

Discriminating rigid from nonrigid motion: Minimum points and views

MYRON L. BRAUNSTEIN, DONALD D. HOFFMAN, and FRANK E. POLLICK
University of California, Irvine, California

Theoretical investigations of structure from motion have demonstrated that an ideal observer can discriminate rigid from nonrigid motion from two views of as few as four points. We report three experiments that demonstrate similar abilities in human observers: In one experiment, 4 of 6 subjects made this discrimination from two views of four points; the remaining subjects required five points. Accuracy in discriminating rigid from nonrigid motion depended on the amount of nonrigidity (variance of the interpoint distances over views) in the nonrigid structure. The ability to detect a rigid group dropped sharply as noise points (points not part of the rigid group) were added to the display. We conclude that human observers do extremely well in discriminating between nonrigid and fully rigid motion, but that they do quite poorly at segregating points in a display on the basis of rigidity.

Human observers report seeing three-dimensional (3-D) relationships in certain changing two-dimensional (2-D) images—as, for example, in images that represent projections of rotating solid objects (Wallach & O'Connell, 1953) or projections of rotating patterns of dots (Braunstein, 1962; B. F. Green, 1961). There has been recent interest in the minimum numbers of points and views that must be presented in order for subjects to make accurate judgments about 3-D structure from 2-D images. This interest stems in part from theoretical analyses of the minimum conditions under which an ideal observer can infer 3-D structure from 2-D coordinates. In this paper, we relate psychophysical data to theoretical analyses for a particular judgment: the discrimination of rigid from nonrigid motion.¹

Lappin, Doner, and Kottas (1980) studied the ability of subjects to judge 3-D relationships on the basis of only two views. They added noise to polar projections of rotating rigid spheres by varying the number of points that were in correspondence between the views. They concluded that two views were sufficient for discriminating between different levels of noise applied to rigid structures. Braunstein, Hoffman, Shapiro, Andersen, and Bennett (1987) asked subjects to discriminate between same and different rigid structures on the basis of from two to six views of from two to five points. They found that human performance exceeded theoretical expectations, al-

though some of the accuracy may have resulted from subjects exploiting the correlation that exists between 3-D and 2-D interpoint distances: 2-D interpoint distances tend to be more similar for two projections of the same 3-D object than for two projections based on different 3-D objects.

Todd (1988) has provided further evidence that two views are sufficient for distinguishing between rigid and nonrigid motion. He had subjects rate the rigidity of the depicted motion for two, four, or eight views of 14 connected line segments. The nonrigid displays were created by having each line segment's end point rotate about an axis whose position and orientation with respect to the picture plane was selected at random. The mean ratings given by subjects for nonrigid and rigid displays were at opposite ends of a 5-point rating scale. This clear discrimination between rigid and nonrigid displays did not increase with views, possibly because the effect had already reached a ceiling in the two-view condition.

In psychophysical experiments based on Ullman's (1979) theorem (that 3-D structure can be recovered from three views of four noncoplanar points), Petersik (1987) studied discrimination of rigid from nonrigid motion and found that subjects could indeed make that discrimination with three views of four points. This study included only rotations about a vertical axis. Nonrigid motion was produced by taking rigid displays and displacing points horizontally or vertically in the 2-D projection. This method, however, does not provide a clear indication of a subject's ability to discriminate rigid from nonrigid motion. When nonrigid displays are produced by perturbing the 2-D trajectories of points in a rigid display, it may be possible to distinguish between rigid and nonrigid displays on the basis of the trajectories of individual points. The most obvious case is that of a parallel projection of dots rotating about a vertical axis with a perturbation inserted in the vertical direction. All of the unperturbed

This research was supported by Office of Naval Research Contract N00014-88-K-0354, National Science Foundation Grants BNS-8819565 and IRI-8700924, and DOD Grant N00014-87-G-0135. We would like to thank Bruce Bennett, Jill Nicola, Chetan Prakash, and Whitman Richards for helpful discussions, Laura Bertin for programming assistance, George Andersen for comments on an earlier draft of this paper, and Lionel Shapiro for conducting a series of experiments preliminary to those reported here. Correspondence should be addressed to Myron L. Braunstein, Department of Cognitive Sciences, University of California, Irvine, CA 92717 (e-mail: MLBRAUNS@orion.oac.uci.edu).

trajectories are horizontal lines; any perturbed trajectories in the nonrigid display may be detected merely because they deviate from horizontal lines. It is important that any task involving discriminations between rigid and nonrigid motion should require subjects to use relationships between points; the trajectories of individual points should not be discriminable between rigid and nonrigid displays.

Theoretical Developments

Theoretical investigations of structure from motion have proceeded in two directions. In the first, investigators have developed specific theorems, stating specific conditions in which image motion can be given a 3-D interpretation. In the second, investigators have developed a general framework, within which the specific theorems can be seen as special cases. In this section, we briefly review the progress on specific theorems. Discussion of the general framework is beyond the scope of this paper, and may be found elsewhere (Hoffman & Bennett, 1988; Bennett, Hoffman, & Prakash, 1989). We then discuss the distinction between detecting and recovering rigid structures in motion, a distinction critical to the experimental work we present.

The specific theorems can be distinguished, for convenience, along three dimensions: constraints, projection, and temporal mode. Constraints are required because of the fundamental ambiguity of structure from motion—namely, that any given dynamical image, no matter how rich in features or extended in time, has not just one 3-D interpretation, but, in principle, infinitely many. A dynamic 2-D image does not, by itself and without further constraints, specify a unique 3-D interpretation; it does not, because the components of motion and position along the lines of sight are lost in projection. So if, for example, one collected the data output by a video camera, it would not make sense to ask, without further constraints, what *the* 3-D interpretation of that data was. For this reason, each theorem employs some constraint.

