The perception of 3-dimensional affine structure from minimal apparent motion sequences

JAMES T. TODD
Brandeis University, Waltham, Massachusetts
and Boston University, Boston, Massachusetts

and

PAOLA BRESSAN
Università di Padova, Padova, Italy

The research described in the present article was designed to identify the minimal conditions for the visual perception of 3-dimensional structure from motion by comparing the theoretical limitations of ideal observers with the perceptual performance of actual human subjects on a variety of psychophysical tasks. The research began with a mathematical analysis, which showed that 2-frame apparent motion sequences are theoretically sufficient to distinguish between rigid and nonrigid motion and to identify structural properties of an object that remain invariant under affine transformations, but that 3 or more distinct frames are theoretically necessary to adequately specify properties of euclidean structure such as the relative 3-dimensional lengths or angles between nonparallel line segments. A series of four experiments was then performed to verify the psychological validity of this analysis. The results demonstrated that the determination of structure from motion in actual human observers may be restricted to the use of first order temporal relations, which are available within 2-frame apparent motion sequences. That is to say, the accuracy of observers' judgments did not improve in any of these experiments as the number of distinct frames in an apparent motion sequence was increased from 2 to 8, and performance on tasks involving affine structure was of an order of magnitude greater than performance on similar tasks involving euclidean structure.

When a visible object moves relative to an observer (or vice versa), there are continuous deformations of optical stimulation that can often provide perceptually salient information about an object's 3-dimensional structure. This can be demonstrated most clearly with displays that contain no other sources of static pictorial information. Suppose, for example, that an observer views a video display of an unbounded configuration of randomly positioned dots or lines. If such a configuration is presented at a stationary orientation, observers inevitably report that the display appears as a 2-dimensional pattern of random noise. If the same configuration begins to rotate in depth, however, its 3-dimensional structure becomes immediately apparent—even to naive observers (e.g., see Wallach & O'Connell, 1953).

During the past decade, numerous attempts have been made to explain this phenomenon, by developing formal models of how an object's structure can be computationally determined from its pattern of projected motion. Although there is considerable variation in the specific details of these models, several generic characteristics are shared by many of them. Most analyses assume, for example, that the observed object is composed of identifiable discrete points, and that its motion is constrained to rigid rotations in depth. They also share a common goal of computing the 3-dimensional distance between each pair of points, solely on the basis of the measured positions of those points within a sequence of 2-dimensional orthographic projections. Almost all of these analyses have been focused on the theoretical limitations of an "ideal observer," who can identify the corresponding points in successive views, and who can measure the projected position of each point and perform necessary mathematical operations with perfect accuracy. Within this context, it can be proven mathematically that three distinct views of four noncoplanar points are both necessary and sufficient for the obtaining of a unique interpretation of an object's 3-dimensional structure (see Bennett, Hoffman, Nicola, & Prakash, 1989; Huang & Lee, 1989; Ullman, 1979, 1983).

It is tempting to conclude from these mathematical analyses that the perception of structure from motion should

be theoretically impossible for displays that depict fewer than four points or fewer than three views. Such a conclusion would be highly misleading, however. It is important to keep in mind that these theoretical limitations of ideal observers are specifically concerned with computing the metric distance between each pair of points in 3-dimensional space. It does not follow that other, more abstract properties of object structure cannot be optically specified with smaller amounts of information. Indeed, there is a growing amount of psychophysical evidence that 2-frame apparent motion sequences of objects rotating under orthographic projection provide sufficient information for observers to obtain compelling kinetic depth effects, and to accurately discriminate between different 3-dimensional structures (e.g., see Braunstein, Hoffman, Shapiro, Andersen, & Bennett, 1987; Todd, Akerstrom, Reichel, & Hayes, 1988).

In light of these observations, it is useful to reexamine the computational analysis of structure from motion in an effort to obtain a more detailed understanding of what types of tasks are theoretically possible from 2-frame apparent motion sequences. Let us begin by adopting a moving Cartesian coordinate system (x, y, z), where the origin is attached to some visible point P_o, the z-axis (i.e., the line of sight) is defined by a line connecting P_o to the point of observation, and the xy plane (i.e., the picture plane) is perpendicular to the z-axis. In order to remain consistent with previous analyses, we shall assume that an ideal observer is able to measure the position of each point within an object's orthographic projection in the picture plane. That is to say, we assume that the x- and y-coordinates of each point are given at each moment in time, but that their z-coordinates are unknown. The problem is to determine which aspects of an object's 3-dimensional structure, if any, can be reliably computed from these measures within a 2-frame apparent motion sequence.

When defined relative to some visible feature point P_o, any rigid displacement of an object can be decomposed into two independent components: a rotation ϵ about an axis in the picture plane through P_o, followed by a rotation θ about the line of sight (see Figure 1). As was first noted by Ullman (1977, 1983), this particular method of decomposition is especially useful for the computational analysis of structure from motion, because it is possible to isolate the effects of each component within the overall pattern of optical flow.

Consider, for example, the instantaneous configurations O_1 and O_2 of a moving set of points at two different moments in time, and their orthographic projections O_1' and O_2'. Let us assume initially that $\theta = 0$ (i.e., the configurations are related by a pure rotation about an axis in the picture plane). In that special case, the image trajectories between each pair of corresponding points in O_1' and O_2' will all be parallel to one another in a direction that is perpendicular to the axis of rotation (see Figure 2). This parallelism will be destroyed in the more general case of
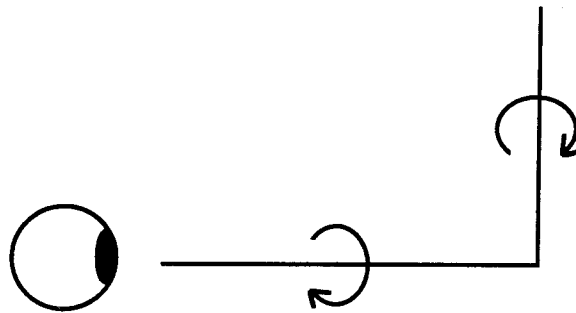

Figure 1. Any rotation of an object in 3-dimensional space can be decomposed into two components: a rotation about an axis in the picture plane, followed by a rotation about the line of sight.

rigid body motion where $\theta \neq 0$, but it can always be restored by rotating O_2' in the picture plane about the origin P_o, effectively transforming the overall pattern of motion back to the special case where $\theta = 0$. (A specific algorithm for computing the required amount of rotation solely on the basis of image data is provided in the appendix.) Although it is mathematically possible to construct nonrigid configurations that will project to a pattern of parallel trajectories following image rotation, the probability of ever encountering such a configuration in natural vision is negligible. Thus, since the probability of false targets is vanishingly small, the decomposition of 2-frame sequences can provide a reliable test to determine whether an object is undergoing rigid motion (see also the related analyses of Bennett et al., 1989; and Koenderink & van Doorn, in press).

