poses the most difficulty for an analysis of affine structure from motion. Because two-frame apparent motion sequences are inherently ambiguous up to an affine stretching transformation along the line of sight, they cannot uniquely specify a particular value of absolute depth, slant, or curvature. Although this is consistent with the overall inaccuracy of the observers' judgments, it cannot account for their reliability. Why should the surfaces be systematically overestimated? Indeed, why should they appear reliably to have any specific depth at all?

The available evidence from the present series of experiments and other previous investigations does not provide a definitive answer to these questions. There is evidence to suggest, however, that the magnitude of perceived surface amplitude in these displays could not have been based on a simple linear mapping between the depth of each point relative to the rotation axis and its projected displacement in the image plane. If perceived relative depth were determined directly as a simple linear mapping of projected displacement, it should also have varied proportionally with the amount of rotation at each frame transition of an apparent motion sequence. The available evidence does not confirm this prediction. Although there are significant effects of rotary displacement on perceived relative depth, they are many times smaller than what would be expected on the basis of a simple linear mapping hypothesis. One likely explanation of this finding is that the perceptual analysis of structure from motion incorporates a form of automatic gain control to provide some degree of constancy over possible variations in the angular velocities of moving objects. The precise mechanisms by which this is achieved have yet to be determined, although some plausible hypotheses have been proposed by Loomis and Eby (1988) and by Tittle et al. (1990).

Detection of Rigidity

Still another commonly used task for studying the perception of structure from motion involves judgments of rigidity or coherence (e.g., see Braunstein, 1962; Braunstein et al., 1990; Cutting, 1987; Doner et al., 1984; Dosher et al., 1989; Green, 1961; Lappin et al., 1980; Petersik, 1979, 1980, 1987; Todd, 1982; Todd et al., 1988; Todd & Bressan, 1990). The theoretical possibility of detecting rigidity from two-frame apparent motion sequences under orthographic projection was first noted by Ullman (1977). The method that he proposed for accomplishing this task involves mathematically removing the component of an object's rotation about the observer's

line of sight. If this results in a pattern of parallel image trajectories, there is an overwhelming probability that the depicted motion is rigid. If it does not, the motion must be nonrigid (see also Bennett et al., 1989; Koenderink & van Doorn, in press).

Most previous experiments on the discrimination of rigid from nonrigid motion by actual human observers have employed displays that are theoretically discriminable by using Ullman's test, and there is growing evidence that the relevant information for the perception of rigidity in these displays is fully available within two-frame apparent motion sequences. As with other types of discrimination tasks involving the perception of structure from motion, the available evidence indicates that observers are surprisingly poor at integrating information over multiple views to perceptually discriminate rigid from nonrigid motion. Recent studies by Braunstein et al. (1990) and Todd and Bressan (1990) have shown that increasing the number of distinct frames in an apparent motion sequence beyond two can have little or no effect on performance when the amount of nonrigid distortion to be detected is equated over different sequence lengths.

Some investigators have also examined the perception of nonrigid motions that are theoretically undetectable by using Ullman's test. For example, in one such study by Norman and Todd (1991), observers were presented with random configurations of moving lines that were rotated in depth about a fixed axis in the picture plane. For some of the trials, the depicted object was rotated rigidly with a variable angular displacement at each frame transition. For other trials, the depicted object was rotated with a fixed angular displacement, but it was stretched along the line of sight at each frame transition. Observers' rigidity ratings revealed that these two types of motion were perceptually indistinguishable, even when the objects were presented in continuous rotation with apparent motion sequences composed of several hundred distinct frames. Because the large depicted distortions in the stretching displays could easily have been detected by using an analysis of euclidean structure from motion, this result provides especially strong evidence that human observers may be restricted to a more abstract analysis of affine structure.

Another degenerate form of nonrigid motion that cannot be detected by using Ullman's test has been investigated by Todd (1982). Observers in this study were again presented with random configurations of connected line segments rotating in depth about a fixed axis in the picture plane. Nonrigid distortions were produced in some of the displays by having each vertex rotate at a different angular velocity. Although any two discrete views from the resulting patterns of projected motion had possible rigid interpretations, observers had no difficulty identifying the motion as nonrigid if they were allowed to view a sufficient proportion of its overall trajectory. It is clear in this case that observers must be able to integrate information over more than two frames of an apparent motion sequence. It should also be noted, however, that accurate

performance on this task does not require a precise measure of second-order derivates of position with respect to time. Observers need only note that the oscillatory frequency of each point is different in order to correctly label the motion as nonrigid. This finding suggests strongly that observers may be sensitive to some types of higher order spatiotemporal relations but not others. Developing a more precise inventory of which properties can be perceptually integrated over time and which ones cannot is an interesting and important question that will have to await future research.

REFERENCES

ALOIMONOS, J., & BROWN, C. M. (1989). On the kinetic depth effect. *Biological Cybernetics*, **60**, 445-455.

ANDERSEN, G. J., & CORTESE, J. M. (1989). 2-D contour perception resulting from kinetic occlusion. *Perception & Psychophysics*, **46**, 49-55.

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. (1962). Depth perception in rotating dot patterns: Effects of numerosity and perspective. *Journal of Experimental Psychology*, **64**, 415-420.

BRAUNSTEIN, M. L., & ANDERSEN, G. J. (1984). Shape and depth perception from parallel projections of three-dimensional motion. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 749-760.