The constraints employed so far can be grouped into two categories: those applicable to rigid motion and those applicable to nonrigid motion. The constraint of rigid motion has been proposed by many perceptual psychologists (Gibson & Gibson, 1957; B. F. Green, 1961; Hay, 1966; Johansson, 1975; Ullman, 1979; Wallach & O'Connell, 1953). The idea is that, of all possible 3-D interpretations of dynamic 2-D images, the rigid interpretations should be among the ones preferred. Kruppa (1913), building on work of Chasles (1855), first stated rigorously conditions in which image motion can be given a 3-D interpretation using a constraint of rigidity. Kruppa's result and others (Faugeras & Maybank, 1989; Huang & Lee, 1989; Longuet-Higgins & Prazdny, 1980; Ullman, 1979) allow arbitrary rigid motions. Other results have restricted the type of rigid motion: rotation about the vertical axis (Longuet-Higgins, 1982); rotation about an arbitrary fixed axis (Bobick, 1986; Hoffman & Bennett, 1986; Webb & Aggarwal, 1981); rotation at a constant angular velocity

(Hoffman & Bennett, 1985, 1986); and rotation in a single plane (Hoffman & Flinchbaugh, 1982). Nonrigid motion has been studied less (Bennett & Hoffman, 1985; Grzywacz & Hildreth, 1987; Koenderink & van Doorn, 1986; Ullman, 1984).

The projections employed in the theorems are two: orthographic and perspective. In orthographic projection, the distance of an object from the imaging surface has no effect on its image. If one uses cartesian coordinates (x, y, z) such that coordinates x and y lie in the imaging plane and z is orthogonal, then orthographic projection is the map $(x, y, z) \mapsto (x, y)$. In perspective projection, the distance of an object from the imaging surface does have an effect on its image, with greater distance leading to a smaller image. A simple model of this is given by the map $(x, y, z) \mapsto (x/z, y/z)$. Some analyses use a combination of orthographic and perspective projections, as in Ullman's (1979) polar-parallel projection.

The temporal modes employed are two: discrete and continuous. Discrete time analyses treat motion much like a video camera does—as a sequence of frames (Hoffman & Flinchbaugh, 1982; Huang & Lee, 1989; Longuet-Higgins, 1982; Ullman, 1979). Continuous time analyses treat motion in terms of vector fields and their spatial and temporal derivatives (Hoffman, 1982; Koenderink & van Doorn, 1975, 1976, 1981; Longuet-Higgins & Prazdny, 1980; Waxman & Worn, 1987). Again, combinations of these temporal modes are possible, though not common (Bobick, 1986).

We now consider one structure-from-motion theorem in modest detail—as an example of this line of investigation, and as an aid to understanding the distinction between detection and recovery. Bennett, Hoffman, Nicola, and Prakash (1989) prove the following result: Suppose that four points are moving in space. Suppose that one is given two distinct orthographic views of the points. And suppose that, between the two views, the four points move rigidly and are noncoplanar. Then the two views contain sufficient information to restrict the possible rigid interpretations to a one-parameter family. Moreover, if the four points do *not* move rigidly between the views, then, almost surely, the views have no possible rigid interpretation (this last statement was proved by Ullman, 1977).

We can now make clear, with the aid of this example theorem, the distinction between detection and recovery of a structure in motion. Informally, to *detect* rigid structures is to discriminate successfully between image data (here, the two views of four points) that have rigid interpretations from those that do not; to *recover* rigid structures is to assign a rigid interpretation to each set of image data that is compatible with a rigid interpretation. Detection is necessary for recovery, but not vice versa. In our example theorem, two views of four points are sufficient for one to detect rigid structures, but once detection has occurred, there is still an uncountable set of rigid interpretations that could be assigned. Rigidity alone, under these conditions, is an insufficient constraint to pick out one interpretation from this uncountable collection.

Hence one cannot, under these conditions, recover a rigid structure.

The example theorem states that for two views of any four moving points it is theoretically possible to determine whether or not the points could lie on a rigid structure. Our first objective in the present study was to determine whether human observers can discriminate rigid from nonrigid structures at this minimum combination of points and views. A finding that human observers can make this discrimination on the basis of two views of four points would provide support for the hypothesis that human observers can exploit a rigidity constraint. Our second objective in this research was to examine the robustness of this discrimination as noise is added to a rigid display. To do this, we measured the reduction in accuracy of this discrimination when points that were not part of the rigid structure were added to the display.

The first experiment was intended to demonstrate that rigid structures can be discriminated from nonrigid structures at the theoretical minimum level of two views of four points. In the second experiment, we studied how increasing the number of views affects accuracy of this discrimination. In the third experiment, we studied how the detection of rigid structures is affected by adding points that were not part of the rigid structure. This addresses a fundamental question: Is a rigidity constraint likely to be useful in human vision for segregating rigid from nonrigid motion, so that 3-D structure can be recovered for points that are moving rigidly?

EXPERIMENT 1

The principal objective in the first experiment was to establish that subjects can discriminate rigid from nonrigid structures on the basis of two distinct views of as few as four points. The discriminability of rigid and nonrigid structures depends of course on the set of nonrigid structures that are used in the noise trials. As we noted earlier, the nonrigid trials must be generated in a way that does not allow discrimination between rigid and nonrigid displays on the basis of the motions of individual points. This precludes the use of random motions in the nonrigid displays and of methods in which the 2-D trajectories of moving points are perturbed. Instead, both the rigid and nonrigid displays should be generated from the same individual dot motions. This suggests two methods of generating nonrigid displays, methods which can be used separately or in combination. The first assigns a different angular velocity to each point in the nonrigid displays, but has all points rotating about the same axis. The second assigns the same angular velocity to each point, but has each point rotate about a different axis. We chose the latter method because the first produces a relationship between 2-D and 3-D nonrigidity—namely, greater 2-D nonrigidity for the displays that are nonrigid in 3-D. (The measure of 2-D nonrigidity that we used is described in the Stimuli section.) Producing 3-D nonrigidity by vary-

ing the axis of rotation, as in the latter method, does not result in a consistent relationship between 2-D and 3-D nonrigidity.

Method

Subjects. The subjects were 4 undergraduate students, 1 graduate student, and 1 staff member, who were paid for their participation. Acuity of at least 20/40 (Snellen eye chart) was required in the eye used throughout the experiment. Three of the undergraduate students were run without feedback. These subjects had no knowledge of the purposes of the experiment. The remaining subjects were run with feedback. One of these subjects was the third author; the other 2 were generally familiar with the purposes of the experiment.