Let us now consider how the pattern of projected motion can be used to determine an object's 3-dimensional structure. For the purposes of this analysis we shall assume that two views of an object occur close together in time, so that each element's displacement provides a reasonable approximation of its instantaneous velocity; and that one of the views has been rotated appropriately (as described above), so that all of the projected element trajectories are parallel to one another. If we align the x-axis in the same orientation as these transformed element trajectories, then our problem will be reduced to analyzing the special case of rotation in depth about the y-axis (see Braunstein & Todd, 1990). Within this context, the instantaneous position of any point (i) at each moment in time (t) can be defined by the following set of parametric equations:

$$x_i = a_i \sin(2\pi t/p + \phi_i) \tag{1}$$

$$y_i = k_i \tag{2}$$

$$z_i = a_i \cos(2\pi t/p + \phi_i) \tag{3}$$

where x_i, y_i, and z_i are its instantaneous position coordinates, a_i is its radial distance from the axis of rotation, k_i is its vertical distance from the origin, ϕ_i is its initial phase when $t = 0$, and p is the period of rotation. If we

differentiate Equation 1 with respect to time and substitute Equation 3, the following relation is obtained:

$$z_i = (dx_i/dt)(p/2\pi) \qquad (4)$$

This relation demonstrates that the instantaneous depth of any point relative to the rotation axis is optically specified by its transformed velocity in the image plane times a constant, whose value is determined by the period of rotation. Because this constant term is not specified instantaneously, Equation 4 cannot provide a unique interpretation of an object's metric structure. This is consistent with previous findings of Bennett et al. (1989) and Huang and Lee (1989), who have demonstrated independently that a 2-frame apparent motion sequence under orthographic projection will always be compatible with an infinitely large, one-parameter family of rigid interpretations.[1]

This *does not* suggest, however, that a 2-frame apparent motion sequence is necessarily inadequate for making accurate judgments about an object's 3-dimensional structure. To better understand why this is so, it is useful to consider how the family of rigid interpretations compatible with Equation 4 are related to one another. It is important to note in particular that varying the parameter p for a given pattern of optical motion is mathematically equivalent to subjecting the depicted object to an affine "stretching" transformation along the z-axis (cf. Koenderink & van Doorn, in press). It follows, therefore, that any structural property of an object that remains invariant under affine transformations can be optically specified from 2-frame apparent motion sequences if the depicted displacements are sufficiently small to approximate a first-order temporal derivative. This is a surprisingly broad set of properties. On the basis of Equation 4, it is possible 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 detect structural differences between any pair of objects that cannot be made congruent by an affine stretching transformation in depth (cf. Ullman, 1983).

We shall refer to these properties hereafter as *affine structure*, but it is important not to be misled by this nomenclature. The inherent ambiguity of 2-frame apparent motion sequences is restricted to affine stretching transformations along the z-axis, which does not encompass the entire class of possible affine transformations. Thus, two objects that are affine equivalent in the more general sense could still be discriminated from 2-frame apparent motion sequences if they are related by a stretching transformation in any direction that is not parallel to the z-axis.

The theoretical analysis of affine structure from motion has important methodological implications. In previous investigations of the kinetic depth effect, several basic tasks have typically been employed. These include judg-
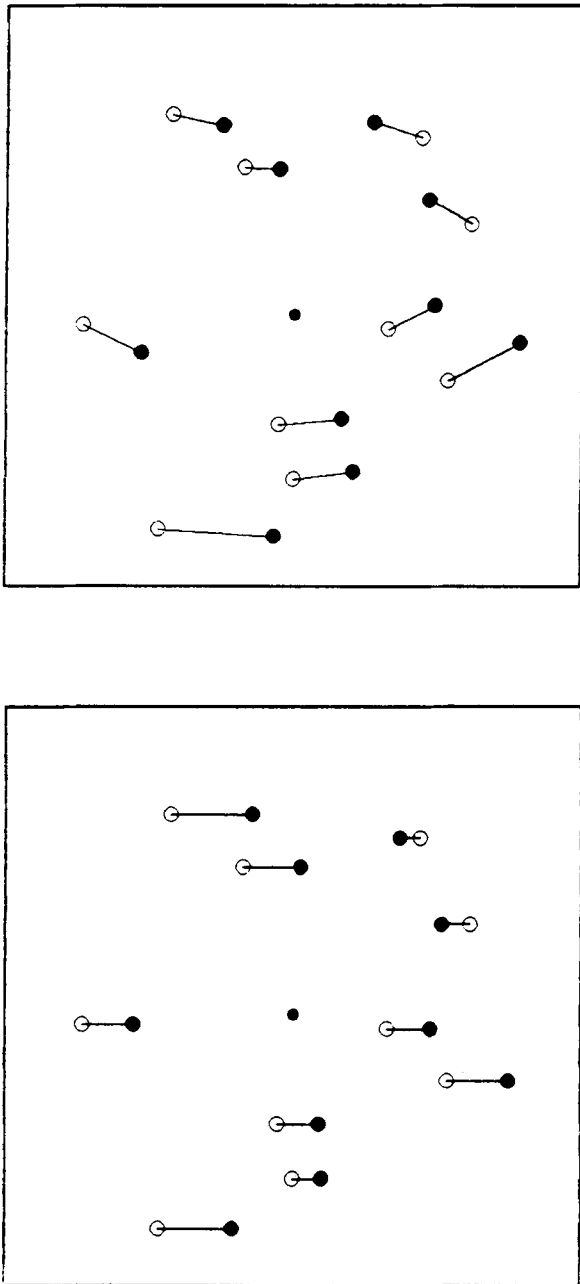

Figure 2. Two patterns of projected trajectories relative to a fixed feature point P_0 for a random configuration of elements rotating rigidly in 3-dimensional space. Each connected pair of open and filled circles represents a single element at different moments in time. The upper pattern of trajectories with varying orientations is typical of an object rotating about an axis that is slanted with respect to the image plane. The lower pattern of parallel trajectories was obtained by rotating the filled circles in the upper pattern about P_0. Barring the occurrence of improbable coincidences, the creation of parallel trajectories through image rotation can only be achieved in natural vision if the depicted configuration in 3-dimensional space is undergoing rigid motion—see Ullman, 1977.

ments 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, Hoffman, & Pollick, 1990; Cutting, 1987; Doner, Lappin, & Perfetto, 1984; Lappin, Doner, & Kottas, 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 3-dimensional forms (e.g., Braunstein, et al., 1987; Dosher, Landy, & Sperling, 1990; Sperling, Landy, Dosher, & Perkins, 1989)—all of which can be performed accurately by ideal observers on the basis of an analysis of affine structure from 2-frame apparent motion sequences. Other investigators have employed tasks involving psychometric ratings of surface depth, slant, or curvature (e.g., Braunstein & Andersen, 1984; Loomis & Eby, 1988, 1989; Todd, 1984, 1985), for which an affine analysis can produce reasonably accurate performance based on the overall range of transformed velocities in each display. Indeed, it would appear from this brief survey that none of these existing tasks requires a complete analysis of euclidean metric structure such as the one proposed by Ullman (1979), and that most if not all of the existing psychophysical data can be adequately explained using a more abstract analysis of affine structure.