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.

BRAUNSTEIN, M. L., & TODD, J. T. (1990). On the distinction between artifacts and information. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 211-216.

CORNILLEAU-PÉRÈS, V., & DROULEZ, J. (1989). Visual perception of surface curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics*, **46**, 351-364.

CUTTING, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 323-334.

DONER, J., LAPPIN, J. S., & PERFETTO, G. (1984). Detection of three-dimensional structure in moving optical patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 1-11.

DOSHER, B. A., LANDY, M. S., & SPERLING, G. (1989). Ratings of kinetic depth in multidot displays. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 816-825.

DOSHER, B. A., LANDY, M. S., & SPERLING, G. (1990). Kinetic depth effect and optic flow: I. 3D shape from Fourier motion. *Vision Research*, **29**, 1789-1814.

GOTTSDANKER, R. M. (1952). The accuracy of prediction-motion. *Journal of Experimental Psychology*, **43**, 26-36.

GOTTSDANKER, R. M. (1955). A further study of prediction-motion. *American Journal of Psychology*, **68**, 432-437.

GOTTSDANKER, R. M., FRICK, J. W., & LOCKARD, R. B. (1961). Identifying the acceleration of visual targets. *British Journal of Psychology*, **52**, 31-42.

GREEN, B. F., JR. (1961). Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology*, **62**, 272-282.

HILDRETH, E. C., GRZYWACZ, N. M., ADELSON, E. H., & INADA, V. K. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception & Psychophysics*, **48**, 19-36.

HOFFMAN, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America A*, **72**, 880-892.

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

HUANG, T., & LEE, C. (1989). Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **11**, 536-540.

HUSAIN, M., TREUE, S., & ANDERSEN, R. A. (1989). Surface interpolation in three-dimensional structure-from-motion perception. *Neural Computation*, **1**, 324-333.

KOENDERINK, J. J., & VAN DOORN, A. J. (1975). Invariant properties of the motion parallax field due to the motion of rigid bodies relative to the observer. *Optica 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 flow. In G. Hauske & F. Butenandt (Eds.), *Kybernetik* (pp. 224-247). Munich: Oldenberg.

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

LAPPIN, J. S. (1990). Perceiving metric structure of environmental objects from motion, self-motion and stereopsis. In R. Warren & A. H. Wertheim (Eds.), *The perception and control of self-motion* (pp. 541-576). Hillsdale, NJ: Erlbaum.

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.

LAPPIN, J. S., & FUQUA, M. A. (1983). Accurate visual measurement of three-dimensional moving patterns. *Science*, **221**, 480-482.

LAPPIN, J. S., & LOVE, S. R. (in press). Metric structure of stereoscopic form from congruence under motion. *Perception & Psychophysics*.

LONGUET-HIGGINS, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, **293**, 133-135.

LONGUET-HIGGINS, H. C., & PRAZDNY, K. (1984). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London: Series B*, **208**, 385-397.

LOOMIS, J. M., & EBY, D. W. (1988). Perceiving structure from motion: Failure of shape constancy. In *Proceedings from the Second International Conference on Computer Vision* (pp. 383-391). Washington, DC: IEEE.

LOOMIS, J. M., & EBY, D. W. (1989). Relative motion parallax and the perception of structure from motion. In *Proceedings from the Workshop on Visual Motion* (pp. 204-211). Washington, DC: IEEE.

NORMAN, J. F. (1990). *The perception of curved surfaces defined by optical motion.* Unpublished doctoral dissertation, Vanderbilt University.

NORMAN, J. F., & TODD, J. T. (1991). The perception of rigid motion for affine distortions in depth. *Investigative Ophthalmology & Visual Science*, **32**, 958.

PETERSIK, J. T. (1979). Three-dimensional object constancy: Coherence of a simulated rotating sphere in noise. *Perception & Psychophysics*, **25**, 328-335.

PETERSIK, J. T. (1980). The effects of spatial and temporal factors on the perception of stroboscopic rotation simulations. *Perception*, **9**, 271-283.

PETERSIK, J. T. (1987). Recovery of structure from motion: Implications for a performance theory based on the structure-from-motion theorem. *Perception & Psychophysics*, **42**, 355-364.

REICHEL, F. D., & TODD, J. T. (1990). Perceived depth inversion of smoothly curved surfaces due to image orientation. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 653-664.

SPERLING, G., LANDY, M. S., DOSHER, B. A., & PERKINS, M. E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 826-840.

TITTLE, J. S., BRAUNSTEIN, M. L., & LITER, J. C. (1990). Recovering surface slant from orthographic projections. *Investigative Ophthalmology & Visual Science*, **31**, 523.

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

TODD, J. T. (1982). Visual information about rigid & nonrigid motion: A geometric analysis. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 238-252.

TODD, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, **36**, 97-103.

TODD, J. T. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 689-710.

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.

TREUE, S., HUSAIN, M., & ANDERSEN, R. A. (1991). Human perception of structure from motion. *Vision Research*, **31**, 59-76.

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.

ULLMAN, S. (1983). Recent computational studies in the interpretation of structure from motion. In J. Beck & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 459-480). New York: Academic Press.

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

WAXMAN, A., & ULLMAN, S. (1985). Surface structure and three-dimensional motion parameters from image flow kinematics. *International Journal of Robotics Research*, **4**, 79-94.