Design. We examined two independent variables: the number of points in a simulated object, and the presence or absence of feedback. The number of points was four, five, or six. Each subject responded to 80 signal trials and 80 noise trials at each of the three levels of points.

Stimuli. A stimulus display consisted of two views of four, five, or six light-green dots changing position against a dark background. The two views represented a sequence of orthographic projections of points undergoing rotations in three dimensions. Initial point positions were selected at random within the volume of a sphere. The axes of rotation were determined as follows: A total of 272 points were placed at approximately equal distances on the surface of a sphere. (This was done using a three-frequency dodecahedron approximation. See Pugh, 1976.) A set of potential axes of rotation was defined by connecting each of these points to the center of the sphere, with the constraint that the slant angle of each axis (relative to the viewing direction) fell within the range 45° - 90° . There were 34 axes that met this constraint. For *rigid* displays, all points in the display rotated about the same axis (which was randomly selected from the set of 34 axes). For *nonrigid* displays, each point rotated about a different axis, each randomly selected without replacement from the set of 34 axes. For both rigid and nonrigid displays, an angle of rotation was selected from a uniform distribution over integer values between 6° and 18° and was the same for all points in the display.

Displays were used in the experiment only if they met three criteria: (1) nearest neighbor correspondence, (2) minimum 2-D motion, and (3) minimum 3-D spacing. The nearest neighbor criterion required the 2-D position of each point in each view to be closer to the 2-D position of that point in the other view than to the position of any other point in the other view. The minimum 2-D motion criterion required that each point move between views a distance of at least 5% of the radius of the generating sphere. The minimum 3-D spacing criterion required all pairs of points, in any given view, to be separated by at least 5% of the radius of the generating sphere. These three criteria were imposed to help assure (1) correct correspondence matching, (2) clearly visible motion of all points, and (3) clear separation of all points.

We developed a measure of 2-D nonrigidity to determine whether the 2-D projections of nonrigid displays were less rigid than the 2-D projections of rigid displays. First, we computed the variance, across views, of the projected interpoint distances of each pair of points in a display. Then we computed the mean of these variances across pairs. This mean gave the measure of nonrigidity in the 2-D projection. An analysis of variance (ANOVA) was conducted on the stimulus displays, using the measure of 2-D nonrigidity for each randomly generated display as the dependent variable. The independent variables were 3-D rigid versus 3-D nonrigid displays and number of points. The 2-D nonrigidity did not differ significantly for the rigid and nonrigid displays [$F(1,79) < 1$]. The main effects of number of points on 2-D nonrigidity, and the interaction, were also not significant.

The stimulus onset asynchrony (SOA) between views was 400 msec. There was no interstimulus interval between views. In order to allow sufficient time for subjects to make a judgment, the two views were repeated until the subject responded, up to a maximum of 60 sec.

Apparatus. The stimuli were presented on a Hewlett-Packard Model 1321B X-Y Display with a P-31 phosphor, under the control of a PDP-11/83 computer. The maximum projected diameter of each simulated object occupied 821 plotting positions on the screen and subtended a visual angle of 2°. Points were refreshed at a rate of 17.5 Hz. The dot and background luminances at the screen were approximately 5 and 0.02 cd/m², respectively. Subjects viewed the displays through a tube that limited the field of view to a circular area 7.9° in diameter. A 0.5 neutral-density filter was inserted in the tube to remove any apparent traces on the CRT. The eye-to-screen distance was 1.7 m.

A metal and plastic model consisting of four white spheres rigidly connected by thin black rods was used to instruct the subjects. The subjects responded by pressing one of two switches, one labeled "rigid" and the other "nonrigid." The responses (and response latencies) were recorded by the PDP-11/83.

Procedure. Each subject participated in one practice session followed by four experimental sessions. Each session began with 9 practice trials followed by a random sequence of 120 trials, consisting of 20 signal and 20 noise trials at each of the three point levels. The trials were presented in three blocks of 43 trials each. There was a 2-sec delay between each trial and a 1-min rest period between each block.

Subjects were instructed to press the "rigid" switch if the display consisted of a group of dots that was moving rigidly and to press the "nonrigid" switch otherwise. A group of dots was defined as moving rigidly if "the distance from any dot to any other dot remains the same, no matter how the group is moved." The model was used to demonstrate the rigid group condition. Subjects who were to receive feedback were told that a single tone would indicate a correct response and that two tones would indicate an incorrect response. The room was darkened 2 min before the trials began.

Results

A signal detection paradigm (D. M. Green & Swets, 1966) was used to analyze the results, with the trials containing a rigid group serving as signal trials. (We consider some of the implications of this definition of signal trials in the Discussion section.) A d' measure was computed for each subject and stimulus condition, using the proportion of rigid group responses on signal (3-D rigid display) trials as the hit rate and the proportion of rigid

group responses on noise (no rigid group) trials as the false-alarm rate. Each d' was based on 160 trials, half of which were signal trials.

The significance of the d' scores was calculated for each subject and number of points, using Marascuilo's (1970, pp. 238-240) one-signal significance test. Table 1 lists these d' values. Of a total of 18 d' s (6 subjects, three numbers of points) 14 were significantly different from zero ($p < .05$). For feedback subjects, 8 (of a total of 9) were significant. For nonfeedback subjects, 6 (of a total of 9) were significant. The d' s for all feedback subjects and for one nonfeedback subject were significant at two views of four points. The d' s for all subjects were significant at two views of five points. The mean d' for the subjects given feedback was higher than for those not given feedback (0.84 vs. 0.67) and lower for four points (0.51) than for five and six points (0.90 and 0.85), but these differences were not statistically significant.