In order to provide a clear empirical demonstration that human observers are capable of perceiving euclidean structure from motion, it would be necessary to design an appropriate psychophysical task for which an analysis of affine structure would be inherently inadequate. It is especially useful to note in this regard that an affine analysis of structure from motion can reveal certain metric properties of an object but not others. That is to say, it is capable of determining the metric length ratio between any pair of parallel line segments, but it is incapable of determining the relative lengths (or angles) between line segments oriented in different directions. With respect to the specific analysis presented above, this limitation arises from the unknown parameter p in Equation 4. By assigning an arbitrary nonzero value to p, it would be possible to obtain a perfectly accurate ratiometric scale of the relative distances between each point in depth, but there would be no way to relate this scale to measures of distance in the xy plane (i.e., the resulting distance measure would be anisotropic). Are human observers capable of detecting these euclidean properties of structure from motion? The present series of experiments began with an effort to address this question.

GENERAL METHOD

Apparatus

Experiments 1 and 3 were run on 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 1.2'. All responses were recorded by pressing keys on the computer keyboard.

Experiments 2 and 4 were run on an Iris 3130 graphics workstation. The stimuli were presented within a 34 × 26-cm rectangular region of the display screen with a spatial resolution of 1,024 × 768 pixels. A chinrest was used to maintain a constant viewing distance of 86 cm, such that each pixel spanned a visual angle of approximately 1.3'. All responses were recorded by pressing buttons on a handheld mouse.

Stimuli

The stimuli were designed to simulate random configurations of dots or lines rotating in depth under orthographic projection. Each display was composed of a sequence of discrete images that cycled back and forth in rapid succession in order to create the visual impression of motion. On the basis of the earlier findings of Todd et al. (1988), preliminary experiments were performed to determine the optimal timing parameters for each possible sequence length. Observers in these experiments compared the relative salience of perceived structure from motion for similar displays with different stimulus onset asynchronies (SOA), which could be varied in increments of $\frac{1}{60}$ sec as constrained by the 60-Hz raster refresh rate. For sequence lengths of 2, 3, 4, 6, and 8 distinct frames, the optimal SOAs were 200, 150, 117, 100, and 83 msec, respectively. These values were employed in all subsequent experiments, with an interstimulus interval (ISI) of 0 msec.

In a further attempt to optimize the displays, the patterns of projected motion were constrained so that the average displacement of each feature point would fall within a range between 3 and 10 pixels, and that each individual displacement would fall within a range between 1 and 12 pixels (cf. Braunstein et al., 1987). Due to the limited spatiotemporal resolution of raster displays, the projected displacement of an unconstrained feature point will be rounded to zero during certain phases of its rotary motion in depth. If this is allowed to occur for 2-frame apparent motion sequences, the depicted element will be perceived as stationary, and its relation to other moving elements will appear nonrigid. Similarly, we have also observed that when projected displacements become too large, the perception of rotation in depth is diminished, and the depicted pattern of motion can appear as a 2-dimensional nonrigid deformation within the picture plane (see Todd et al., 1988). By avoiding these extreme cases in our experimental displays, we hoped to ensure that the relevant information for perceiving structure from motion would remain well within the range of human sensitivity in all conditions.

EXPERIMENT 1

Experiment 1 was designed for the purpose of investigating the ability of human observers to determine the relative 3-dimensional lengths of moving line segments oriented in different directions.

Method

Each stimulus display depicted two solid line segments connected at the origin, which rotated in depth about a vertical axis. One of these line segments pointed upward relative to the horizontal, whereas the other pointed downward. The observer's task on each trial was to indicate which line segment (upward or downward) appeared longer in 3-dimensional space. The displays were generated with computer simulation such that the relative 3-dimensional length ratio (longer/shorter) could have two possible values of 1.1 or 1.3.

In order to investigate the effects of stimulus complexity, two different types of structural configuration were employed: *3-point displays*, in which the two solid line segments were presented in

Figure 3. A 2-frame sequence of images from the 23-point condition of Experiment 1. The two solid line segments have a relative 3-dimensional length ratio of 1.3, and are specifically oriented so that the longer segment in 3-dimensional space has the shorter projected image within the picture plane. The 3-dimensional structure of the depicted configuration can be observed by fusing the two images stereoscopically.

isolation; and *23-point displays*, in which they were accompanied by a connected sequence of five dotted lines, each containing four equally spaced dots (see Figure 3).

We also manipulated the length of the apparent motion sequences from a set of possible sequence lengths of 2, 3, 4, 6, or 8 distinct frames, with a 3° rotary displacement of the depicted object at each frame transition. Each sequence was presented in continuous oscillation until an appropriate response was recorded.

The displays were generated at random on each trial, subject to the following constraints: the initial projected lengths of the two solid line segments varied between 100 and 400 pixels, with a 2-dimensional projected length ratio (longer/shorter) between 1.01 and 1.40. From the initial frame of an apparent motion sequence until it reversed direction, the line segment with the longer projected length became progressively longer and slower at each frame transition, with an average displacement between 3 and 6 pixels. Similarly, the line segment with the shorter projected length became progressively shorter and faster, with an average projected displacement between 7 and 10 pixels. The displays were further constrained so that on half of the trials of each condition the longer segment in 3-dimensional space would projectively correspond to the shorter (and faster) segment within the 2-dimensional plane of the display screen. The latter constraint was maintained, moreover, on every frame of the apparent motion sequence. As a consequence of this control, the observers' performance would be no greater than chance if their responses were based solely on the relative projected length of the segments or on their relative projected velocity.

Figure 3 shows a possible 2-frame apparent motion sequence of a 23-point display. In generating this sequence, the simulated line segments in 3-dimensional space had a relative length ratio of 1.3, and were specifically oriented so that the longer segment in 3-dimensional space would be projectively shorter within the picture plane. If this were an actual experimental display, the 2 distinct frames would be presented sequentially over time. In order to obtain a general sense of the overall appearance of structure from motion, a phenomenally similar effect can be obtained by fusing the two patterns stereoscopically.