A measure of 3-D nonrigidity was developed to determine whether the amount of 3-D nonrigidity in the noise displays affected the d' results. This measure was the mean across pairs of points of the variances of the 3-D interpoint distances across views. (Specifically, let $\mathbf{p}_{ij} = (x_{ij}, y_{ij}, z_{ij})$ denote the position in space of point i in view j . Let $d_{ii'j}$ be the 3-D distance between \mathbf{p}_{ij} and $\mathbf{p}_{i'j}$. Let $\sigma^2_{ii'}$ be the variance of $d_{ii'j}$ over all views j . Then our 3-D nonrigidity measure is the mean of the $\sigma^2_{ii'}$ for all distinct i and i' .) The nonrigid displays were separated into two categories—*high* and *low* 3-D nonrigidity—according to whether nonrigidity was greater than or less than the median value. The proportion of false alarms was calculated separately for each category. The proportion of correct responses for the entire rigid group was used to calculate the hit rate. This provided separate measures of d' for nonrigid displays with low and high amounts of nonrigidity. Fifteen (of 18) d' s were significantly different from zero when the high nonrigidity displays were used in calculating the false-alarm rate, and 8 (of 18) were significantly different from zero when the low nonrigidity displays were used. The d' values were higher for the high nonrigidity displays than for the low nonrigidity displays in 16 of 18 comparisons (6 subjects \times 3 numbers of points). The mean d' s for the high nonrigidity and low nonrigidity displays were 0.99 and 0.54, respectively.

These results indicate that human observers can discriminate rigid from nonrigid structures at or near the minimum level at which this discrimination is theoretically possible: two views of four points. (This is the minimum level if one assumes orthographic projection and if no constraints other than rigidity are applied.) The discriminability of rigid from nonrigid motion depends on the nonrigidity in the noise trials, as reflected in our 3-D nonrigidity measure.

EXPERIMENT 2

In Experiment 2, we examine accuracy in the four-point condition as the number of views increases. Previous

Table 1
 d' Scores in Experiment 1

Subject	Number of Points		
	4	5	6
Feedback Group			
F.	0.865*	1.235*	1.635*
A.	0.550*	0.735*	0.280
T.	0.505*	0.800*	0.925*
No Feedback Group			
G.	0.345	0.715*	1.060*
L.	0.475*	1.210*	0.805*
O.	0.290	0.705*	0.405

* $p < .05$.

studies present mixed results for the effects of number of views on judgments related to recovery of 3-D structure and discrimination of rigid from nonrigid motion. Doner, Lappin, and Perfetto (1984) found increased accuracy with increasing numbers of views in discriminations between different levels of spatiotemporal correlation in polar projections of rotating dot spheres. Braunstein et al. (1987) found increasing accuracy with increasing numbers of views in discriminations between same and different 3-D structures. On the other hand, Todd (1988) found no increase in the discriminability of rigid from nonrigid structures as the number of views was increased beyond two. Theoretically, two views do contain sufficient information for discriminating rigid from nonrigid structures (Bennett, Hoffman, Nicola, & Prakash, 1989; Ullman, 1977), but a third view is required before a specific rigid structure can be recovered (Ullman, 1979). It is possible that human observers are more accurate in discriminating rigid from nonrigid motion when there is sufficient information to recover a specific structure. If this is the case, an increase in accuracy should be expected in the three-view over the two-view condition.

Number of views, however, cannot be studied in isolation. Only two of the following three variables can be held constant as the number of views is varied: (1) rate of presentation of the views, (2) amount of rotation between views, and (3) total amount of rotation in the sequence of views. We chose to hold the first two variables constant and to allow the total amount of rotation to vary with number of views. For our nonrigid displays, this resulted in an increase in our measure of 3-D nonrigidity with increasing numbers of views. It is thus possible that an increase in d' with increasing views could be attributed to an increase in nonrigidity in the noise trials (suggested by Todd, personal communication, May 1, 1989). If the effect of number of views was due to the increase in 3-D nonrigidity in the noise trials, we would expect that d' would increase steadily with increasing numbers of views, and that the increase in d' would result from a decrease in the false-alarm rate rather than an increase in the hit rate.

Method

Subjects. The subjects were 4 of the 6 subjects who had served in Experiment 1. Two subjects had received feedback in Experiment 1 and 2 had not.

Design. We examined two independent variables: number of views (2, 3, 4, 5, or 6) and SOA (66 or 400 msec). (Two levels were used, because Todd, Akerstrom, Reichel, and Hayes, 1988, found an interaction between number of views and SOA in determining ratings of rigidity.) All displays contained four points. Each subject responded to 60 signal trials and 60 noise trials at each of the 10 combinations of SOA and number of views.

Stimuli. The method of generating the stimuli was the same as that used in Experiment 1, with the following exceptions: The SOAs were 66 msec and 400 msec. The refresh rate for both SOAs was 15 Hz. The angles of rotation between views were randomly selected from a uniform distribution over integer values between 5° and 9°. For rigid displays having more than two views, a new axis of rotation was randomly selected for each additional view. This was done

so that the presence of different axes of rotation in a sequence could not be used to distinguish rigid from nonrigid displays. For nonrigid displays having more than two views, a new axis of rotation was selected for each point in each additional view.

An ANOVA was conducted on the stimulus displays, using the 2-D nonrigidity measure as the dependent variable. The independent variables were 3-D rigidity, SOA, and number of views. The 2-D nonrigidity was significantly different for the 3-D rigid and 3-D nonrigid displays [$F(1,59) = 10.8, p < .01$]. The 2-D nonrigidity measure increased significantly with number of views [$F(4,236) = 178.3, p < .01$]. There were no other significant effects or interactions. The significant effect of 3-D nonrigidity indicates that it was theoretically possible for subjects to discriminate 3-D rigid from 3-D nonrigid displays on the basis of 2-D nonrigidity. This seems unlikely, however, since the variance in the 2-D nonrigidity measure accounted for by 3-D nonrigidity was 0.3%, compared with 38.2% accounted for by number of views. The means of the 2-D nonrigidity measures were .0053 for the 3-D rigid displays and .0058 for the 3-D nonrigid displays. The means for the displays with two to six views were .0012, .0029, .0053, .0077, and .0107, respectively. The units are squared distances in a unit sphere.

Procedure. Each subject participated in 1 practice session followed by 10 experimental sessions. Each session began with 9 practice trials followed by a random sequence of 120 trials, consisting of 12 signal and 12 noise trials at each of the 5 view levels. The trials were presented in three blocks of 43 trials each. Half the experimental sessions were at the short SOA, the other half at the long SOA. The order of SOAs was alternated between sessions, with half the subjects beginning with the long SOA and the other half beginning with the short SOA. The procedure was otherwise the same as in Experiment 1.

Results

A d' was computed for each subject and stimulus condition (Table 2). For the short SOA, 15 of the 20 d' s were significantly different from zero ($p < .05$). Of the five that were not significant, three were at the two-view level and two were at the three-view level. For the long SOA, all 20 d' s were significantly different from zero ($p < .05$).

A two-way ANOVA was conducted with SOA and number of views as the independent variables. There were two significant effects. The main effect of SOA [$F(1,3) = 16.83, p < .05, \omega^2 = 0.08$] showed an increase in d' with longer SOA (1.46 vs. 1.09). The main effect of number of views [$F(4,12) = 29.16, p < .01, \omega^2 = 0.44$] showed

Table 2
 d' Scores in Experiment 2

Subject	Number of Views				
	2	3	4	5	6
66-msec SOA					
F.	0.905*	1.075*	1.530*	1.620*	2.030*
A.	0.200	0.390	0.460*	1.190*	1.315*
G.	0.000	0.260	1.315*	1.470*	1.575*
L.	0.490	1.630*	1.210*	1.810*	1.295*
400-msec SOA					
F.	1.190*	1.520*	2.225*	2.300*	3.035*
A.	0.670*	0.860*	1.110*	1.165*	1.745*
G.	0.715*	1.120*	1.400*	1.045*	1.460*
L.	0.825*	1.330*	1.620*	1.420*	2.495*

Note—SOA = stimulus onset asynchrony. * $p < .05$.

an increase in d' with greater numbers of views. Post hoc comparisons (Tukey's *HSD* test) showed significant differences for two views versus three, four, five, and six views; three views versus five and six views; and four views versus six views.

As in Experiment 1, d' s were calculated with the nonrigid displays divided into high and low 3-D nonrigidity subgroups. For the high nonrigidity displays, 36 of 40 d' s were significantly different from zero, with a mean d' of 1.50. For the low nonrigidity displays, 29 of 40 were significantly different from zero, with a mean d' of 1.07. The d' values were greater for the high nonrigidity displays than for the low nonrigidity displays in 37 of 40 comparisons (4 subjects \times 2 SOAs \times 5 numbers of views).

The relationship between number of views and 3-D nonrigidity, d' , hit rate, and false-alarm rate is shown in Figure 1. The 3-D nonrigidity measure increased with number of views. There was a corresponding decrease in the false-alarm rate. The hit rate remained constant, indicating that the increase in d' was due to a decrease in the false-alarm rate. This is the pattern of results that would be expected if the effect of number of views was due to the increase in the 3-D nonrigidity that occurred with increasing numbers of views. This provides a further indication of the subjects' sensitivity to variations in 3-D nonrigidity and confirms the usefulness of the 3-D nonrigidity measure as a predictor of performance in discriminating rigid from nonrigid motion.

EXPERIMENT 3

Two orthographic views of four points are theoretically sufficient to determine whether or not a 3-D motion is rigid (Bennett, Hoffman, Nicola, & Prakash, 1989; Ullman, 1977), and the results of Experiments 1 and 2 indicate that subjects can make this discrimination at these minimum levels of points and views. For displays containing more than four points, the same theoretical analysis can be used to determine whether a display contains any subset of four points that is moving rigidly. It is important to know whether subjects can also determine whether rigid motion is present under these conditions; the usefulness of a rigidity constraint would be severely limited if such a constraint could be applied only when all moving elements were part of the same rigid structure. Experiment 3 included displays in which four points were moving rigidly but which, in addition, had from one to four points that were not part of the rigid structure. The subject's task, rather than indicate whether the observed structure was rigid or nonrigid as in Experiments 1 and 2, was to determine whether the display contained at least four points that moved together rigidly.

Method

Subjects. The subjects were 3 of the 4 subjects from Experiment 2 and 1 graduate student who had not served in Experiments 1 or 2. Three of the subjects were naive as to the purposes of the experiment; 1 subject was the third author. As a precondition for par-

ticipating in this experiment, each subject was required to achieve a d' of 1.2 or better in a screening session, in which they responded to 100 trials of 12 views of four points. This criterion assured that subjects were performing, on trials with no noise points, at a level comparable to performance in Experiment 2. One of the 4 subjects failed to meet the criterion in the first screening session but succeeded in doing so in a second screening session.

Design. We examined two independent variables: number of views (2, 3, 4, or 12) and number of noise points (0, 1, 2, 3, or 4). Each subject responded to 60 signal trials and 60 noise trials at each of the 20 combinations of number of views and noise points.

Stimuli. The method of generating the stimuli was the same as that used in Experiments 1 and 2, with the following exceptions: The 2-D minimum motion criteria for a display had to be met for each point for at least one transition between views rather than for all transitions. This change was made because of difficulty in generating 12-view displays that satisfied the more stringent criterion. Also, there was a change of two parameters: SOA and range of rotation angle for transitions. Two SOAs were used, 80 msec and 240 msec. (These were selected on the basis of Todd's observations, personal communication, November, 1988, and our own observations of the SOAs required for perception of smooth motion for two-view and multiple-view displays.) The refresh rate for both SOAs was 12.5 Hz. The long SOAs were used for the two-view displays and the short SOA for the 3-, 4-, and 12-view displays. The angles of rotation were randomly selected from a uniform dis-

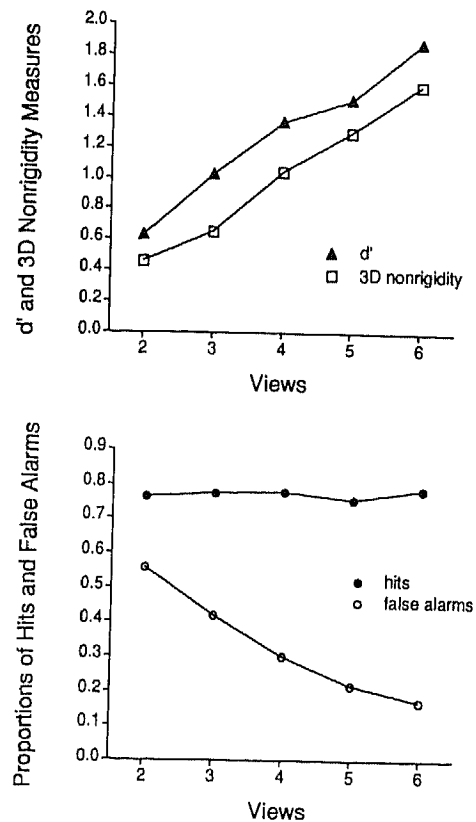


Figure 1. d' , 3-D nonrigidity, proportion of hits, and proportion of false alarms as functions of the number of views (Experiment 2). (In order to use the same ordinate values for d' and the 3-D nonrigidity measure, the nonrigidity measure is multiplied by 1,000 in this figure and in Figures 2 and 3.)

tribution of integer values between 5° and 7°. The larger rotation angles used in the previous experiments were eliminated, because they appeared to interfere with the perception of smooth motion at the 80-msec SOA.