Using all possible combinations of these different manipulations, the resulting experimental design had 20 distinct conditions (2 length differences × 2 levels of structural complexity × 5 sequence lengths). One of the authors (J.T.) and 4 naive observers evaluated the displays over eight experimental sessions. Altogether, each condition was presented 48 times to each observer.

At the beginning of the first experimental session, the observers were given 20 trials of practice to familiarize themselves with the task. All of the observers reported that the vast majority of displays appeared quite clearly as a configuration of line segments rotating rigidly in 3-dimensional space. They all complained, however, that the task was quite difficult, and that they had little confidence

in the accuracy of their judgments. No feedback regarding their performance was provided until all eight sessions were completed.

Results and Discussion

The combined results of all 5 observers are presented in Figure 4. As is evident from the figure, the 3-dimensional length ratio of the simulated line segments had a significant effect on the observers' performance [$F(1,4) = 34.2, p < .01$], accounting for over 87% of the treatment sum of squares. Observers were 62.3% accurate when the length ratio was 1.1, and 79.1% accurate when the length ratio was 1.3. The structural complexity of the depicted objects and the number of frames in the apparent motion sequences had no significant effects on the observers' judgments. There were also no significant interactions, and no improvements with practice over the eight experimental sessions.

It is clear from these data that observers' perceptions of euclidean structure from motion do not have a high degree of accuracy. Because of the various constraints on our stimulus generation procedure, the largest 3-dimensional length ratio (longer/shorter) we were able to generate was 1.3. This would be considered a relatively large difference in most natural contexts—that is, it is analogous to judging whether a 5-ft.-tall jockey appears bigger than a 6.5-ft.-tall professional basketball player—yet observers' judgments in this condition were incorrect on over 20% of the trials. Indeed, if the shorter line segment of each pair is adopted as a standard, then the estimated Weber fraction for this task is 0.25, which is an order of magnitude larger than those obtained for many other types of sensory discrimination (e.g., see Teghtsoonian, 1971).

From a theoretical point of view, the most surprising aspect of these results is that performance does not improve as the number of distinct frames in an apparent motion sequence increases from 2 to 8. Our results are similar
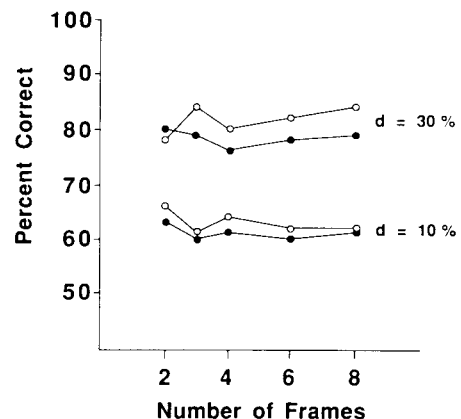
Figure 4. The percentage of correct length discrimination judgments for 5 observers as a function of the number frames in an apparent motion sequence for the two possible 3-dimensional length ratios employed in Experiment 1. The 3-point and 23-point conditions are represented by filled and open circles, respectively.

in this regard to those reported previously by Braunstein et al. (1987; Braunstein et al., 1990), but they contrast sharply with a recent series of experiments by Hildreth et al. (1990), in which the number of frames in an apparent motion sequence had a significant influence on judgments of 3-dimensional structure. We believe that this discrepancy is probably due to a fundamental difference in the method of stimulus presentation. In the displays employed by Hildreth et al. (1990), each apparent motion sequence was terminated after a single cycle, so that all variations in the number of distinct frames were confounded by variations in total stimulus duration. For the displays employed in the present experiment and by Braunstein et al. (1987; Braunstein et al., 1990), in contrast, each sequence was presented in continuous oscillation—first in one direction, then in reverse—until an appropriate response was recorded. Thus, the observers in these experiments could sample the available information for as long as was necessary in order to arrive at a maximally accurate judgment. Within this context of unlimited viewing time, the available evidence shows that increasing the number of distinct frames of an apparent motion sequence beyond two has little or no effect on observers' perceptions of structure from motion.

It is important to keep in mind that existing algorithms for computing euclidean structure from motion require a minimum of three distinct views in order to obtain a unique interpretation of an object's 3-dimensional form (see Bennett et al., 1989; Huang & Lee, 1989; Ullman, 1979, 1983). Although some algorithms may make an initial estimate of an object's structure from the first 2 frames of an apparent motion sequence (e.g., Grzywacz & Hildreth, 1987; Ullman, 1984), they are designed specifically to refine these estimates with greater and greater precision as each new frame is presented. Since all of the information used by actual human observers seems to be available within 2-frame apparent motion sequences, it is reasonable to conclude that these existing algorithms may have little in common with the basic processes of human perception.

Let us now consider these results from the perspective of an affine analysis. On the basis of Equation 4, the 3-dimensional length L_3 of a rigidly moving line segment with one fixed endpoint can be determined by:

$$L_3^2 = L_2^2 + (D_2/\epsilon)^2, \qquad (5)$$

where L_2 is its projected length in the picture plane, D_2 is the projected displacement of its moving endpoint, and ϵ is the rotary displacement of the segment in 3-dimensional space, which is unknown. Note that the relative contributions of L_2 and D_2 in this equation are scaled by the unknown parameter $1/\epsilon$. If $1/\epsilon$ is assumed to be very small, then the line segment's extension in depth will tend to be underestimated, and its perceived 3-dimensional length will be primarily determined by L_2. If $1/\epsilon$ is assumed to be very large, on the other hand, then the line segment's extension in depth will tend to be overestimated,
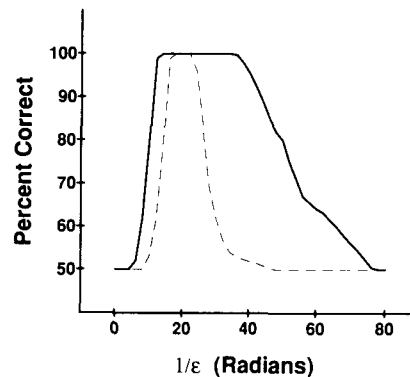

Figure 5. The expected performance of an ideal observer based on Equation 5 for both possible length ratios of Experiment 1 over a wide range of assumed values for the unknown scaling constant $1/\epsilon$. The solid and dashed lines represent length ratios of 1.3 and 1.1, respectively.

and its perceived 3-dimensional length will be primarily determined by D_2.

Figure 5 shows the expected performance of an ideal observer as a function of the assumed value of $1/\epsilon$ for the entire set of stimulus displays employed in the present experiment. The predicted levels of accuracy based on Equation 5 are plotted separately for both possible length ratios, with a range of values for the unknown scaling constant that spans the continuum between complete reliance on L_2 and complete reliance on D_2 in the determination of relative 3-dimensional length. Note in the figure that there is no one value of $1/\epsilon$ that can account for the actual data presented in Figure 4. Suppose, however, that the assumed value of the scaling constant was selected at random on each trial. The expected levels of accuracy for this strategy can be estimated by computing the area under each curve and dividing by the sampling range of $1/\epsilon$. On the basis of this calculation, the predicted accuracy would be 60% for displays with a length ratio of 1.1 and 82% for displays with a length ratio of 1.3—both of which are in close correspondence with the observers' judgments. It would appear to be the case, therefore, that an analysis of affine structure from motion can account for all of the major results of this experiment, including the specific levels of accuracy, and that the absence of any significant improvements in performance as the number of frames in an apparent motion sequence is increased beyond 2.

In an effort to explore the overall generality of these results, we have also examined a number of variations on the same basic experimental paradigm. For example, one such variation is to add a component of curl to the overall pattern of image motion, to test the hypothesis that performance might be impaired when the individual image velocities must be appropriately transformed before they can be used for an analysis of affine structure. Another is to embed the moving line segments within a planar grid of dots (see Figure 6), to test a hypothesis of

STRUCTURE FROM MOTION   425

Figure 6. A 2-frame sequence of images used to study the perception of relative 3-dimensional length in different directions along a planar surface. The results obtained using this type of display are identical to those reported for Experiment 1 of the present series.

Lappin (1990) that observers may be more attuned to metric distances along smooth surfaces than they are to distances in empty euclidean space. None of these manipulations seem to have any significant effect on performance, however. That is to say, we have always obtained the same basic pattern of results as is shown in Figure 4, regardless of the specific type of rigid transformation by which the line segments are displaced in 3-dimensional space or the specific type of structural configuration in which they are embedded.[2]

One other surprising aspect of these results that deserves to be highlighted concerns the possible effects of "flatness cues" (e.g., the absence of any changes in accommodation). From previous research on static form perception, it would be reasonable to expect that the presence of salient flatness cues should significantly reduce an object's perceived extension in depth. With respect to the present experiment, we would expect in that case that the observers' relative length judgments would be primarily determined by L_2, and that the line segment with the longer projected length would therefore appear to be 3-dimensionally longer as well on a majority of the trials. The data, however, do not support this hypothesis. Out of 4,800 possible responses in the present experiment, the observers selected the projectively longer line segment on only 45% of the trials.[3] This finding indicates that their errors were biased toward seeing *more* depth than was appropriate for the actual geometry of the displays.

EXPERIMENT 2

Experiment 2 was designed to provide converging evidence for the preceding conclusions, using an angle discrimination task.

Method

Each stimulus display depicted two connected solid line segments and four dotted line segments, which rotated in depth about an axis that was slanted 30° with respect to the picture plane. The observer's task on each trial was to indicate whether the 3-dimensional angle between the two solid line segments was larger or smaller than 90°. Two levels of difficulty were employed on alternating blocks of trials: a 45° difference condition, in which the two possible angles were 67.5° and 112.5°; and a 30° difference condition, in which the two possible angles were 75° and 105°.

The patterns of apparent motion had three possible sequence lengths of 2, 4, or 8 distinct frames, with a 3° rotary displacement of the depicted object at each frame transition. Each sequence was presented in continuous oscillation until an appropriate response was recorded.

The displays were generated at random on each trial subject to the following constraints: The 3-dimensional lengths of the moving line segments varied between 225 and 375 pixels. For each possible 3-dimensional angle of the two solid line segments, the corresponding 2-dimensional angle in the picture plane was allowed to vary between 30° and 80° on half the trials, and between 100° and 150° on the remaining half. (The latter constraint was maintained on every frame of an apparent motion sequence.) Thus, as a consequence of this control, the observers' performance would have been no greater than chance if their responses were based solely on the 2-dimensional projected angle between the moving line segments.

Including all possible combinations of these different manipulations, the resulting experimental design had six distinct conditions (2 levels of difficulty × 3 sequence lengths). The 2 authors and 3 naive observers evaluated the displays over four experimental sessions. Altogether, each condition was presented 80 times to each observer.

At the beginning of the first experimental session, the observers were given 20 trials of practice to familiarize themselves with the task. As in the previous experiment, all of the observers agreed that the vast majority of the depicted configurations appeared quite clearly to be rotating rigidly in 3-dimensional space. No feedback regarding their performance was provided until all four sessions had been completed.

Results and Discussion

The combined results for all 5 observers are shown in Figure 7. An analysis of variance for these data revealed a significant difference between the two levels of difficulty [$F(1,4) = 44.55, p < .01$] and a significant effect of sequence length [$F(2,8) = 29.75, p < .01$]. It is important to point out, however, that although the latter effect was statistically reliable, it had a negligible influence on the overall accuracy of the observers' judgments. Indeed, the level of performance improved by only 5% as the number of frames in the apparent motion sequences was increased from 2 to 8.
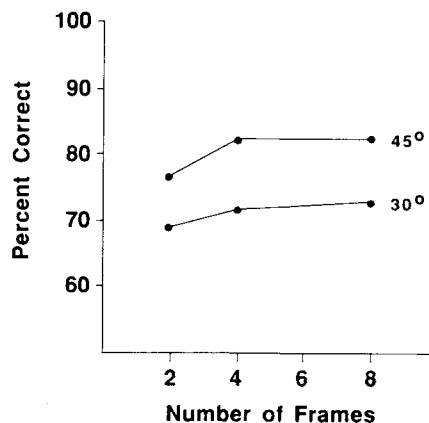

Figure 7. The percentage of correct angle discrimination judgments for 5 observers as a function of the number of frames in an apparent motion sequence for the two possible difference conditions employed in Experiment 2.

Although the observers' accuracy on this task was far above chance, their overall level of performance was surprisingly poor, considering the large differences between the angles they were asked to discriminate. On the basis of an interpolation of the data, the 75% difference threshold for discriminating the depicted angles was 36°. If the smaller of the two angles in each block of trials is adopted as a standard, then the estimated Weber fraction for this task is 0.5. Like the Weber fraction for judgments of 3-dimensional length in Experiment 1, this value is many times larger than those obtained for other types of sensory discrimination.