The method used to generate the noise points added to the rigid displays was the same as that used to generate the points in the non-rigid displays, with the following additional restrictions: For each pair of views, the angle of rotation of the noise points was the same as that of the rigidly moving points, but no noise point was rotated about the same axis as that of the rigidly rotating points.

An ANOVA was conducted on the 2-D nonrigidity measure, with 3-D rigidity, number of views, and number of noise points as the independent variables. The only significant effect was the main effect of number of views [$F(3,177) = 1,510.0, p < .01$]. The means for 2, 3, 4, and 12 views were 0.0009, 0.0019, 0.0029, and 0.0133.

Procedure. Each subject participated in one or more screening sessions (described above), one practice session, and 24 experimental sessions. Each experimental session began with 5 practice trials followed by a random sequence of 100 trials, consisting of 10 signal and 10 noise trials at each of the 5 noise point levels. The trials were presented in three blocks of 35 trials each. There were 6 sessions at each of the 4 levels of number of views. The number of views across the 24 sessions was in the order 12, 4, 3, 2, 2, 3, 4, and 12 views, repeated three times.

As in Experiment 1, there was a 2-sec delay between each trial and a 1-min rest period between each block. The subjects were instructed to press the "rigid" switch if the display contained a group of dots that was moved together rigidly, and to press the "non-rigid" switch otherwise. A group of dots was defined as moving together rigidly if "at least four dots maintain constant distances from each other regardless of how the entire group moves."

Results

A d' was computed for each subject and stimulus condition (Table 3). Of 80 d' s, 48 were significantly different from zero ($p < .05$). For zero noise points, 15 (of 16) d' s were significantly different from zero. For four

Table 3
d' Scores in Experiment 3

Subject	Number of Noise Points				
	0	1	2	3	4
Two Views					
F.	0.740*	0.420	0.170	0.125	0.645*
M.	0.300	0.545*	0.000	-0.135	0.000
G.	0.695*	0.320	0.895*	0.000	0.105
L.	0.740*	0.380	0.555*	-0.045	0.305
Three Views					
F.	1.200*	0.815*	0.725*	0.160	0.320
M.	0.630*	0.505*	0.245	0.175	0.090
G.	0.550*	0.445	0.730*	0.490*	0.305
L.	1.045*	0.515*	0.310	0.385	0.715*
Four Views					
F.	0.950*	0.595*	0.465*	0.415	0.375
M.	1.040*	0.505*	0.260	0.565*	0.530*
G.	0.505*	0.375	-0.205	0.650*	0.510*
L.	0.940*	0.595*	0.275	0.180	0.220
Twelve Views					
F.	1.560*	0.850*	0.695*	0.945*	0.815*
M.	2.005*	0.865*	0.870*	0.480*	0.660*
G.	1.345*	0.465*	0.685*	0.660*	0.250
L.	1.330*	0.375	0.550*	0.800*	0.555*

* $p < .05$.

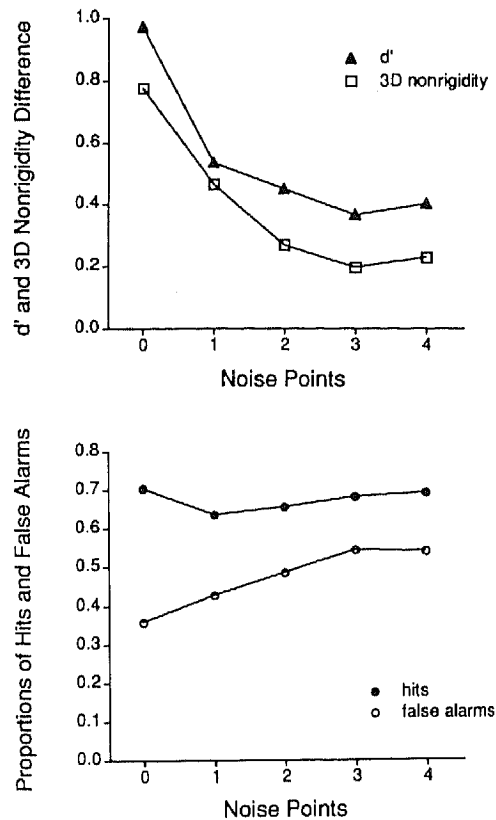


Figure 2. d' , difference in 3-D nonrigidity (noise nonrigidity - signal nonrigidity), proportion of hits, and proportion of false alarms as functions of the number of noise points in the signal displays (Experiment 3).

noise points, 7 (of 16) d' s were significantly different from zero.

The independent variables in the ANOVA were number of noise points and number of views. There were two significant effects. The main effect of number of noise points [$F(4,12) = 26.79, p < .01, \omega^2 = 0.34$] showed a decrease in d' with more noise points. The mean d' values for 0, 1, 2, 3, and 4 noise points were 0.97, 0.54, 0.45, 0.37, and 0.40, respectively. Post hoc comparisons showed only the differences between zero noise points and nonzero noise point conditions to be significant. The main effect of number of views [$F(3,9) = 10.43, p < .01, \omega^2 = 0.21$] showed an increase in d' with greater numbers of views. The mean d' s for 2, 3, 4, and 12 views were 0.34, 0.52, 0.49, and 0.84, respectively. Post hoc comparisons showed only the differences between 12 views and smaller numbers of views to be significant.