Another important aspect of these results that confirms the findings of Experiment 1 is that there were negligible improvements in performance as the number of frames in the apparent motion sequences was increased from 2 to 8. It should again be noted that existing algorithms for computing euclidean structure from motion require a minimum of 3 distinct frames in order to be effective. If all of the perceptually useful information for observers' judgments is available within 2-frame apparent motion sequences, as suggested by our results, it is reasonable to conclude that these existing algorithms may have little relevance to the processes of human perception.

One plausible interpretation for all of the data reported thus far is that observers' perceptions of kinetic depth displays are based primarily on an analysis of affine structure. Because this type of analysis is designed to be used with 2-frame apparent motion sequences, there would be no reason to expect that performance must necessarily improve as the number of distinct frames in a sequence is increased beyond 2. Moreover, because an affine analysis is inherently inadequate for the tasks employed in Experiments 1 and 2, it would be reasonable to expect that observers' performance on these tasks should be relatively inaccurate. As we pointed out in the introduction, however, there are other possible tasks, such as the detection of rigidity or planarity, for which an affine analysis of structure from motion is better suited. If our hypothesis is correct, then these tasks should be significantly easier than those involving euclidean metric structure. The remaining two experiments of the present series were designed to test this prediction.

EXPERIMENT 3

Experiment 3 was designed for the purpose of investigating the ability of human observers to discriminate planar and nonplanar configurations of moving line segments.

Method

Each stimulus depicted two configurations of three connected line segments—one in the upper half of the display screen and one in the lower half—rotating in depth about a vertical axis. Each chain of three segments formed two planar facets, whose relative orientation could be systematically manipulated. For one of the configurations in each display, these facets were coplanar. For the other they were not. The observers' task on each trial was to indicate which configuration (upper or lower) was coplanar. The difficulty
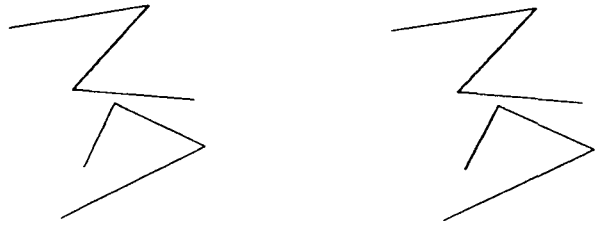

Figure 8. A 2-frame sequence of images from the displays employed in Experiment 3. Each connected sequence of three solid line segments defines two planar facets in 3-dimensional space. As can be observed by fusing the images stereoscopically, the two facets in the lower configuration are separated by an angle of 4°, whereas those in the upper configuration are coplanar.

of this discrimination was varied across trials by selecting the angle between the nonplanar facets from possible values of 2° or 4° (see Figure 8).

The patterns of apparent motion had three possible sequence lengths of 2, 4, or 8 distinct frames, with a 3° rotary displacement of the depicted object at each frame transition. Each sequence was presented in continuous oscillation until an appropriate response was recorded.

The displays were generated at random on each trial such that the 2-dimensional projected length of each line segment was allowed to vary between 100 and 300 pixels, and the 2-dimensional projected angle between each connected pair of line segments was allowed to vary between 30° and 150°.

Including all possible combinations of these different manipulations, the resulting experimental design had six distinct conditions (2 levels of difficulty > sequence lengths). One of the authors (J.T.) and 3 naive observers evaluated the displays over three experimental sessions. Altogether, each condition was presented 75 times to each observer. At the beginning of the first experimental session, the observers were given 20 trials of practice to familiarize themselves with the task. No feedback regarding their performance was provided until all three sessions were completed.

Results and Discussion

The combined results for all 4 observers are shown in Figure 9. An analysis of variance for these data revealed a significant difference between the two levels of difficulty $[F(1,3) = 78.05, p < .01]$ and a significant effect of sequence length $[F(2,6) = 14.08, p < .01]$. Although the latter effect is statistically reliable, it should also be pointed out that it is based on an improvement of only 9% as the number of frames in the apparent motion sequences was increased from 2 to 8.

The main thing to note in evaluating these data is the striking contrast with the results obtained in Experiment 2. The 75% threshold for detecting deviations from planarity in the present experiment was 2.6°, whereas the 75% threshold for the angle discrimination task employed in Experiment 2 was 36°. Although these tasks may appear at first blush to be superficially similar, they differ by more than an order of magnitude when compared with respect to their perceptual difficulty.

One possible explanation to account for this difference is that the perceptual analysis of structure from motion by actual human observers is restricted to the use of first
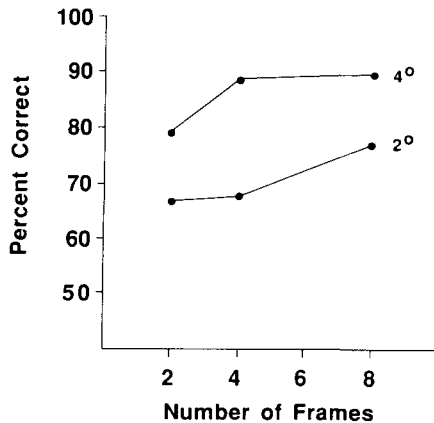
Figure 9. The percentage of correct planarity discriminations for four observers as a function of the number frames in an apparent motion sequence for the two possible levels of angular deviation employed in Experiment 3.

order temporal relations that are available within 2-frame apparent motion sequences. It is important to recognize that the detection of planarity is theoretically possible within this constraint. Any set of moving points with parallel image trajectories (see the Appendix) must be coplanar if there exists a set of constant coefficients a, b, c, and d, such that for any point i:

$$ax_i + by_i + cv_i + d = 0 \qquad (6)$$

where x_i and y_i are its instantaneous position coordinates within the image plane, and v_i is its transformed projected velocity. For the angle discrimination task employed in Experiment 2, in contrast, it would be necessary to perform an analysis of euclidean structure from motion requiring a minimum of 3 distinct frames (i.e., it would be necessary to detect second-order relations over multiple time intervals). Thus, the large differences in performance on these tasks, and the absence of any substantive improvements in performance with increasing sequence length, provides strong converging evidence that observers' perceptions of kinetic depth displays may be primarily based on an analysis of affine structure.

EXPERIMENT 4

Experiment 4 was designed to measure the ability of human observers to discriminate rigid from nonrigid configurations within apparent motion sequences of variable length.

Method

Each stimulus display depicted two solid line segments—one red and one green—and four white dotted line segments, which rotated in depth about an axis that was slanted 30° with respect to the picture plane. One of the solid line segments maintained a constant 3-dimensional length throughout the rotation sequence, whereas the length of the other was increased or decreased by a constant percentage at each frame transition. The observers' task on each trial was to indicate which line segment (red or green) was undergoing

a nonrigid change in length. To manipulate the difficulty of this discrimination, we varied the percentage length change at each frame transition across trials from possible values of 1%, 2%, 3%, or 4%.