In the previous experiments, we examined the relationship between accuracy of discrimination and a measure of 3-D nonrigidity for the noise trials. For those experiments, the 3-D nonrigidity for the signal trials was always zero. In Experiment 3, 3-D nonrigidity increased for the signal trials as additional noise points were added.

It is likely that discriminability in this experiment was based on a relationship between 3-D nonrigidity in the signal trials and 3-D nonrigidity in the noise trials. We examined two obvious relationships: the ratio of the nonrigidity measure (signal trials/noise trials) and the difference in the measure (noise trials - signal trials). The correlations with d' , across the 20 combinations of views and noise points, were $-.65$ for the ratio measure and $.87$ for the difference measure. We therefore present the difference measure in Figures 2 and 3. Figure 2 shows the effects of number of noise points on d' and on the difference between noise and signal trials in 3-D nonrigidity. The hit rate and false-alarm rate are also shown. Figure 3 presents these effects as the number of views increases from 2 to 12. These results suggest that the difference in nonrigidity, or some related quantity, accounts both for the effects of points and for the effects of views. These effects are due primarily to changes in the false-alarm rate.

GENERAL DISCUSSION

On the basis of the rigidity constraint alone, human observers can discriminate rigid motion from nonrigid motion at the minimum level of points and views at which

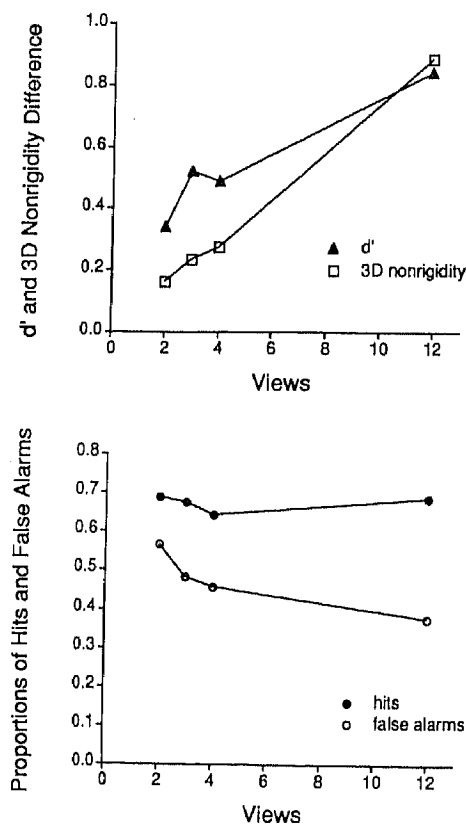


Figure 3. d' , difference in 3-D nonrigidity (noise nonrigidity - signal nonrigidity), proportion of hits, and proportion of false alarms as functions of the number of views (Experiment 3).

such discrimination is theoretically possible: two views of four points. For discriminations between displays in which all points were either moving rigidly or rotating about separate axes, accuracy depended on the deviation of the nonrigid displays from rigid motion. Our measure of this deviation, the mean across pairs of points of the variance in the interpoint distance over views, was related to the discriminability of rigid from nonrigid displays. This measure is based on the 3-D structure used to generate the displays. The usefulness of this measure is especially interesting in the case of the two-view displays, because the same two-view displays can be generated from an infinite number of rigid 3-D structures (Bennett, Hoffman, Nicola, & Prakash, 1989).

Increasing the number of points in a rigidly moving group does not lead to a clear increase in accuracy, although there was a nonsignificant increase from four to more than four points. It is certainly possible that an effect of points would be found for larger numbers of points—numbers sufficient to give the configuration a clear shape. Increasing the number of views did increase accuracy of discrimination, but this can be attributed to the increase in nonrigidity of the nonrigid displays. With points rotating about separate axes, the variance of the distances between pairs of points increases with number of views. Our measure of 3-D nonrigidity, based on these variances, correlated $.985$ with d' across the five levels of views.

Although human subjects can discriminate rigid from nonrigid structures at the minimum level of points and views at which this discrimination is theoretically possible, accuracy drops sharply when even one point that is not part of the rigid structure is added to a rigid display. It appears that human observers are not proficient at analyses that require testing subgroups of points to determine whether one subgroup is present that is moving rigidly. (With five points there would be five such subgroups to test. This may not seem to be much of a processing load from a computational viewpoint, but five subgroups involving six distances each in one display may be difficult for human subjects to process.) These results may appear to be in conflict with Ullman's (1979) well-known demonstration that two concentric cylinders differing in diameter are easily segregated by the human visual system. Ullman's stimuli, however, are not directly comparable to the present stimuli. Ullman used a large number of points and views, not the minimal numbers used in the present research. Perhaps more importantly, the motion in the demonstration was rotation about a fixed axis at a constant angular velocity. Bennett and Hoffman (1985) have shown that a fixed-axis constraint is sufficient mathematically for recovering 3-D structure from four orthographic views of two points or three orthographic views of four points; a rigidity constraint is not necessary. Demonstrations by Braunstein (1983) and Ramachandran, Cobb, and Rogers-Ramachandran (1988) also indicate that the perceptual segmentation of two rotating cylinders may not be based entirely on the use of a rigidity constraint.

The sharp drop in accuracy in detecting the presence of a rigid structure when noise points were added to the structure is consistent with Lappin et al.'s (1980) results with larger numbers of dots. In that study, accuracy in determining which of two displays had more coherent motion was highest when one of the displays was completely rigid, but dropped sharply when both displays contained nonrigid motion. If the subjects in the present experiments were primarily engaged in detecting nonrigid motion, rather than detecting rigid groups of points, it is not surprising that accuracy should have dropped sharply when both the signal trials and noise trials included nonrigid motion.

Discrimination between rigid and nonrigid structures, at least on the basis of small numbers of points and views, does not appear to be an easy task for human subjects. Subjective reports indicate that this task requires careful attention. It is possible that the task could be performed with less effort if the nonrigid motions differed even more from the rigid motions. In our displays, the same center of rotation was used for all points, whether or not they were part of a rigid structure. Generically, feature points that are moving independently would probably not have the same center of rotation. This probably made discriminations especially difficult in the present study, but it was necessary, to prevent a consistent relationship between nonrigidity in the 2-D projection and nonrigidity in 3-D.