The patterns of apparent motion had three possible sequence lengths of 2, 4, or 8 distinct frames, with a 3° rotary displacement of the depicted object at each frame transition. Each sequence was presented in continuous oscillation until an appropriate response was recorded.

The displays were generated at random on each trial, subject to the following constraints: The 2-dimensional projected lengths of the different line segments were allowed to vary between 100 and 300 pixels. The two solid line segments were always connected at the origin, and the angle formed by their 2-dimensional projections in the picture plane was restricted to a range between 30° and 150°.

Using all possible combinations of these different manipulations, the resulting experimental design had 12 distinct conditions (4 levels of difficulty × 3 sequence lengths). The 2 authors and 2 naive observers evaluated the displays over eight experimental sessions. Altogether, each condition was presented 100 times to each observer. At the beginning of the first experimental session, the observers were given 20 trials of practice to familiarize themselves with the task. No feedback regarding their performance was provided until all eight sessions had been completed.

Results and Discussion

The combined results for all 4 observers are shown in Figure 10. An analysis of variance for these data revealed that there was a significant difference between the four levels of difficulty [$F(1,3) = 299.41, p < .001$], but that there were no significant effects of sequence length and no significant interactions.

In evaluating the observers' performance on this task, it is useful to contrast the results with those obtained in Experiment 1. Note that, in Figure 4, the difference threshold for detecting the longer of two moving line segments in Experiment 1 was 25%, which is an order of magnitude larger than the 2% threshold obtained in Experiment 4 for detecting length changes of a single line segment over time. Although these tasks appear superficially similar, they are fundamentally different when considered from the perspective of a computational analysis.
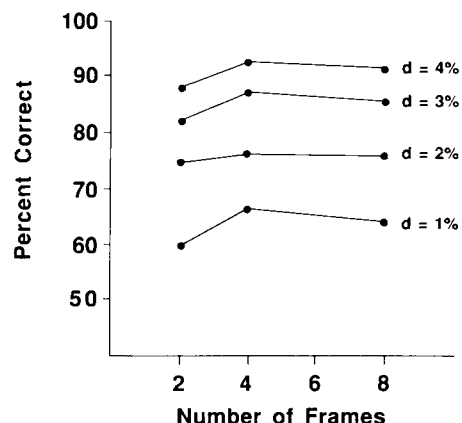

Figure 10. The percentage of correct rigidity discriminations for four observers as a function of the number of frames in an apparent motion sequence for the four possible levels of nonrigid length change employed in Experiment 4.

To accurately compare the lengths of two moving line segments, it would be necessary to perform an analysis of euclidean structure from motion, using a minimum of 3 distinct frames. However, to detect a nonrigid length change of a single line segment within an otherwise rigid configuration, it would only be necessary to perform an analysis of affine structure with as few as 2 distinct frames. The dramatic difference in performance on these tasks provides strong evidence that the processes of human perception may be limited to the analysis of affine structure (cf. Todd & Reichel, 1989).

It is important to keep in mind when considering these results that the length of a nonrigid line segment was increased or decreased by a constant percentage at each frame transition, so that the total amount of length change varied systematically with the number of frames in an apparent motion sequence (e.g., if the frame-to-frame length change was 3%, the total change for an 8-frame sequence would be 23%). We had originally expected that the accuracy of the observers' judgments would be primarily determined by the total length change in each display, but this is not what occurred. Although performance was influenced by the amount of length change per frame transition, the total change over an entire sequence had no detectable effect. It is especially interesting in this regard that a similar result has been obtained independently by Braunstein et al. (1990), using a very different paradigm to measure observers' sensitivity to nonrigid deformations. Taken together, these findings suggest that observers' perceptions of rigidity are determined solely from first-order relations between adjacent frames of an apparent motion sequence.

GENERAL DISCUSSION

Since the pioneering analysis of Ullman (1979), it has been generally accepted among perceptual psychologists that there are strict theoretical limits on the amount of information required to determine an object's structure from motion, including a minimum of three distinct views in all conditions that do not impose additional restrictions on the possible variations of an object's structure or its specific pattern of motion (e.g., see Hoffman & Flinchbaugh, 1982). During the past several years, however, there has been a growing amount of psychophysical evidence that these theoretical limits may have little relevance to the perceptual performance of actual human observers. Of particular importance in this regard are the recent findings of Braunstein et al. (1987; Braunstein et al., 1990) and Todd et al. (1988) that 2-frame apparent motion sequences under orthographic projection provide sufficient information for observers to obtain compelling kinetic depth effects and to accurately discriminate between different 3-dimensional structures (see also the earlier findings of Lappin et al., 1980, and Doner et al., 1984, for 2-frame sequences under polar projection).

The research described in the present article was designed for the purpose of investigating this phenomenon from a variety of perspectives. We began by presenting a mathematical analysis, extending the work of Ullman (1977, 1983), in which we attempted to delineate which specific tasks are theoretically possible from 2-frame apparent motion sequences and which ones are not. We then described a series of experiments designed to verify the psychological validity of this analysis by measuring how the length of an apparent motion sequence influences the accuracy of observers' judgments for a variety of different psychophysical tasks.

There are several noteworthy aspects of these experiments that are important to highlight. Each task involved a simple discrimination of some basic constituent of 3-dimensional form, such as the length of a line, the angle between two lines, or the angle between two planes. The stimulus displays were carefully controlled, moreover, so that above chance performance could not be achieved based on simple judgments of 2-dimensional optical properties such as projected length, projected angle, or projected velocity. Within these constraints, we attempted to optimize all other parameters in an effort to produce the best possible kinetic depth effects in each condition: We restricted the range of projected displacements to conform to the limits of human motion sensitivity; we optimized the timing parameters for each sequence length; and we presented these sequences in continuous oscillation so that observers could sample the available information for as long as was necessary to arrive at a maximally accurate judgment.

The results of these experiments provide compelling evidence that existing analyses of euclidean structure from motion may have little relevance to the processes of human perception. It is important to keep in mind that all such analyses require a minimum of three distinct views of an arbitrary moving object in order for a unique interpretation of its 3-dimensional form to be obtained, but that the information used by actual observers in the present experiments was readily available within 2-frame apparent motion sequences. That is to say, there were no appreciable improvements in performance on any of the four tasks as the number of frames in the apparent motion sequences was increased from 2 to 8.

One possible explanation for this theoretically surprising result is that human observers have poor sensitivity to the higher order spatiotemporal relations that are of fundamental importance to an analysis of euclidean structure from motion. Previous research has demonstrated, for example, that observers generally perform poorly on tasks such as the detection 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). On the basis of this evidence, it would appear to be the case that the human visual system is ill equipped

to perform higher order temporal derivatives, and that any analysis of structure from motion must therefore be restricted to the use of first-order temporal relations (i.e., those that are available within 2-frame apparent motion sequences).

Another source of converging evidence to support this hypothesis is provided by a comparison of the observers' performance on the four different experimental tasks. In our mathematical analysis of 2-frame apparent motion sequences, we demonstrated that first-order temporal relations are theoretically sufficient to determine properties of affine structure, such as rigidity or planarity, but that they are inherently inadequate for determining properties of euclidean structure, such as isotropic lengths or angles. The psychological validity of this theoretical distinction was clearly confirmed, moreover, in that the observers' performance on tasks involving affine structure was of an order of magnitude greater than their performance on similar tasks involving euclidean structure.

One important limitation of this analysis that needs to be highlighted is that in natural vision a pattern of parallel trajectories can only be obtained from image rotation when viewing distance is sufficiently large to approximate an orthographic projection. Although most previous investigations of the visual perception of structure from motion have involved displays that satisfy this condition, it is also possible to obtain compelling kinetic depth effects when displays are generated with large amounts of polar perspective (e.g., see Doner et al., 1984; Lappin et al., 1980; Lappin & Fuqua, 1983). It is especially interesting in this regard that a recent analysis by Longuet-Higgins (1981) has shown that 2-frame apparent motion sequences under polar projection provide sufficient information for observers to obtain a unique interpretation of an object's euclidean metric structure. Thus, it is quite possible that the performance limitations observed in the present experiments for parallel projections may not generalize to moving objects viewed at close range.

It is important to keep in mind, however, that it is also theoretically possible to determine an object's euclidean metric structure under parallel projection if the apparent motion sequence contains 3 or more frames, but that human observers are apparently incapable of making use of this information. Indeed, there is a growing amount of evidence to suggest that this is not an isolated phenomenon. An analogous insensitivity to euclidean metric structure has previously been demonstrated in the recognition of line drawings (Biederman, 1987) and in the visual perception of smoothly curved surfaces from shading or texture (Todd & Reichel, 1989). When considered as a whole, these findings provide strong evidence that the geometry of perceived 3-dimensional form may be much more abstract than is generally taken for granted.

REFERENCES

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.

BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

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.

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.

GRZYWACZ, N., & HILDRETH, E. (1987). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. *Journal of the Optical Society of America A*, **4**, 503-518.

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., & 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.

KOENDERINK, J. J., & VAN DOORN, A. J. (in press). Affine structure from motion. *Journal of the Optical Society of America A*.

LAPPIN, J. S. (1990). Perceiving metric structure of environmental objects from motion, self-motion and stereopsis. In R. Warren and 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.

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

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.

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.

TEGHTSOONIAN, R. (1971). On the exponents of Stevens' law and the constant in Ekman's law. *Psychological Review*, **78**, 71-80.

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

TODD, J. (1982). Visual information about rigid and 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., & REICHEL, F. R. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, **96**, 643-657.

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.

ULLMAN, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, **13**, 255-274.

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

NOTES

1. Longuet-Higgins (1981) has demonstrated, however, that a unique interpretation of an object's 3-dimensional structure can be obtained from 2-frame apparent motion sequences under polar projection.

2. This excludes, of course, the degenerate case of rotation about the line of sight where $\epsilon = 0$.

3. This effect was obtained for 4 out of the 5 individual observers.

APPENDIX

The following analysis is adapted from Ullman (1977). It is assumed that an arbitrary configuration of points is rotating rigidly in 3-dimensional space, and that the instantaneous position of each point is defined with respect to a moving Cartesian coordinate system (x, y, z), where the origin is attached to some arbitrary feature point P_o, the z-axis is defined by a line connecting P_o with the point of observation, and the xy plane is perpendicular to the z-axis. Within this coordinate system, any rigid displacement of an object can be decomposed into two independent components: a rotation ϵ about an axis in the picture plane through P_o, followed by a rotation θ about the z-axis.

Let $P_{i,j} = (x_{i,j}, y_{i,j}, z_{i,j})$ be the instantaneous position of a point i at time j, and let $P'_{i,j} = (x_{i,j}, y_{i,j})$ be its orthographic projection in the xy plane. If we measure the position of a given point i at two different moments in time ($j = 0,1$), the slope m_i of its projected trajectory is given by

$$m_i = \frac{y_{i,0} - y_{i,1}}{x_{i,0} - x_{i,1}}. \tag{A1}$$

If $P'_{i,1}$ is rotated about the origin by an angle $-\theta$, the value of m_i will be transformed to

$$m_i = \frac{y_{i,0} - y_{i,1} \cos\theta - x_{i,1} \sin\theta}{x_{i,0} - x_{i,1} \cos\theta + y_{i,1} \sin\theta}. \tag{A2}$$

Let us now consider the projected trajectories of a pair of points ($i = 0,1$), and the specific value of θ needed to make them parallel to one another. This will occur when

$$\frac{y_{0,0} - y_{0,1} \cos\theta - x_{0,1} \sin\theta}{x_{0,0} - x_{0,1} \cos\theta + y_{0,1} \sin\theta}$$
$$= \frac{y_{1,0} - y_{1,1} \cos\theta - x_{1,1} \sin\theta}{x_{1,0} - x_{1,1} \cos\theta + y_{1,1} \sin\theta}. \tag{A3}$$

With appropriate rearrangements and substitutions this can be reduced to a quadratic in $\sin\theta$:

$$0 = (a^2 + b^2)\sin^2\theta + (2ac)\sin\theta + (c^2 - b^2), \tag{A4}$$

where

$$a = x_{0,0}\, x_{1,1} - x_{0,1}\, x_{1,0} + y_{0,0}\, y_{1,1} - y_{0,1}\, y_{1,0}, \tag{A5}$$

$$b = x_{0,0}\, y_{1,1} + x_{0,1}\, y_{1,0} - y_{0,0}\, x_{1,1} - y_{0,1}\, x_{1,0}, \tag{A6}$$

$$c = y_{0,0}\, x_{1,0} + y_{0,1}\, x_{1,1} - x_{0,0}\, y_{1,0} - x_{0,1}\, y_{1,1}. \tag{A7}$$

Barring the occurrence of degenerate cases (see Ullman, 1977), Equation A4 will have two possible solutions, only one of which will be consistent with the solutions obtained for other point pairs. If there are no common solutions among all possible point pairs, then the observed pattern of motion must necessarily be nonrigid.