In presenting a signal-detection analysis of the present experiments, we chose to define displays containing groups of at least four points moving together rigidly as signal displays, and displays lacking such rigid groups as noise displays. Our results suggest that the opposite interpretation may be worth considering. Discrimination of rigid motion from nonrigid motion may be conceived of as detecting deviations from constant interpoint distances in 3-D—that is, as detecting nonrigidity. Thus, in Experiments 1 and 2, the rigid displays might have been defined as the “noise displays” and the nonrigid displays as the “signal-plus-noise displays.” Increasing the 3-D nonrigidity of the nonrigid displays by increasing the number of views in Experiment 2 could then be described as increasing the signal strength, with the expected result of increasing d' . In Experiment 3, subjects may have been discriminating between levels of nonrigidity (i.e., between two levels of signal) rather than detecting rigid groups. Introspective reports suggest that subjects were both looking for rigid groups and looking for deviations from rigidity. The relationship between signal detection concepts and the discrimination of rigid from nonrigid motion would be worth exploring further with additional experimental manipulations.

In conclusion, these experiments reveal that human subjects are surprisingly good at some aspects of analyzing 3-D structures and surprisingly poor at others. Human subjects can discriminate rigid from nonrigid motion at exactly the minimum levels of points and views specified by theoretical analyses, suggesting that such analyses may be of relevance to the study of human vision. But when

the task is changed to determining whether a rigid structure is present in noise, performance falls off sharply with even one noise point. We need to look further into the issue of whether a rigidity constraint is useful in perceptual grouping, or whether other constraints must determine grouping before a rigidity constraint can be applied.

REFERENCES

- BENNETT, B., & HOFFMAN, D. (1985). The computation of structure from fixed axis motion: Nonrigid structures. *Biological Cybernetics*, *51*, 293-300.
- BENNETT, B., HOFFMAN, D., NICOLA, J., & PRAKASH, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A*, *6*, 1052-1069.
- BENNETT, B., HOFFMAN, D., & PRAKASH, C. (1989). *Observer mechanics: A formal theory of perception*. New York: Academic Press.
- BOBICK, A. (1986). A hybrid approach to structure from motion. In N. I. Badler & J. K. Tsotsos (Eds.), *Motion: Representation and perception* (pp. 91-109). New York: North-Holland.
- 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. (1983, June). *How flexible is the rigidity assumption?* Paper presented at the Second International Conference on Event Perception and Action, Nashville, TN.
- 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.
- CHASLES, M. (1855). Question No. 296. *Nouvelles Annales de Mathématiques*, *14*, 50.
- 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.
- FAUGERAS, O., & MAYBANK, S. (1989). Motion from point matches: Multiplicity of solutions. In *Proceedings of the IEEE Workshop on Visual Motion* (pp. 248-255).
- GIBSON, J., & GIBSON, E. (1957). Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology*, *54*, 129-138.
- GREEN, B. F., JR. (1961). Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology*, *62*, 272-282.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- 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*, *A4*, 503-518.
- HAY, C. (1966). Optical motions and space perception: An extension of Gibson's analysis. *Psychological Review*, *73*, 550-565.
- HOFFMAN, D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America*, *72*, 888-892.
- HOFFMAN, D., & BENNETT, B. (1985). Inferring the relative 3-D positions of two moving points. *Journal of the Optical Society of America*, *A2*, 350-353.
- HOFFMAN, D., & BENNETT, B. (1986). The computation of structure from fixed-axis motion: Rigid structures. *Biological Cybernetics*, *54*, 71-83.
- HOFFMAN, D., & BENNETT, B. (1988). Perceptual representations: Meaning and truth conditions. In S. Schiffer & S. Steele (Eds.), *Cognition and representation* (pp. 87-128). Boulder, CO: Westview Press.
- HOFFMAN, D., & FLINCHBAUGH, B. (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.
- JOHANSSON, G. (1975). Visual motion perception. *Scientific American*, *232*(6), 76-88.

- KOENDERINK, J., & VAN DOORN, A. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, **22**, 773-791.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1976). Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, **66**, 717-723.
- KOENDERINK, J., & VAN DOORN, A. (1981). Exterspecific component of the motion parallax field. *Journal of the Optical Society of America*, **71**, 953-957.
- KOENDERINK, J., & VAN DOORN, A. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America*, **A3**, 242-249.
- KRUPPA, E. (1913). Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Akademie der Wissenschaften in Wien: Mathematisch-naturwissenschaftliche Klasse Sitzungsberichte*, **122**, 1939-1948.
- LAPPIN, J. S., DONER, J. F., & KOTTAS, B. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science*, **209**, 717-719.
- LONGUET-HIGGINS, H. C. (1982). The role of the vertical dimension in stereoscopic vision. *Perception*, **11**, 377-386.
- LONGUET-HIGGINS, H. C., & PRAZDNY, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London*, **208B**, 385-397.
- MARASCUILO, L. A. (1970). Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrika*, **35**, 237-243.
- 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.
- PUGH, A. (1976). *Polyhedra: A visual approach*. Berkeley: University of California Press.
- RAMACHANDRAN, V. S., COBB, S., & ROGERS-RAMACHANDRAN, D. (1988). Perception of 3-D structure from motion: The role of velocity gradients and segmentation boundaries. *Perception & Psychophysics*, **44**, 390-393.
- TODD, J. T. (1988, November). *Minimal conditions for the perception of structure from motion*. Paper presented at the Twenty-Ninth Annual Meeting of the Psychonomic Society, Chicago, IL.
- 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.
- ULLMAN, S. (1977). *The interpretation of visual motion*. Unpublished PhD dissertation, Massachusetts Institute of Technology.
- ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT 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.
- WAXMAN, A., & WOHN, K. (1987). Contour evolution, neighborhood deformation, and image flow: Textured surfaces in motion. In W. Richards & S. Ullman (Eds.), *Image understanding 1985-86* (pp. 72-98). Norwood, NJ: Ablex.
- WEBB, J. A., & AGGARWAL, J. K. (1981). Visually interpreting the motion of objects in space. *Computer*, **14**(8), 40-46.

NOTES

1. Points move rigidly if all of their 3-D interpoint distances remain constant over time.

(Manuscript received June 16, 1989;
revision accepted for publication October 11, 1989.)